www.ijecs.in
*International Journal Of Engineering And Computer Science ISSN: 2319-7242*
*Volume 4 Issue 12 Dec 2015, Page No. 15202-15207*

# Syntactic Error Correction System

***A.D.Shah[1], A.D.Shah[2],L.D'mello[3]***
Dwarkadas J. Sanghvi College Of Engineering, Mumbai, India
[1]aasthashahdjsce14@gmail.com
[2]aashanashahdjsce14@gmail.com
[3]lopeslynn@gmail.com

## 1. ABSTRACT

The paper represents the various types of syntactic errors, their detection and further their correction system. This method of detection and correction of the syntactic errors is inspired by the need to correct them for a correct grammatical sentence formation. The best systems, however, results in syntactic errors due to of the lack of linguistic knowledge. An attempt to optimize the task of detecting and correcting the syntactic errors is reflected in this paper.
Keywords: Syntactic errors, Structural, morpho-syntactic, detection, correction, competence, strict, tolerant, FB-LTAG, XTAG

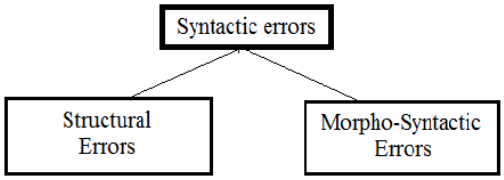
## 2. INTRODUCTION

A very important fraction of the time and efforts for developing a program is dedicated to detect and correct the errors. The types of the frequently occurring errors is studied closely. The method to detect and further correct them is formulated. Syntax errors are frequently occurred and its correction is a part of debugging process. This process is referred as Text data mining as most of the information is stored as text. It is a process of delivering high quality data from the low quality data i.e. the normal text. It is also called as text mining. A typical application is to scan the document and search for errors and correct them. The problem of error detection and correction is in process since long time and various methods for the same has been proposed to get the text into valid collection of words grammatically correct.

In addition to the methods of syntactic error detection and correction, we have first discuss the type of errors in the following sections of this papers. Also, the terms or the grammar used is discussed for the basic understanding before going into the depth.

## 3. SYNTACTIC ERRORS

Basically, the errors that form the basis are syntactic errors and semantic errors. The syntactic errors can
be defined as an error of language resulting from code that does not conform to the syntax of the programming language. This paper includes mainly 2 types of syntactic errors[1]:
Structural errors (Linguistic errors) and Morpho-syntactic errors
The above mentioned two types of errors are explained in detail in the following sections.




### 3.1 Structural errors

The structural errors also known as linguistic errors are the most frequently occurring errors. Linguistic errors signifies the lack of linguistic knowledge or misconception of the languages. The error correction system constructed should respond to these errors and those produce the text without the linguistic errors. The method focuses on the detection and correction of syntactic errors of competence within sentences. Syntactic errors are related to the order and the relationships between the constituents present in a sentence. A sentence whose constituent structure is not according to the rules of the languages are said to possess structural errors[2].
Examples:

He said that he ill was.
(He said that he was ill)
Was raining last night.
(It was raining last night.)

The above examples represent the sentence with the structural errors and the sentences in the brackets represent the correct formation of those statements.

### 3.2 Morpho-syntactic errors

Morpho-syntactic errors refer to the other class of syntactic errors. These errors are basically caused due to the misconception of the syntactic rules. These misconceptions affect the morphology of a word. The errors concerning the tense, agreement or case requirements within a sentence are the examples of the morpho-syntactic errors[3].
Examples:
He go to the school yesterday.
(He went to the school yesterday.)

I did not went to the market.
(I did not go to the market.)

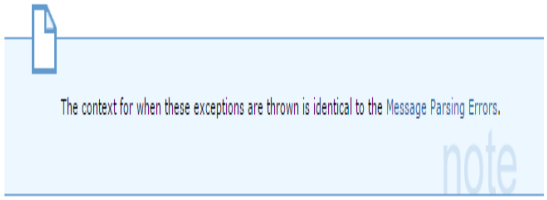
The above two statements are the examples of morpho-syntactic errors with their corrections in the brackets.

