

Grammar Checker for Asian Languages: A Survey

Misha Mittal
Ph.D. Research Scholar
Guru Kashi University
Talwandi Sabo

Dinesh Kumar
Associate Professor
Guru Kashi University
Talwandi Sabo

Sanjeev Kumar Sharma
Assistant Professor
DAV University
Jalandhar

ABSTRACT

Grammar checker is one of the natural language processing applications that is used with most of the word processors. Various grammar checkers have been developed for Indian and foreign languages. In this research article author has performed a survey on the grammar checkers developed for Asian languages. In Asian languages mainly developed grammar checker are for Punjabi language, Urdu language, Nepali language and Bangla language. The architecture of all these grammar checker has been discussed in this paper.

Keywords

Hindi Grammar Checker, Punjabi Grammar Checker, Bangla Grammar Checker, Urdu Grammar Checker, Nepali, Chinese.

1. INTRODUCTION

A grammar checker is one of the basic Natural Language Processing (NLP) application or tools for checking the syntax of language. The natural language processing field is relatively new in Asian languages and a lot of tools have yet to be developed. One of these is a grammar checker. In general rule based and machine learning techniques have been used for developing a grammar checker. Resources in the field of natural language processing for Asian languages are limited. This, along with the complex grammar, presents some problems when developing a grammar checking system for the Asian languages. Besides these problems grammar checking system for some of the Asian languages have been developed. Some of the popular grammar checking system is a Punjabi grammar checker, a grammar checker for Urdu language, a Chinese grammar checker, a Nepali grammar checker, a Hindi grammar checker etc. Different techniques have been used for the development of these systems. The general architecture of grammar checker is shown in figure 1:

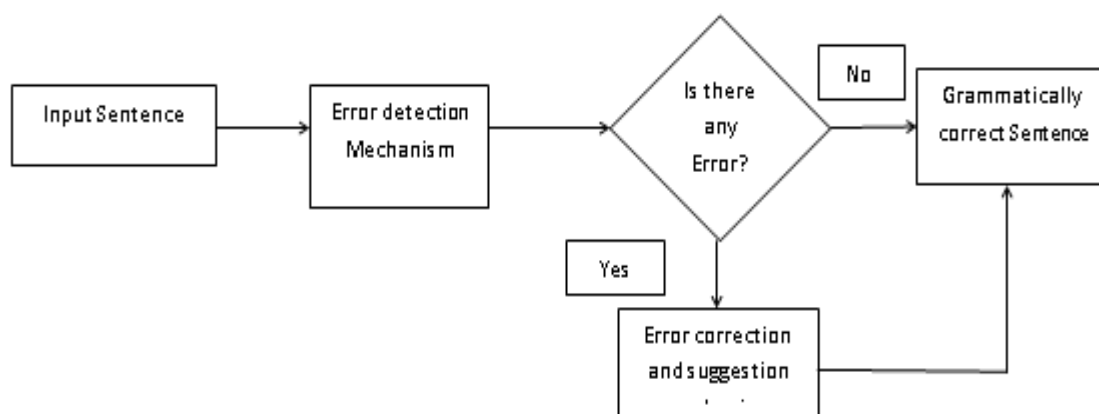


Fig 1: Simplified functional diagram of grammar checker

2. PUNJABI GRAMMAR CHECKER

Gill et al developed and represented a system for Punjabi grammar checker. This system uses a full form lexicon for morphology analysis and rule based systems for part of speech tagging and phrase chunking. The system is also supported by a set of carefully devised error detection rules, which can detect and suggest rectification for a number of grammatical errors, resulting from each of agreement, order of words in various phrases etc., in literacy style Punjabi texts. As this system

is a rule based system, so rules provided in the system can be turned on and off individually. This system also provides details of errors detected and suggestions for those errors. It can also detect any agreement errors in compound or complex sentences. The architectural diagram for this grammar checker is shown in figure 1.

As shown in figure 2, main components of Punjabi grammar checker are pre-processing, full form lexicon based morphological analyzer, rule based part of speech (POS) tagger, rule based phrase chunker and a rule based system for error detection and correction.

