



A Rule-Based Grammar Approach to Detect Noun Number Errors

Shahla Mansoori¹, Yogesh Kumar Vishwakarma², Madhvi Soni³, Jitendra Singh Thakur⁴

Department of Computer Science and Engineering, Jabalpur Engineering College, Jabalpur – 482011, India¹⁻⁴

Abstract: English is the world's most dominant language, making automated grammar checking a vital task in natural language processing. Among various grammatical mistakes, noun number errors represent a critical issue that requires accurate detection and correction. This paper presents a rule-based, feature-driven approach designed to detect and resolve noun number errors in English sentences. To evaluate performance, the method was tested on two datasets: a custom dataset of 500 sentences containing noun number errors and the NUS Corpus of Learner English (NUCLE). Additionally, the model's accuracy on the custom dataset was comparative to the widely used Grammarly application. Results: The proposed rule-based model achieved an accuracy of approximately 78% on the custom 500-sentence dataset. On the same custom dataset, the proposed method outperformed Grammarly. But, on the broader NUCLE dataset, the model achieved an accuracy of 38%. The results indicate that while a pure rule-based model effectively addresses noun number error patterns, its performance drops on complex corpus data. Future efficiency can be improved by transitioning to a hybrid model that combines rule-based methods with deep learning techniques, such as bidirectional LSTM dependency parsing.

1. INTRODUCTION

To do Error is the nature of humans, and the same is applicable for languages also. We make mistakes while writing English.[1] This error needs to be corrected. This task is called grammar checking. A Grammar checker is a software or program or idea which is designed to spot grammatical mistakes in the given text with the capability of either automatically correcting the mistakes or suggesting correct options.[1]

In English, uncountable or mass nouns do not have plurals. So noun number error occurs when a plural form of uncountable nouns is used in the text [7]. There are various methods for the detection and correction of noun number error. This paper represents a rule-based method for noun number grammar checker. The accuracy of this method is checked on our own dataset and on NUS Corpus of Learner English (NUCLE)[2]. In this paper, we also compare our method with the Grammarly app which is known as the most used app for grammar checking.

The types of noun number error discussed in the paper –

1. Countable Noun-Singular form - Words that can be counted and used in singular form (Chair, bottle, student)
2. Countable Noun-Plural form - Words that can be counted and used in plural form (Chair, bottle, student)
3. Uncountable Noun Singular form-Nouns used in the singular form as they don't have a plural form (Scenery, poetry, furniture, advice, information)
4. Uncountable Noun - Some nouns which are always singular but look like plural(news)
5. Uncountable Noun - Some nouns have the same form whether singular or plural (Sheep, Species, Score)
6. Uncountable Noun - some collective nouns are always used in the plural with plural verbs. (Poultry, odds, cattle)
7. Uncountable Noun - Noun indicating numerals must be used in singular form(A hundred rupees note.)
8. Uncountable Noun - Uncountable or 'mass' nouns do not go with numbers, and rarely with plural forms or articles. (severe winter)
9. Determiners usage - Some determiners only go with countable nouns (many, few, several, every)
10. Mixed Uses - Lots of nouns have both countable and uncountable forms , (paper)
11. Determiners usage - In a compound noun, if a noun qualified by a quantitative adjective is used as an adjective of another noun then it is kept in singular form since adjectives are never used in plural forms.
12. Determiners usage - If the same noun is to be used after and before a preposition then the noun remains singular. (days)
13. Determiners usage - The expressions are always used as SINGULAR nouns. (A piece of furniture, a bottle of)
14. Uncountable Noun - Abstract nouns are always used in singular form (charity, kindness)
15. Uncountable Noun - Material nouns are always used in singular form (copper, silver)
16. Uncountable Noun - Some collective nouns are always used in singular with singular verbs. (committee)
17. Determiners usage - A plural noun or a plural pronoun is used after these phrases (boys).



2. RELATED WORK

2.1 Frequency-based Spell Checking and Rule-based Grammar Checking

This paper used two techniques, for grammar checking and prediction (1) Frequency Based Suggestion Prediction (2) Grammar Checker.

In Frequency Based Suggestion Prediction, for each word, in the dictionary, they maintain the usage frequency. For the words, which are within a threshold edit distance of the given incorrect word, they look at the frequency of each suggested word. The word with a higher frequency is presented higher in the suggestion list.