## 4. DETECTION AND CORRECTION OF ERRORS OF COMPETENCE

This section basically describes the method for error detection and correction for errors of competence. This method maps the incorrect sentences into correct sentences using grammar[8]. The terms used in this method are strict grammars, tolerant grammars. They are described below in detail:

### 4.1 Strict grammars and tolerant grammars

A strict grammar provides with the correct sentences for the language fragment in the question. The sentences that do not lie in these set of statements are considered to be incorrect. Hence, the correct sentences by the strict grammar are of utmost importance[7].
When strict grammar checking for segment grammars is enabled, exceptions are thrown as noted in the situations detailed in the following diagram.




Note that only the first error in the message is reported.
The tolerant grammars describe the errors that may occur or are likely to occur in the given language fragment. The tolerant grammar thus contains the rules describing incorrect syntactic constructions. Syntactic errors of competence are predictable and recognized easily. Anticipating the syntactic errors using tolerant grammars by writing its codes are useful in detecting and correcting errors. Using tolerant grammars, the correction of syntactic errors are permitted.

### 4.2 Method

The detection and correction of syntactic errors using tolerant and strict grammar proceeds in the following way:

Step 1: The sentence is parsed through strict grammar. If the parsing is successful, the sentence is correct and contains no syntactic errors.

Sentence

↓

Strict Grammar

↓

Structural description

Step 2: If the above parsing isn't successful, it produces error. This sentence is then parsed through tolerant grammar as it contains references to the rules contained in the tolerant grammar

Sentence

↓

Tolerant Grammar

↓

Structural description

Step 3: The structural description thus formed contains structure of incorrect rules hence this is then mapped into the correct rules referring to the strict grammar.

Step 4: This structural description is the parsed through the strict grammar and at the end of parsing, The correct sentence is obtained.

Sentence

↓

Tolerant Grammar

↓

Structural description

↓

Structural description

↓

Strict grammar

↓

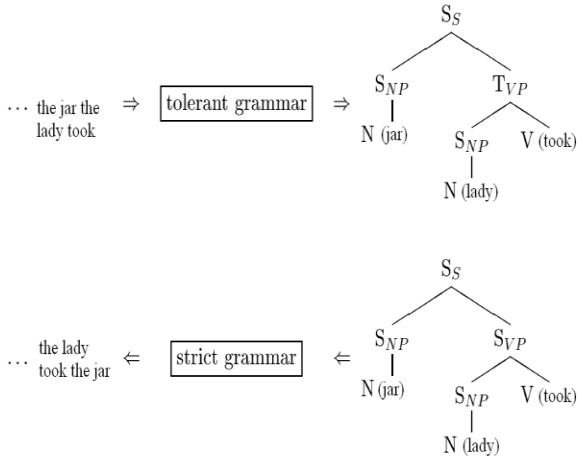
Sentence

Step 5: If when the sentence is parsed through tolerant grammar and it doesn't produce structural description, the sentence cannot be corrected.

DOI: 10.18535/Ijecs/v4i12.20

Sentence → Tolerant grammar → X

Thus, all the sentences that are not correct according to strict grammar cannot necessarily be corrected. But the sentences that are detected by the tolerant grammar will be corrected.
Example:
A sentence with incorrect grammar is selected. This sentence is parsed through strict grammar which gives error. SO, then it is parsed through tolerant grammar and latter it is corrected as shown in the figure.
Consider the statement: The jar the lady took




## 5. DETECTION AND CORRECTION OF MORPHO-SYNTACTIC ERRORS

This section focuses on other type of syntactic errors that is morpho-syntactic errors[8]. It provides the method for detection and correction of this type of syntactic errors. The steps are as follows:

Step 1: The sentence is first passed through the word level. Here, the sentence is first checked for the spellings. The incorrect spellings are corrected.

Step 2: The corrected sentence from the word level is then passed through the sentence level. Here, the other corrections are made and the errors are removed[1].

Sentence

↓

Word level

↓

Sentence level

↓

Sentence

Thus, the correction of morpho-syntactic errors is very important and cannot be neglected.