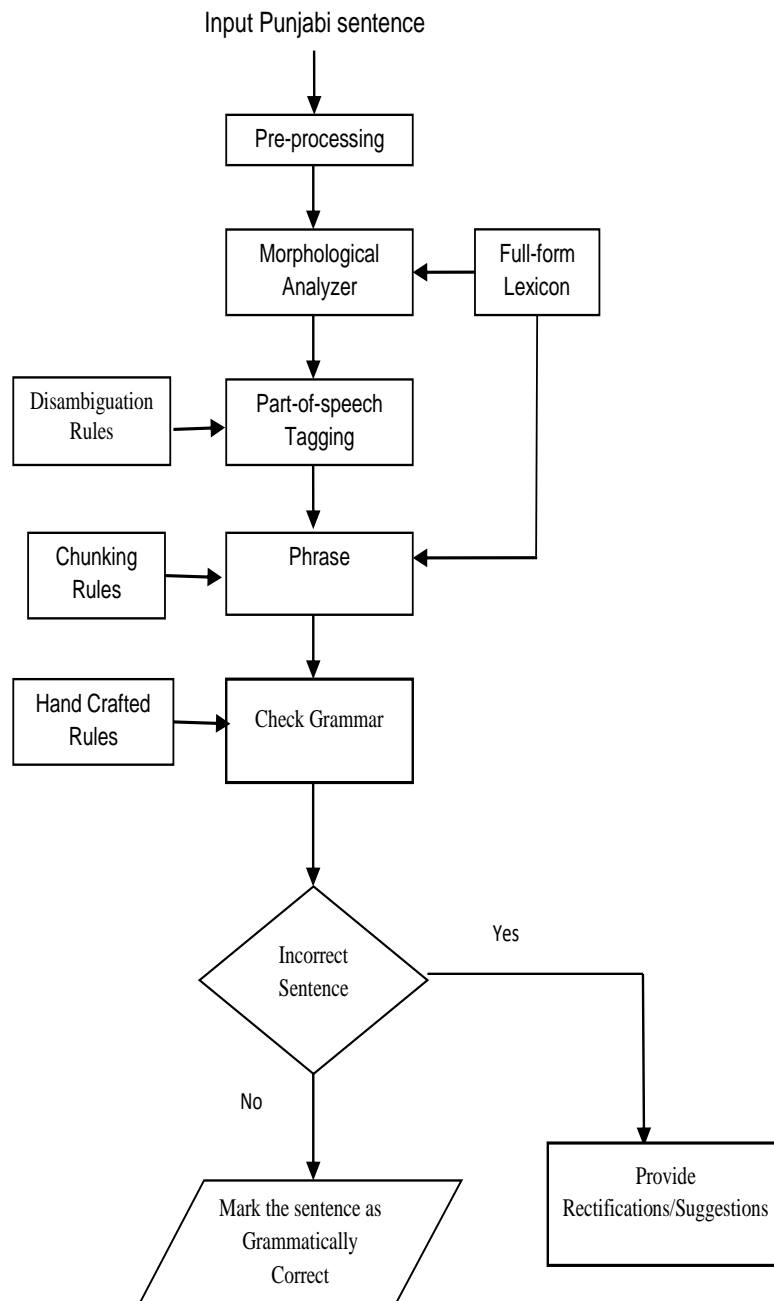


Fig 2. Architecture of Punjabi Grammar checker

3. HINDI GRAMMAR CHECKER

Rule based Morphological process grammar checker for Indian language: Bopche et al described a novel method for Hindi grammar checker. A full-form lexicon for morphology analysis and rule based system are used to form the grammar checker

system. The authors proposed a system, which uses a set of rules and those rules are match against the input Hindi sentences, which are at least POS tagged. As per authors, grammar checker has been successfully implemented for simple sentences and the results are promising. However, the grammar checker only checks those patterns, which have same number of words present in input sentence and this system does not provide any suggestion for the errors. The architectural diagram of the system has been shown in fig 3.