And for grammar checker on the basis of tense rules, certain rules are formed which can be understood by the system.[1]

2.2 Grammar Checker by MitaliBhurani

MitaliBhurani made a grammar checker using a rule-based approach. This approach includes (1) Sentence Boundary Detection: Here she used nltk's sent_tokenize to split the text into sentences. (2) Part-of-speech Tagging: Here she used nltk's pos_tag to assign pos tags to each word of the sentence. (3) Spell Check: Here she used pyenchant. (4) Indefinite Article and following word agreement: She used inflect which also considers exceptions like "a European" is correct and not "an European" and hence is perfect for this purpose.[3]

2.3 Efficient Rule-Based Grammar Checker with Word Sequencing.

In this thesis, they present a rule-based grammar checker, which is capable of correcting the word order of a sentence in the English language. Their tool first identifies the tense of the input sentence. For every tense, they have different categories of sentences. Then based on the tense, it further categorizes the sentence into the category it belongs to, and depending on the category of the sentence, the corresponding function is called which corrects the sentence.[4] They used their method on 36 different sentences which involve different types of error. They compare their result with the 9 grammar checking tools.

2.4 A Hybrid Approach to Grammatical Error Correction

In this paper, they describe a grammar correction system for the CoNLL-2013 shared task. Their system corrects three of the five error types specified for the shared task - noun-number, determiner and subject-verb agreement errors. For noun-number and determiner correction, they apply a classification approach using rich lexical and syntactic features. This system obtained an F-score of 11.03 on the official test set using the M2 evaluation method (the official evaluation method).[5] For noun number error this system accuracy is 14.66%.

2.5 LFG-based Features for Noun Number and Article Grammatical Errors

This paper introduces a participating system of the CoNLL-2013 Shared Task "Grammatical Error Correction". This focused on the noun number and article error categories and constructed a supervised learning system for solving these tasks. This system carried out feature engineering and they found that (among others) the f-structure of an LFG parser can provide very informative features for the machine learning system.[6] For noun number error this system accuracy is 18.03%.

3. OUR PROPOSED METHOD

Noun Number error deals with correction of nouns depending on numbers, the wrong number of nouns (Nn) meant that either a singular noun should occur in the plural form or a plural noun should occur in the singular form or sometimes left unchanged in some exceptional cases.

The major factors which determine the number of nouns are-

- The intended meaning of the text.
- Types of the noun that are used in the sentence.
- Singular or Plural sense of the sentence.

For Noun number error detection and correction, we are using a rule-based technique using Natural language Processing in the domain.



3.1 Model

1. Noun number error detector using a feature-based technique
2. Error Correction by determining the types of noun and their correction.

3.2 Tools used for development

1. NLTK(POS tag, Lemmatizer).
2. INFLECT library for noun pluralization.
3. Wiktionary list of mass nouns, pluralia tantum.
4. List of Uncountable Nouns.
5. Jupyter Notebook.

3.3 Types of nouns

Types of nouns	Examples
Common Nouns	<i>house, cat, girl, foot, country</i>
Proper Nouns	<i>Spain, Fido, Sony</i>
Singular Nouns	<i>house, cat, girl, foot, country</i>
Plural Nouns	<i>houses, cats, girls, countries</i>
Collective Nouns	<i>crowd, flock, committee, a hundred dollars</i>
Countable Nouns	<i>table, apple, rabbit, ear</i>
Uncountable Nouns	<i>salt, seafood, luggage, advice</i>
Abstract Nouns	<i>love, creativity, democracy</i>
Compound Nouns	<i>dry-cleaning, toothpaste, haircut, output</i>
Concrete Nouns	<i>table, apple, rabbit, ear</i>

3.4 Rules for determining types of noun and their corrections

1. Collective nouns such as – Jury, choir, committee, council, crowd, herd, orchestra, team, government, mob, community, union, club, opposition, firm, flock etc, to denote a group. They are considered to be singular and a singular verb is used with them.

Example - The committee has submitted its report

2. The unit of measurement (such as - hour, pound, kilo, mile...etc.) is always used in the singular form in the structure – ‘Half + a/an + unit of measurement’; as, ‘Half a kilo’, ‘Half an hour’.

Example - Only Half an hour left to finish this work.

3. The structure – ‘Numeral Adjectives + plural noun + and + a half’ or Numeral Adjectives + and + a half + plural noun is used in the English Language. Numeral Adjectives: One, two, three, four etc. some, all, many, few etc. are called Numeral Adjectives.

Example- ‘Two kilos and a half’ ‘Five hours and a half’ ‘Two and a half kilos’

4. Generally, the plural of a proper noun is not possible. But the plural of a proper noun can be formed (=made) by adding ‘s’ according to need.

Example - There are two Mohans in my class.