Example:
Consider He go to the schol everyday

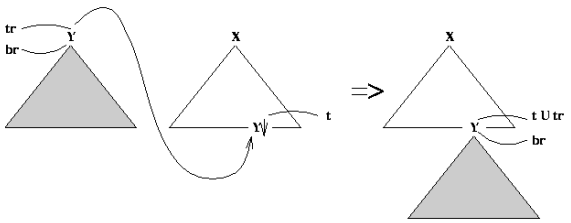
He go to the schol everyday
↓
He go to the school everyday
↓
He goes to the school everyday
↓
He goes to the school everyday

## 6. DETECTION AND CORRECTION OF ALL THE SYNTACTIC ERRORS SIMULTANEOUSLY

The following section describes a general method for detection and correction of the syntactic errors simultaneously. First, the terms used in the method are described in detail. The terms used are Feature-Based Lexicalized Tree Adjoining Grammars (FB-TAG), XTAG, Attribute Value Matrices (AVMs)[4].

### 6.1 FB-LTAG and XTAG

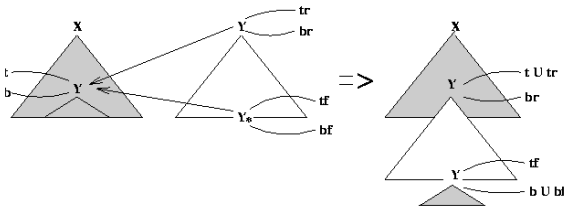
FB-LTAG is basically Feature-Based Lexicalized Tree Adjoining Grammars[2]. The TAG formalism consists of elementary trees. Elementary trees are of 2 types: initial trees and auxiliary trees. Initial trees are those for which the non-terminal nodes are substitutable[5]. Auxiliary trees are those having exactly one non-terminal node is a foot-node that is marked with "*". For adjoining elementary trees, 2 operations are available: substitution and abjunction. FB-LTAG has following characteristics:
1. It is a lexical TAG: This means that elementary tree is associated with at least one lexical item
2. It is Feature-Based Lexicalized TAG: It means that each node in the elementary tree has 2 sets of feature value pairs also known as Attribute Value Matrices (AVMs).




**Substitution of FB-LTAG**

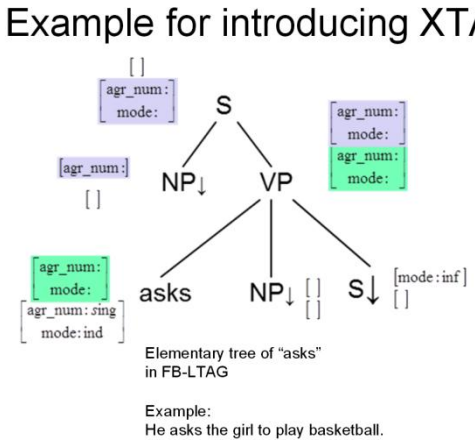
In Fig 1, the feature structure of a new node created by substitution inherits the union of the features of the original nodes. The top feature of the new node is the union of the top features of the two original nodes, while the bottom feature of the new node is simply the bottom feature of the top node of the substituting tree.
In Fig 2, the node being adjoined into splits, and its top feature unifies with the top feature of the root adjoining

node, while its bottom feature unifies with the bottom feature of the foot adjoining node.




**Adjunction using LTAG**

XTAG English grammar is designed using FB-LTAG

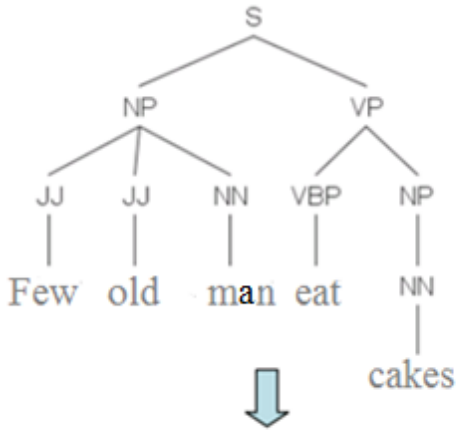


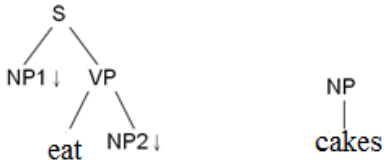

The example in detail is explained later in this paper. The above diagram represents how the XTAG is represented.

