

# Comparative analysis of Grammar Checkers of various Asian Languages

Vikas Verma<sup>1\*</sup>, S.K. Sharma<sup>2</sup>

<sup>1,2</sup>Department of Computer Science and Applications, DAV University, Jalandhar, India.

\*Corresponding Author: vikas2005verma@yahoo.co.in, Tel.: +91-86997-57315

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 15/Oct/2018, Published: 31/Oct/2018

**Abstract**— The processing of natural languages using Computational Linguistics is an important domain of NLP (Natural Language processing). Many sentences are written in a natural language which may be defined as units, explicitly "linguistic units" which are meaningful and involve one or more words linked together in accordance with a set of predefined rules called 'Grammar'. Grammar checking is the task of detection and correction of grammatical errors in the text. Though a lot of work is done particularly for English language yet scant work is done for various Asian languages. This paper explores the contribution made by various researchers in this sphere. This paper critically analyses advancements made in relevance to global context. This facilitates better understandability, comparison and evaluation of previous research.

**Keywords**— Computational Linguistics, Natural Language Processing, Grammar checking, Grammatical errors.

## I. INTRODUCTION

Human beings communicate through a mode referred to as a "Language". Computational Linguistic is an interdisciplinary field of computer science and linguistics has collaboration with Artificial Intelligence area, and is concerned with computational aspects of human natural language[1]. Although existing technologies are far from attaining human abilities and have major feats in related application development, computers are not able to correspond to human thoughts. As people share knowledge, ideas, thoughts and information with each other using natural language, it is also possible to share the same with computer with the help of applied CL that too, in natural languages only. To complete communication and to make it meaningful, used language must follow set of rules involved in it. A grammar checking tool can provide automatic detection and correction of any faulty, unconventional or controversial usage of the underlying grammar. The generalized functionality of a Grammar Checker is depicted in figure 1.

This paper reports on a systematic review that focuses on various approaches for automatic detection and correction of grammar errors in various Asian language text. While reviewing the literature, summarization of many details is done to possible extent, explaining the complete architecture followed, along with its strengths and limitations. Our intention is to provide a platform for comparing the existing approaches that will help in taking further research decisions.

The paper is organized into following sections: Section II presents the Nuts and Bolts of communication mode. Section III presents the systematic review on the basis of legitimate parameters. Section IV presents a detailed and critical review of various approaches along with a comparative analysis. Finally, section V concludes our paper and suggests some directions for further research.

## II. NUTS AND BOLTS OF COMMUNICATION MODE

Natural languages provide a mode of communication within human beings with main intension of sharing information and knowledge transfer. Information can be made available either in text-form or speech-form. The building blocks are Morphemes, phonemes, phrases, words, vocabulary ,grammar, sentences. Rules of that language (grammar) should be duly followed for naming a sentence as "Valid". On the other hand, "Invalid" sentences are not effective to share knowledge. A Sentence in a natural language follows a structure and can be analyzed at the word level, part-of-speech level and at an arrangement of subject, object, verb etc. [1]. Language grammar rules are explicitly provided through instructions and are helpful for sentence formation. Sentences are generally classified as Simple, Complex and Compound on the basis of the internal structure of phrases.

For effective communication the syntax and semantics of a particular sentence is validated. Common grammatical mistakes

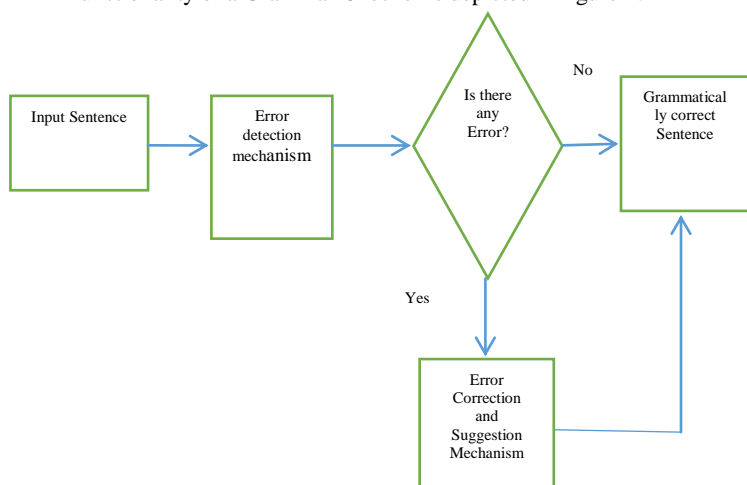


Figure. 1. Functional diagram of a Grammar Checker

include Punctuation position, Disagreement of noun / verb / gender etc., Tense mismatch and other factors [2]. Once rules are predefined, a grammar checker is used to examine the syntax of a given text followed by error reporting and providing suggestions for rectifications, if any [3]. An important constituent phase in the this processing is to perform preliminary operation on the given text and is commonly referred to as pre-processing. The general modules undergoing this process are Sentence Tokenization, Morphological Analysis, Part-of-speech tagging, Parsing.