5. These nouns such as – barracks, corps, crossroads, Innings, headquarters, précis, series, species, Issue, offspring, aircraft, craft, swine are used in the same form both in singular and plural. **Example**- All the police barrack of Gorakhpur are old.

6. The structure – ‘Noun + preposition + same noun’ is always used in the singular. A singular noun is always used before preposition and after a preposition; as’

Example - Village after village has been swept away.



7. A plural noun or a plural pronoun is used after these phrases – one of, each of, either of, neither of, any one of, a few of, very few of, half of, a lot of, a large number of, etc.
Example- One of the boys was innocent.
8. If we add 's' or 'es' to some Adjectives, they become plural nouns; as' **Example-** We have to taste the sweets and bitters of our lives.
9. Some nouns look plural in form but have a singular meaning. Such nouns take a singular verb. These are : news, innings, politics, summons, physics, economics, ethics. mechanics, mathematics, measles, mumps, rickets, billiards, draughts, etc.
10. Some nouns look singular but have a plural meaning. Such nouns take plural verbs. These are : cattle, clergy, cavalry, infantry, poultry, peasantry, children, gentry, police, etc.
11. Some nouns are always used in singular. These are uncountable nouns. We should not use article A/An with such nouns. These are : Scenery, poetry, furniture, advice, information, hair, language. business, mischief, bread, stationery, crockery, luggage, baggage, postage, knowledge, wastage, money, jewellery, breakage, etc. We can not pluralize such nouns by adding 's' or 'es'.
Example- I need three hairs of a black horse.
12. Some nouns have a plural meaning. If a definite numeral adjective is used before them they are not pluralized. e.g., pair, score. Gross, stone, hundred, dozen, thousand, million, billion, etc. Otherwise, these nouns can well be pluralized: Dozens of women, Hundreds of people, Millions of dollars, Scores of shops. Many pairs of shoes, thousands of millions, etc.
13. If a numeral adjective and a fraction are used with a noun, the noun is used with the numeral and the noun will be in the singular.
Example- He gave me one rupee and a half.
As we have seen there are so many rules to consider for noun number error detection and correction and thus we come up with a feature set to resolve this problem step by step.

3.5 Steps

As shown in the flowchart (Figure 1), following are the steps to detect Noun Number errors in a sentence.

Step 1- The given sentence is broken into Tokens, POS and chunk tags using NLTK POS tagger and lemmatizer.

Step 2- Check for determiners class(label : SingularDT, PLuralDT) i.e. determiners used before singular nouns and plural nouns.

Step 3- Check for the presence of any Cardinal Adjectives.

Step 4- For noun (plural to singular and singular to plural), verb (gerund, present & past) and adjective (comparative, superlative) transformations we used INFLECT python library.

Step 5- For special cases – check for uncountable nouns, mass nouns, Pluralia Tantum, named entity, and other cases which are to be checked using specialcheck function.