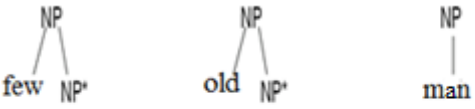
## 6.2 Detection

The steps for detection are as follows:
Step1: Decompose each sentence hypothesis parse tree into elementary trees
Step2: Associate each elementary tree with AVMs
Step3: Reconstruct the original parse tree out of the elementary trees to check if AVMs contradict
Substitution and adjunction operations along with AVM unifications: To simultaneously detect multiple error types and words, a new unification method is proposed[8].
The method is described in detail step by step:

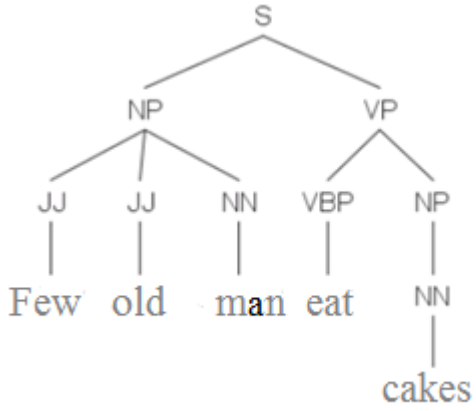
Step 1: Decompose each parse tree into elementary tree



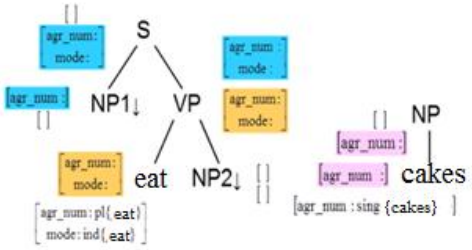



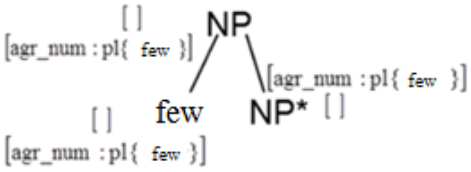



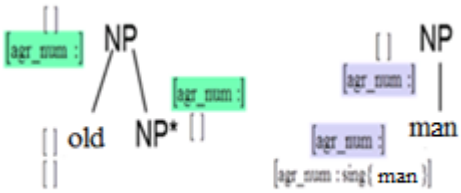

Step 2: Associate each elementary tree with AVMs



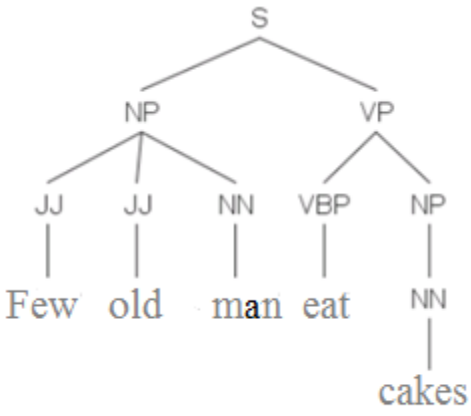
DOI: 10.18535/Ijecs/v4i12.20

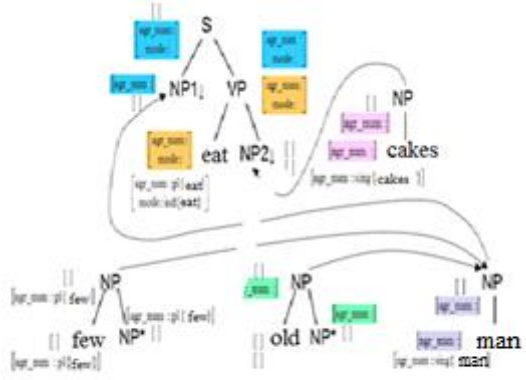


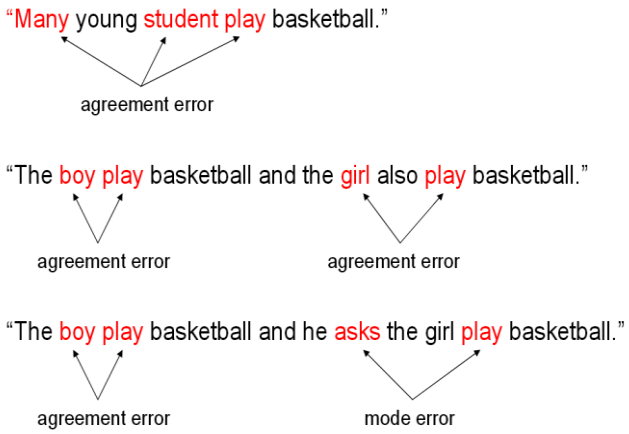




Step3: Reconstruct the original parse tree out of the elementary trees to check if AVMs contradict