### III. SYSTEMATIC REVIEW

A systematic review is different from a conventional review as it summarizes the existing work in comprehensive manner. Such review on grammar checking in Asian language as per the recommended guidelines [4] is provided here. Various parameters like Spelling, Syntax errors, Subject-Verb Agreements, Noun Number, Punctuation, Semantic errors, Error Detection, Error Correction, Sentence Category - simple, Compound or Complex are used for analysis and is depicted in table 1.

Table 1. Asian Languages Grammar Checking error detection criteria

Asian Language	Demographic region	Methodology followed	Number of Speakers (in millions)
Nepali [28]	Nepal, Bhutan, India, Burma	Rule Based	17
Urdu [29]	Pakistan, India, Bangladesh, Afganisthan, Thailand	Rule Based	104
Bangla [30]	Bangladesh, India, Andaman-Nicobar Islands	Data Driven Based	300
Punjabi [31]	India, Pakistan	Rule Based	130
Hindi [32]	India, Nepal	Rule Based	545

Grammar checking methodology may be categorized as Rule based [5], Statistical (Data driven) based [6] and Hybrid based [7]. Various researchers had handled complexity of large sentences in several ways. One such approach was followed while developing a tool referred as "Text Simplification for Reading Assistance" called KURA [8], another tool called "HAPPI" (Helping Aphasic People Process Online Information) [9], a tool for Lexical Simplification called "Putting It Simply: a Context-aware Approach to Lexical Simplification"-SIMPLEX [10], Use of a Psycholinguistic Database in the Simplification of Text for Aphasic Readers- PSET [11], development of sentence simplification tool for children's stories in Italian [12], Lexical Simplification for Spanish - "LexSiS" [13], Generation approach used for automatic Induction of rules for text simplification [14], splitting of long sentence after explanation generation [15], maintaining discourse when performing syntactic simplification [16], text simplification for information seeking applications [17], splitting of Vietnamese sentences for Vietnamese English machine translation [18], simplification of Newspaper text to assist

Aphasic reader [19], developing an authoring tool which provides text simplification techniques whilst writing a document [20], acquisition of syntactic simplification rules for French [21], direct manipulation of parse tree [22], automatic simplification of Bosque complex sentences using dependency tree [23], simplification of Korean sentences for deaf readers [24], sentence simplification to enhance multi-document summaries [25], removing unnecessary parts of sentences [26] and Spanish sentence simplification [27].

Table 2 represents the comparison of Languages, Methodology adopted along with the name of country wherein language is prominent, approximate number of speakers.

Table 2. Asian Languages categorization with demographic and methodology criteria

Concept / Approach	Spelling	Syntax errors	Subject-Verb agreement	Error Detection	Error Correction	Category of Sentence		
						Simple	Compound	Complex
[5]	√	√	√	√		√		
[6]		√		√	√	√		
[7]		√		√		√		
[8]	√	√	√	√	√	√	√	√
[9]		√		√		√		

### IV. CRITICAL ANALYSIS OF ASIAN GRAMMAR CHECKERS

After systematic review, critical analysis emphasizing features of Grammar Checkers along with their evaluation parameters is done and is summarized in table 3.

Table 3. Evaluation of Asian Grammar checkers with distinct features

Language	Methodology Adopted for Checking Grammar	Characteristics	Evaluation Features	Evaluation Parameters		Shortcomings
				Accuracy (%)	Revision (%)	
Nepali [28]	Rule Based	Shares Bangla and Hindi language primitives	Providing error information for simple sentences	Details not provided	Details not provided	Not suitable for Complex and Compound sentences
Urdu [29]	Rule Based	Formulates S-O-	Provides error	Details not	Details not	Lacks

		V agreeme nt	correction by checking structure and grammar	provid ed	provid ed	disa mbi guati on due to Mor phol ogy and POS
Bangl a [30]	Data Driven Based	Formulat es agreeme nt of Word	Provides better result	Uses thresh old value criteri a	Uses thresho ld value criteri a	Not suita ble for Com poun d sent ence s
Punja bi [31]	Rule Based	Word agreeme nt, Case formulat ion and agreeme nt with verb	Provides error correction suggestion	Detail s not provid ed	Detail s not provid ed	Fals e alar m situa tion alon g with chan ge in mean ing or class
Hindi [32]	Rule Based	Rich in inflectio n	Optimal result	Detail s not provid ed	Detail s not provid ed	Not suita ble for Com poun d and Com plex Sent ence