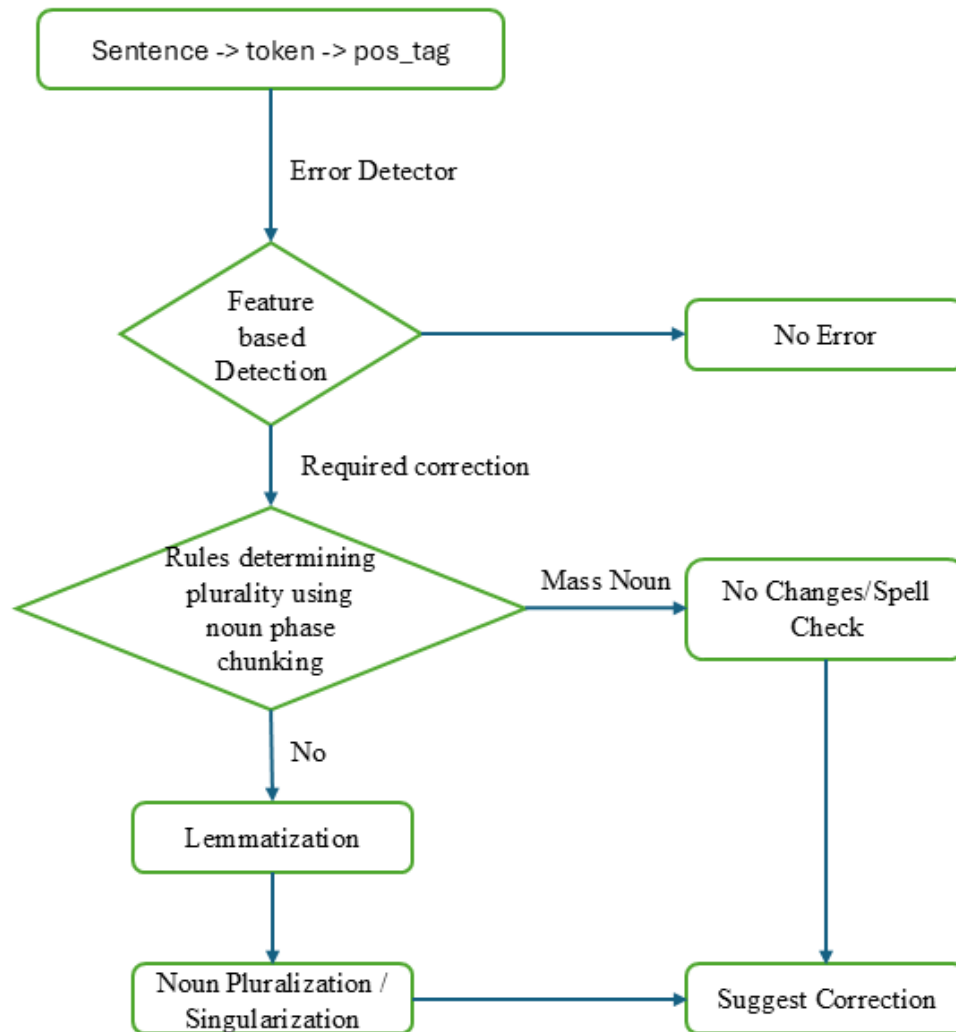


Figure 1

3.6 Dataset Collection –

- 1) **Own Dataset** – We created a dataset of 500 sentences of noun number error. This dataset contains 3 columns correct sentences, incorrect sentences and correct word. This dataset contains all the types of noun number error discussed which were discussed in section 1.

Table 1: Sample dataset of four sentences

Correct sentences	Incorrect sentences	Correct Word
These people are poor because they have been too lazy to work.	These people are poors because they have been too lazy to work.	Poor
He takes care of sixty little blind girls.	He takes care of sixty little blinds girls.	Blind
I took your advice , and I got myself engaged.	I took your advices , and I got myself engaged.	advice
She had a ten dollars bill.	She had a ten dollar bill.	dollars



2) NUS Corpus of Learner English (NUCLE) –

The NUS Corpus of Learner English (NUCLE) was collected in a collaboration project between the National University of Singapore (NUS) Natural Language Processing (NLP) Group led by Prof. Hwee Tou Ng and the NUS Centre for English Language Communication (CELC) led by Prof. Siew Mei Wu. The work was carried out as part of the PhD thesis research of Daniel Dahlmeier at the NUS NLP Group.[2]

This dataset contains a total of 57151 sentences with all types of grammar errors and it has 3140 sentences consisting of only noun number error.

We extract the noun number error sentences from the dataset by using a python script[10]. The python script corrects all types of error in the dataset except noun number error and created a new dataset that contains only noun number error sentences.

4. RESULTS

In every research there is data to study on, so we used 2 datasets that contain noun number error sentences and used them to check the accuracy of our model.

4.1 Result on our own dataset

According to our model and our own dataset, we got an accuracy of around **77.9%** on our own dataset. The goal of our research model to detect noun number error was fulfilled by **389** sentences out of 500 sentences taken.

4.2 Result on NUS Corpus

According to our model and dataset extract from NUS Corpus of Learner English (NUCLE) [2], we got an accuracy of around **38%**. The goal of our research model to detect noun number error was fulfilled by **1196** sentences out of **3140** sentences taken.

Table 2: Method Accuracy

	Total Number of Sentences	Total number of sentences detected and suggested for error	Accuracy (%)
Our Own dataset	500	389	77.9
NUS Corpus	3140	1196	38.0

4.3 Comparison with the Grammarly

We compare the accuracy that we got when we perform our model on our own dataset with the most used app for grammar checking, Grammarly app that performs the same task for the same dataset. We used the free version of the Grammarly app and we test our model on the web interface of the Grammarly app.

Table 3: Comparison with Grammarly

	Total Number of Sentences	Total number of sentences detected and suggested for error	Accuracy (%)
Our rule-based Method	500	389	77.9
Grammarly app	500	184	36.8

5. DISCUSSION AND ANALYSIS

The method proposed by us concludes with 13 rules. Using these 13 rules our method cannot detect and suggest the right word for 111 sentences out of 500 sentences of our own dataset. Hence it is clear that there is a lot of scope to improve our method. Table 4 enlists the issues for which our method cannot detect and suggest a right word.