After these 4 steps, the errors are detected. The types of errors are illustrated using following example[10]:

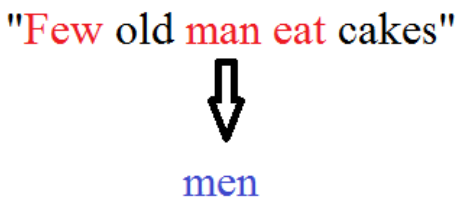



## 6.3 Correction

Once error types and their corresponding words are detected, it is possible to correct errors. It is based on an unified consideration of all related words under the same error types. Given a set of related ungrammatical words, there are two tasks for the correction process: which words should be corrected and how to correct them? This is a simple mechanism available to handle the agreement problem:

The feature value voting: The words whose feature value is in the minority will be selected to be corrected. Take the above example: "student" should be corrected[9]. When we get equal votes, we tend to correct nouns (if nouns are present). Once the corrected words are selected, they can be replaced with their variations but with the same elementary tree type

Example: replacing the above "student" with "students."




## 7. ADVANTAGES OF SYNTACTIC ERROR CORRECTION SYSTEM

Syntactic errors are the most commonly occurring errors. These errors occur in day to day life by the people having less linguistic knowledge and are sometimes done by highly professional too sometimes in haste. The detection and removal of such errors is necessary for proper linguistic understanding. Such systems can also be implemented for other languages where ambiguity is high, like Dutch and Greek[6]. Hence, this syntactic error correction is of utmost importance for inculcating proper linguistic knowledge in people which is the need in today's world.

## 8. CONCLUSION

Thus, we have studied two types of syntactic errors. The methods for detection and correction of each of them, individually is studied. Also, the methods for detection and corrections of combined errors is illustrated with an example.

## References

[1] Automatic Detection and Correction of Syntax Errors in Tutoring Systems for Language Instruction Henk Harkema, harkema@ucla.edu
[2] Detecting and Correcting Syntactic Errors in Machine Translation Using Feature-Based Lexicalized Tree Adjoining Grammars
Wei-Yun Ma∗, and Kathleen McKeown∗,Computational Linguistics and Chinese Language Processing, Vol. 17, No. 4, December 2012, pp. 1-14 *1*, © The Association for Computational Linguistics and Chinese Language Processing
[3] Detecting and Correcting Morpho-syntactic Errors in Real Texts
Theo Vosse*
Nijmegen Institute for Cognition and Information
University of Nijmegen
And Cognitive Technology Foundation
P.O. Box 9104
6500 HE Nijmegen, The Netherlands
e-mail: vosse@nici.kun.nl
[4] Burt, M.K. and C. Kiparsky, The Gooficon, a repair manual for
English Newbury house publishers
[5] Angell, R.C., G.E. Freund, P. Willet. 1983. Automatic spelling correction using a trigram similarity measure. Information Processing and Management (19), pp. 255-261.
[6] Granger, R.H.; Meyers, A.; Yoshii, R.; and Taylor, G. 1983 An Extensible Natural Language Understanding System. Proceedings of the Artificial Intelligence Conference. Oakland University, Rochester, Michigan.
[7] Syntactic Recovery and Spelling Correction of Ill-formed Sentences Kyongho Min and William H. Wilson, School of Computer Science & Engineering, The University of New South Wales
SYDNEY NSW 2052 Australia
Email: {min,billw}@cse.unsw.edu.au
[8] Corpus-Based Syntactic Error Detection Using Syntactic Patterns Koldo Gojenola, Maite Oronoz
Informatika Fakultatea, 649 P. K., Euskal Herriko Unibertsitatea, 20080 Donostia (Euskal Herria)
jipgogak@si.ehu.es, jiboranm@si.ehu.es
[9] The NOMAD System:
Expectation-Based Detection and Correction of Errors during Understanding of Syntactically and Semantically Ill-Formed Text 1
Richard H. Granger Artificial Intelligence Project, Computer Science Department and Cognitive Sciences Program, University of California
Irvine, CA 92717
[10] IEEE explore digital library