## V. CONCLUSION AND FUTURE SCOPE

An constructive application under the umbrella of Natural language Processing in Grammar Checking. Prominent applications include proof-reading and language learning. Though in last ten years, development of Grammar Checker as gained momentum but fever efforts were done as survey for the same. Our paper, presented a comprehensive, unbiased review of grammar checking techniques along with their shortcomings with parameters not limited to their features, performances and shortcomings. An conclusive set of observations highlights the following:

- No approach is full proof.
- Grammar checkers are not widespread for public at large.
- Semantic analysis, still remains a gray area.

d. For real time applications, tools lacks scalability and utility.

e. Amalgamation of various techniques can be carried out keeping in mind the accuracy with size of corpus used.

f. Though performance has improved gradually, yet there's ample scope of improvement.

## REFERENCES

- [1] N. S. Bhirud, R. P. Bhavsar, & B. V. Pawar, "Grammar Checkers For Natural Languages: A Review", International Journal on Natural Language Computing (IJNLC) , Vol. 6, No.4, pp.1-4, 2017.
- [2] C.C. Burch, P. Koehn, C. Monz, M. Post, R. Soricut, & L. Specia, Findings of the 2012 workshop on statistical machine translation, In Proceedings of the Seventh Workshop on Statistical Machine Translation, Montreal, Canada, pp. 10–51, 2012.
- [3] M.S. Gill, G.S. Lehal, "A grammar checking system for Punjabi", In 22nd International Conference on Computational Linguistics: Demonstration Papers. Association for Computational Linguistics, pp. 149-152, 2008
- [4] B. Kitchenham, "What's up with software metrics?—A preliminary mapping study", Journal of systems and software, Vol. 83, No.1, pp. 37-51, 2010.
- [5] D. Naber, "A rule-based style and grammar checker," Master's thesis, University of Bielefeld, 2003.
- [6] B. Manchanda, V.A. Athavale, & S.K. Sharma, "Various Techniques Used For Grammar Checking", International Journal of Computer Applications & Information Technology, Vol.9, No.1, pp.177-181, 2016.
- [7] S.K. Sharma, & G.S. Lehal, "Improving Existing Punjabi Grammar Checker", In International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT) IEEE, pp. 445-449, 2016.
- [8] k. Inui, A. Fujita, T. Takahashi, R. Iida, & T. Iwakura, "Text simplification for reading assistance: a project note", In Proceedings of the second international workshop on Paraphrasing Association for Computational Linguistics, Vol. 16, pp.9-16, 2003.
- [9] S. Devlin, G. Unthank, "Helping aphasic people process online information", In Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility, pp.225-226, 2006.
- [10] O. Biran, S. Brody, N. Elhadad, "Putting it simply: a context-aware approach to lexical simplification", In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers, vol. 2. Association for Computational Linguistics, Stroudsburg, pp. 496–501, 2011.
- [11] C. Napoles, M. Dredze, "Learning simple wikipedia: a cogitation in ascertaining abecedarian language", In Proc. of the NAACLHLT Workshop on Computational Linguistics and Writing. pp. 42–50, 2010.
- [12] G. Barlacchi & S. Tonelli, "ERNESTA: A sentence simplification tool for children's stories in Italian", In International Conference on Intelligent Text Processing and Computational Linguistics, pp. 476-487, Springer, Berlin, Heidelberg, 2013.
- [13] S. Bott, L. Rello, B. Drndarevic & H. Saggion, "Can spanish be simpler? lexis: Lexical simplification for spanish", In Proceedings of COLING , pp.357-374, 2012.
- [14] R. Chandrasekar, B. Srinivas, B, "Automatic induction of rules for text simplification", Knowledge-Based Systems, Vol.10, No.3, pp. 183-190, 1997.

