Identification And Classification Of Reduplication Words In Punjabi Language

Pertik Garg, Anu Marwaha, Manju Bala Goel

Abstract: Identification of reduplication words is a Natural Language Processing task that extracts reduplicative words from various text forms and classifies them according to full, partial and discontinuous type. Over the years, magnificent growth could be observed in the use of regional languages on the web in the terms of news, opinions, tweets, hash tags, reviews, articles and blogs etc. Identification and classification of reduplication words task are very challenging in computational linguistic point of view, especially if the text is written in regional languages. The availability of linguistic resources for Punjabi language is not available such as automatic tools for tokenization, feature selection, stemming and tagging etc. In this paper, we have designed an algorithm and develops Graphical User Interface which accepts input as a Punjabi text and gives output by highlighting reduplicative words and also classified the types of identifying reduplicative words. Corpus based and Rule based approaches are used for implementation of the algorithm and experimental results are evaluated from the implementation.

Keyword: Corpus based approach, NLP, Reduplication, Rule based approach.

1 INTRODUCTION

Natural Language processing (NLP) is a subject of artificial intelligence dealing with mathematical and computational algorithms to automate various aspects of human (natural) language and development of various systems. In other words, Natural Language Processing is a method where the machine can turn out to be progressive human and in this way lessening the separation between the individual and the machine [1]. The study of NLP has been around for more than 50 years and commence in the research with the rise of computers. NLP is based on two main techniques Syntax Analysis and Semantic Analysis. Syntax Analysis or Parsing is the second phase of compiler that comes after lexical analysis. It checks the syntactical structure of the entered input, i.e. whether the entered input is under the rules of predefined grammar or not. Semantic Analysis is the third phase of Compiler that comes after a lexical and semantic analysis. Semantics analysis interprets symbols, their types, and their relations with each other. It confirms that declarations and statements of program are semantically correct. The main task of semantic analysis to judge whether the syntax structure defined in the source program explain any meaning or not. Identification of reduplicated words has an important role in many natural language processing (NLP) applications, such as text summarization, machine translation (MT), identification of multiword expressions, etc. When we want to work on the application related to speech to text recognition, the algorithms for identification of reduplication words must be used. This paper focuses on an algorithm for finding the reduplicated words from the entered text in Punjabi language and identifying the types of reduplication.

1.1 Punjabi Language

Punjabi, often spelled Panjabi, belongs to the Indic gathering of the Indo-European group of Languages [2]. It is spoken in Punjab, neighboring states of Haryana and Himachal Pradesh. In India it is the official Language of Punjab state. In addition, about 25 percent of the public living in the New Delhi metropolitan area speaks Punjabi in daily life. The Punjabi language is almost related to the Sikh religion. Its alphabet, recognized as Gurmukhi, was the vehicle for recording the teachings of the Sikh gurus. It was introduced by the second of the gurus in the 16th century. The word “Gurmukhi” explains “Guru’s mouth.” There is no conception of upper or lower case letter in the Punjabi (Gurmukhi) language. The Gurmukhi characters, unlike the Greek and Roman alphabets, are set in a logical way: vowels first, then the consonants and semi-vowels.

1.2 Reduplication & Its Types

Reduplication means when a word repeat itself completely same or with minor change. It is related with phonological as well as with morphological patterns. In general, reduplication can be seen as a morphological procedure that includes the repetition of a morpheme or whole word, which sometimes results in phonological changes in the new word [3]. Reduplication arises from the Latin word “reduplicate” whose meaning is doubling. The repeated word can be fixed at anywhere either at the start, middle or at the end. Reduplication has a different role in different languages. Reduplication can occur both in spoken words and written text. Lots of reduplication words are used in communication. We can see reduplication in baby talk like “oti” for “Roti” (Food). We can handle reduplication only when we know the syntax and semantics of a word. There are various types of reduplication exists in Punjabi Language. Fig. 1 shows all types of reduplication.

We can define the types of reduplication with the example as:-

Copyright

Form

- *Pertik Garg, Associate Professor, Department of Computer Science & Engineering,SVIET, Ramnagar, Punjab, India*
- *Anu Marwaha, Assistant Professor,Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India*
- *Manju Bala Goel, Department of Electronics & Communication Engineering, PTU, Jalandhar, Punjab, India*

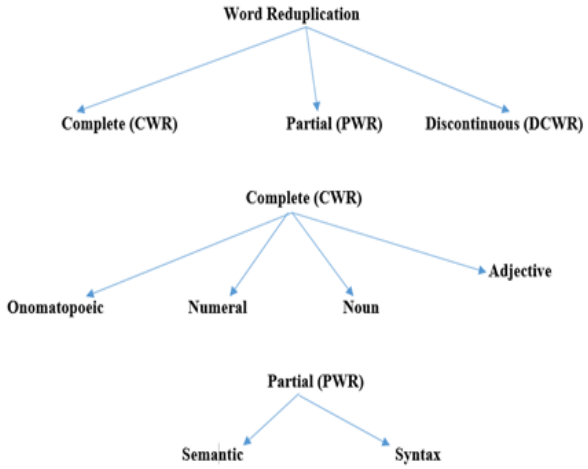

***Fig. 1**. Types of Reduplication*

1.2.1 Complete reduplication

Complete/Full reduplication makes the repetition of a sound or word. When the same word repeats itself the reduplication is complete reduplication. In complete reduplication both the words are important for making the sense or meaning of the sentence as shown in Table 1.