When we performed our method on NUS Corpus the Rule-based model failed in long sentences where noun depends on words which are not covered under grammar parsing applied in our model. Applying a special check to pluralize those nouns by keeping a dictionary of words and pluralizing it causes extra/unwanted suggestions.

Table 4: List of issues with our Method.

Example	Correct word	Issues
She had a ten dollar bill.	Dollars	it can be written as ten-dollar or ten dollars, both are correct but in different situations
Tom put two pieces of breads in the toaster.	Bread	Bread can be countable and uncountable
You should never give out your passwords.	Password	The intended meaning of this sentence doesn't clarify the plural or singular nature of the sentence
He maneuvered the teams around a big stump.	Team	The intended meaning of this sentence doesn't clarify the plural or singular nature of the sentence
You cannot master physics without learning the formula that govern it.	Formulae	As this speaks about all the formulae of physics but this type of implications are not possible in the rule-based model
What kind of perfumes are you wearing, it's Wonderful	Perfume	Our model is unable to distinguish the frame 'kind of perfumes' from 'kind of perfume' as its singular nature is described far later.
I didn't get past the perfumes counter.	Perfume	The intended meaning of this sentence doesn't clarify the plural or singular nature of the sentence
We like the large bottle of mineral water.	Bottles	Lack of information
We lost our way in the wood.	Woods	Change in meaning with 's' as a suffix. As here woods will refer to a forest.
He instructed me to write the telegram in capital letter.	Letters	The intended meaning of this sentence doesn't clarify the plural or singular nature of the sentence
A series of lecture is scheduled.	Lectures	Conflicting rules to distinguish the use of 'a series of lectures' vs 'series of lectures'
To do so would release the Original Being, and then things would really be bad.	Beings	Pronouns that are also nouns create confusion.
Sheila must know time's worth.	worth of time	Model is not trained for phrasal substitution for such errors
Firstly, I believe most of us will agree with the sentences below.	Sentence	Here it can be singular or plural but cannot be very sure without further details.
With the risk having genetic disorder.	Disorders	Lack of information can cause such type of errors.
First of all, from the angle of human right, a genetic risk carrier should not be obligated to tell the truth to others.	Rights	Use of suffix 's' may sometimes cause a change in meaning
That is a skid mark made by the right wheel or the vehicle's tyre.	Tyre of the vehicle	Model is not trained for phrasal substitution for such errors.
They had many chances to scores but they couldn't convert their opportunities.	Score	The intended meaning of the sentence does not clarify the singular or plural nature of the sentence.



6. CONCLUSION

Noun Number error is one of the grammatical errors which may occur in any English sentence. In our research, we had prepared a model with a feature-based approach to resolve noun number error from any sentence.

Our model provides an accuracy of around 78% on our own dataset but on NUS Corpus it provides only 38% accuracy. To make a more robust model which will overcome all the drawbacks of this model the Rule-based model can be converted to a Hybrid model (i.e. Rule-based + Deep learning model). Using bidirectional LSTM Dependency parsing can be achieved. Thus, there is a lot of scope to increase the efficiency of our method in the future.

REFERENCES

- [1] Shashi Pal Singh, Ajai Kumar, Lenali Singh, Mahesh Bhargava, Kritika Goyal, Bhanu Sharma Frequency based Spell Checking and Rule based Grammar Checking International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016
- [2] NUS Corpus of Learner English (NUCLE) <https://www.comp.nus.edu.sg/~nlp/corpora.html>
- [3] MitaliBhurani, Grammar Checker, 28 May 2018, <https://github.com/MitaliBhurani/Grammar-Checker>
- [4] Krishna Kumar Giri, Efficient rule-based Grammar Checking with words sequencing, Computer Science and Engineering Department, Thapar University, Patiala July 2017.
- [5] Anoop Kunchukuttan, Ritesh Shah, Pushpak Bhattacharyya, IITB System for CoNLL 2013 Shared Task: A Hybrid Approach to Grammatical Error Correction, Department of Computer Science and Engineering, IIT Bombay.
- [6] Gabor Berend, Veronika Vincze, Sina Zarriess, Richard Farkas, LFG-based Features for Noun Number and Article Grammatical Errors, Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task.
- [7] Soni, Madhvi, and Jitendra Singh Thakur. "A systematic review of automated grammar checking in English language." *arXiv preprint arXiv:1804.00540* (2018).
- [8] Sanjeev Kumar Sharma, Rule Based Grammar Checking Systems (A Survey) International Journal of Computer Applications & Information Technology Vol. 10, Issue 1, (ISSN: 2278-7720),
- [9] Write Your Best with Grammarly. Grammarly, available at www.grammarly.com/