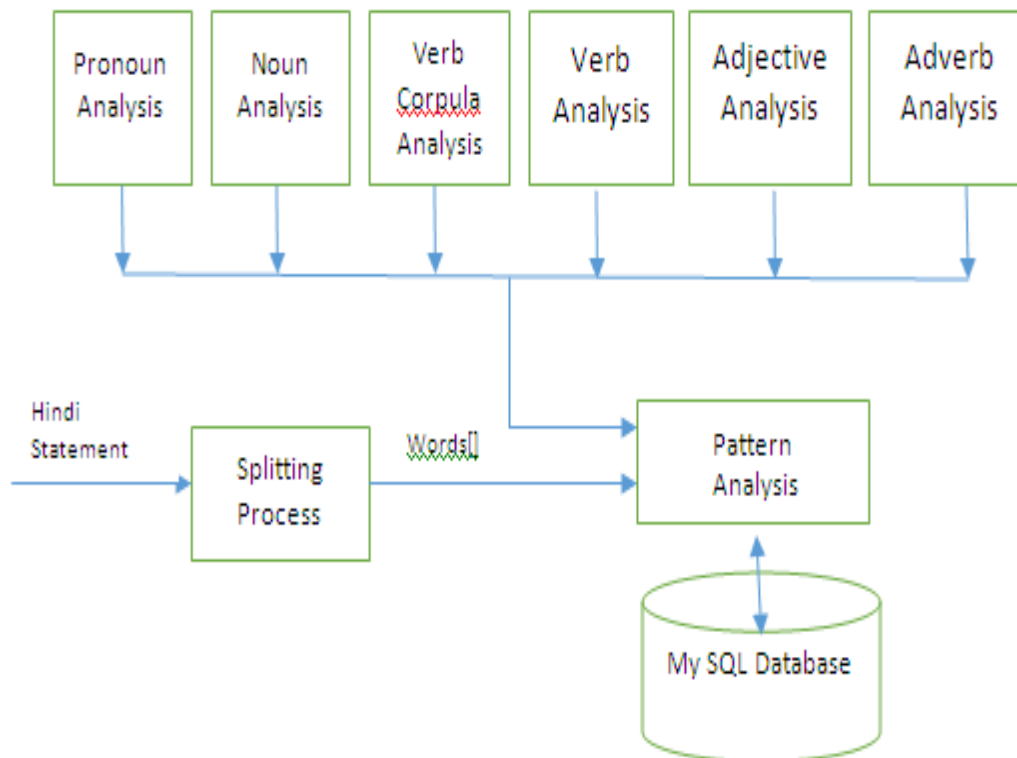


Fig 3. Architecture of Hindi Grammar checker

4. URDU GRAMMAR CHECKER

Kabir et al proposed a model for developing a grammar checker for Urdu. For the sentence analysis, this model uses the proposed two pass parsing approach. This approach was introduced to reduce the redundancy in the phrase structure grammar rules developed for sentence analysis. In the starting to parse the sentence, some base Phrase Structure Grammar Rules are used. Movement Rules are applied and sentence is reparsed in the case of failure. As per the authors, the model was successfully implemented, except for the modules of Morphological Disambiguation and POS Guesser. The model checks the grammatical and structural mistakes in declaration sentences and provides suggestions for the errors[5].

5. BANGLA GRAMMAR CHECKER

Grammar checking System for Bangla and English: Alam et al represented an n-gram based statistical grammar checking system for Bangla and English, which uses the n-gram based analysis of words and POS tags to decide if the sentence is grammatically correct or not. As per the authors, accuracy was very good but there is no such description if suggestions are provided or not for the errors. It only identifies correct sentences from set of sentences given as input. For English, grammar checker performance was 63% and for Bangla it was 53.7%. For both languages, they used manual tagging. For Bangla, they also used automated Bangla POS tagger to check the performance of the grammar checker. They tested this for 34 correct sentences, and the result was 38% correct. The architectural diagram of this system has been shown in figure 4.