- [15] S. Kandula, D. Curtis, Q. Zeng-Treitler, "A semantic and syntactic text simplification tool for health content", In AMIA annual symposium proceedings, pp. 366, 2010.
- [16] A. Siddharthan, "Syntactic simplification and text cohesion", Research on Language and Computation, Vol.4, No.1, pp. 77-109, 2006.
- [17] B.B. Klebanov, K. Knight, & D. Marcu, "Text simplification for information-seeking applications", In OTM Confederated International Conferences On the Move to Meaningful Internet Systems, pp. 735-747, Springer, Berlin, Heidelberg, 2004.
- [18] B.T. Hung, N.L. Minh, & A. Shimazu, "Sentence splitting for Vietnamese-English machine translation", In Fourth IEEE International Conference on Knowledge and Systems Engineering, pp. 156-160, 2012.
- [19] J. Carroll, G. Minnen, Y. Canning, S. Devlin, & J. Tait, "Practical simplification of English newspaper text to assist aphasic readers", In Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology, pp. 7-10, 1998.
- [20] C. Scarton, M.D. Oliveira, A. Candido, C. Gasperin, & S. Aluísio, "SIMPLIFICA: a tool for authoring simplified texts in Brazilian Portuguese guided by readability assessments", In Proceedings of the NAACL HLT 2010 Demonstration Session Association for Computational Linguistics, pp. 41-44, 2010.
- [21] V. Seretan, "Acquisition of syntactic simplification rules for french", 2012.
- [22] D. Febowitz, & D. Kauchak, "Sentence simplification as tree transduction", In Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations, pp. 1-10, 2013.
- [23] M.J. Aranzabe, A. D. Iarraza, & I. Gonzalez-Dios, "Transforming complex sentences using dependency trees for automatic text simplification in Basque", Procesamiento del lenguaje natural, Vol.50, pp. 61-68, 2013.
- [24] J.W. Chung, J.H. Min, J.C. Park, "Enhancing readability of web documents by text augmentation for deaf people", In Proceedings of the 3rd ACM International Conference on Web Intelligence, Mining and Semantics, pp. 30, 2013.
- [25] S.B. Silveira, & A. Branco, "Combining a double clustering approach with sentence simplification to produce highly informative multi-document summaries", In IEEE 13th International Conference on Information Reuse and Integration (IRI), pp. 482-489, 2012.
- [26] S. Klerke & A. Sogaard, "Simple, readable sub-sentences", In 51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop, pp. 142-149, 2013.
- [27] S. Stajner, B. Drndarevic, & H. Saggion, "Corpus-based sentence deletion and split decisions for spanish text simplification", 2013.
- [28] B. K. Bal, P. Shrestha, M.P. Pustakalaya, & N.P. Dhoka, "Architectural and System Design of the Nepali Grammar Checker. PAN Localization Working Paper", 2007.
- [29] H. Kabir, S. Nayyer, J. Zaman, & S. Hussain, "Two pass parsing implementation for an Urdu grammar checker", In Proceedings of IEEE international multi topic conference, pp. 1-8, 2002.
- [30] M. Alam, N. UzZaman, & M. Khan, "N-gram based statistical grammar checker for Bangla and English", Ninth International Conference on Computer and Information Technology (ICCIT), December, 2007.
- [31] G. Singh, "Development of Punjabi Grammar Checker. Phd. Dissertation", 2008.
- [32] G.D. LataBopche, M. Kshirsagar, "Grammar Checking System Using Rule Based Morphological Process for an Indian Language", In Global Trends in Information Systems and Software Applications, 4th International Conference, ObCom, pp. 9-11, 2011.

### Authors Profile

Mr. Vikas Verma is currently pursuing his Ph.D. from DAV University Jalandhar. He has done his MCA and Bachelors from Punjabi University, Patiala. He has 13 years of academic experience and 2 years of Industrial experience. He is currently working with DAV University as an Assistant Professor. He had earlier worked with several reputed Institutes / Universities of Northern India. His areas of interest include Machine Learning, Natural Language Processing, DBMS, Data Warehousing and Mining, Information Systems and Programming. He has presented 4 Research Papers in National Conferences, 6 Research Papers in International Conferences, has authored 10 Research Papers in International and National Journals, has attended 15 Faculty Development Programmes, conducted 4 Faculty Development Programmes, has delivered more than 4 Expert/Guest Lectures, authored 3 books and has been member of Board of Studies and various other professional bodies like UACEE, CSI, ISTE, etc. He is a Paper Setter of various universities of India.



Dr. Sanjeev Sharma is working as an Assistant Professor and Head of CSA Department in DAV University Jalandhar. His specialization is in Natural Language Processing (NLP) which is a sub-branch of Artificial Intelligence (AI). He has teaching experience of more than 16 years. He has published more than 33 research papers in international Journals. He has developed many basic NLP tools like POS Tagger, Clause Identifier, Sentence Identifier, Unknown Word Guesser, Morph etc. and one advance tool i.e. Grammar checker to detect and rectify/suggest the grammatical errors in Compound and Complex sentences in Punjabi language. He is currently guiding four Ph.D students.