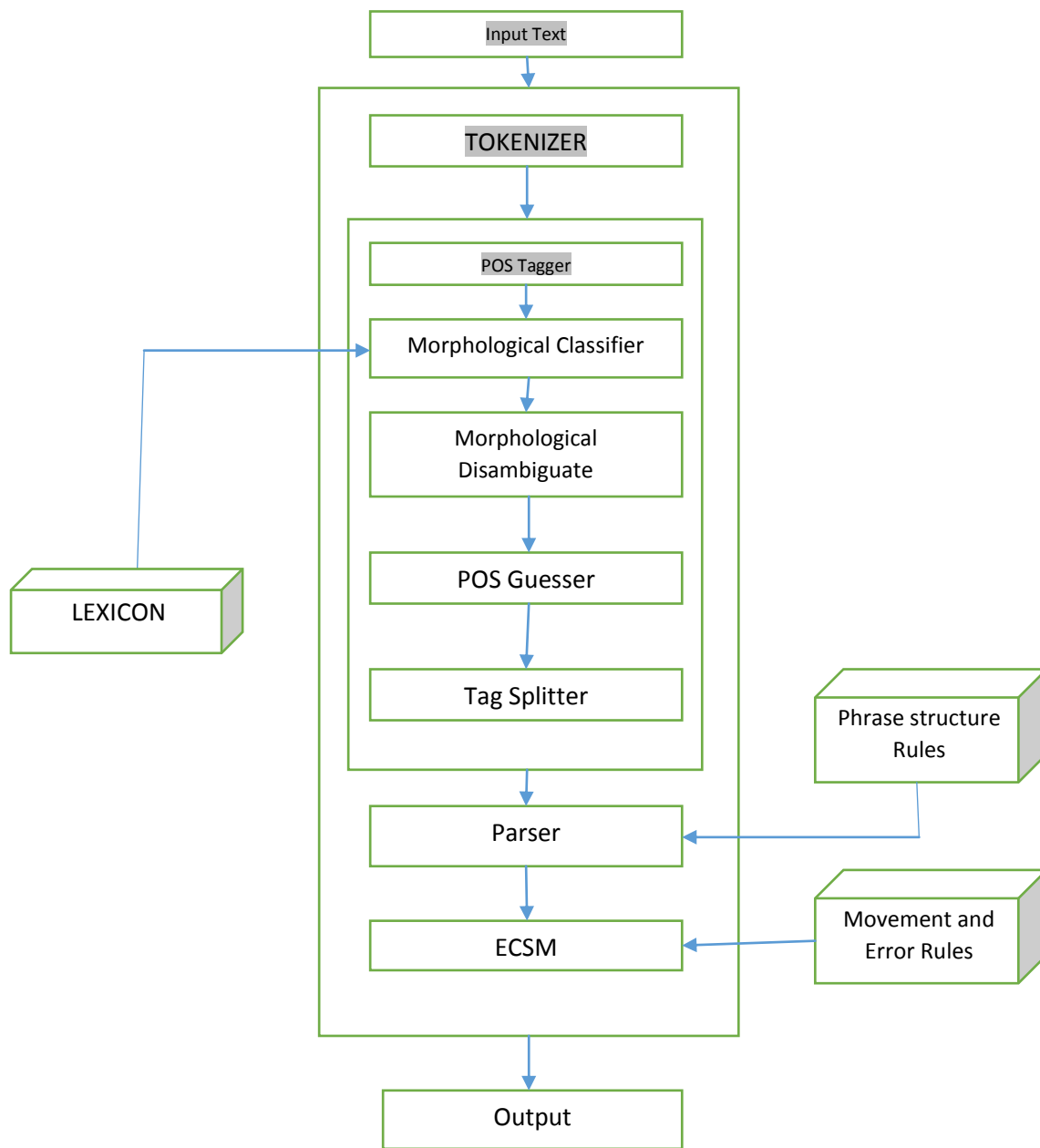


Fig 4. Architecture Diagram of Bangla Grammar Checker

6. CONCLUSION

Different techniques like rule based method and statistical techniques have been used for development of various Asian language grammar checking system. But still a lot of work is to be done for Indian and other Asian languages. For Indian languages only Hindi and Punjabi are only grammar checker developed so far.

REFERENCES

- [1]. Alam, M. J., Uzzaman, N., & Khan, M. (2006). N-gram based Statistical Grammar Checker for Bangla and English. Ninth International Conference on Computer and Information Technology (ICCIT 2006), 3–6.
- [2]. LataBopche, Gauri Dhopavkar, and ManaliKshirsagar, “Grammar Checking System Using Rule Based Morphological Process for an Indian Language”, Global Trends in Information Systems and Software Applications, 4th International Conference, ObCom 2011 Vellore, TN, India, December 9-11, 2011.

- [3]. M. Gill, G. Lehal, and S. Joshi, "A Punjabi Grammar Checker," Acl.Eldoc.Ub.Rug.Nl, pp. 940–944.
- [4]. H. Kabir, S. Nayyer, J. Zaman, and S. Hussain, "Two Pass Parsing Implementation for an Urdu Grammar Checker."
- [5]. J. Kaur, "Hybrid Approach for Spell Checker and Grammar Checker for Punjabi," vol. 4, no. 6, pp. 62–67, 2014.