TABLE 1
COMPLETE REDUPLICATION

S.No.	Reduplicate Word	Transliteration of word - English meaning	Word used in the Sentence
1	ਕਿੱਥੇ ਕਿੱਥੇ	kithe-kithe (where)	ਤੁਸੀਂ ਕਿੱਥੇ ਕਿੱਥੇ ਜਾ ਕੇ ਆਏ ਹੋ?
2	ਆਪਣੇ ਆਪਣੇ	apne-apne (your own)	ਆਪਣੇ ਆਪਣੇ ਕਮਰੇ ਵਿੱਚ ਜਾਓ

1.2.2 Partial reduplication

Partial reduplication involves a repetition of word with minor change. The repeated word either from start, end or in the middle of the word. In linguistic when stem of a word repeated itself with slight change in the morphological process (as shown in Table 2), it is called partial reduplication.

TABLE 2
PARTIAL REDUPLICATION

S.No.	Reduplicate Word	Transliteration of word –English meaning	Word used in the Sentence
1	ਚਾਹ ਚੋਹ	chah choh (tea)	ਚਾਹ ਚੋਹ ਪੀ ਕੇ ਜਾਓ
2	ਜਾਣਾ ਜੋਣਾ	Jana jona (to go)	ਜਾਣਾ ਜੋਣਾ ਕਿੱਥੇ ਹੈ?

1.2.3 Semantic reduplication

When two repeated words are not same, but the meaning of both the words are same is called semantic reduplication. Semantic reduplication may belong to partial reduplication because the second part of the word is slightly different or completely different from the first part of the word, but semantically both the parts define same meaning as shown in Table 3.

TABLE 3
SEMANTIC REDUPLICATION

S.No.	Reduplicate Word	Transliteration of word –English meaning	Word used in the Sentence
1	ਗੱਡੀਆਂ ਕਾਰਾਂ	gadian caran (cars)	ਸਭ ਗੱਡੀਆਂ ਕਾਰਾਂ ਬੰਦ ਨੇ
2	ਖਾਣਾ ਪੀਣਾ	khana pina (Food)	ਖਾਣਾ ਪੀਣਾ ਨਾਲ ਲੈ ਕੇ ਜਾਓ

1.2.4 Syntax reduplication

When the second part of the word repeated with the difference of space or hyphen (-) from the first part of the word it is called syntax reduplication as shown in Table 4. In this type our main focus on syntax. Correct syntax includes tense, word choice and placing words and phrases in the right order.

TABLE 4
SYNTAX REDUPLICATION

S.No.	Reduplicate Word	Transliteration of word –English meaning	Word used in the Sentence
1	ਆਣਾ-ਜਾਣਾ	ana-jana (movement)	ਆਣਾ-ਜਾਣਾ ਲੱਗਿਆ ਰਹਿੰਦਾ ਹੈ
2	ਇੱਧਰ ਉੱਧਰ	idhar udhar (Move over here)	ਇੱਧਰ ਉੱਧਰ ਨਾ ਦੇਖੋ

1.2.5 Onomatopoeic reduplication

Onomatopoeic reduplication focused on sound. When two words have same sound the words refer to as reduplicated words. The word onomatopoeia generates from the combination of two Greek words, onoma meaning "name" and poiein meaning "to make," so onomatopoeia refers "to make a name (or sound)." Example of onomatopoeic reduplication is shown in Table 5.

TABLE 5
ONOMATOPOEIC REDUPLICATION

S. No.	Reduplicate Word	Transliteration of word–English meaning	Word used in the Sentence
1	ਦਗੜ ਦਗੜ	dagd dagdh (noise for running of horse)	ਘੋੜਾ ਦਗੜ ਦਗੜ ਦੌੜਦਾ ਹੈ
2	ਟਿਪ ਟਿਪ	tip tip (noise of droping water)	ਪਾਣੀ ਦੀਆਂ ਬੂੰਦਾਂ ਟਿਪ ਟਿਪ ਕਰ ਰਹੀਆਂ ਹਨ

1.2.6 Numeral reduplication

When two words refer to same numeral, the combination of words is called numeral reduplication (as shown in Table 6). “Numeral” is a synonym of “Number” and assigns all numbers to the part of speech called Numerals. Mainly numeral is used in the language as noun like ਇਕ ਬਹੁਤ ਛੋਟਾ ਨੰਬਰ ਹੈ.

1.2.7 Noun reduplication

A word (other than a pronoun) used to recognize type of people, places, or things (common noun), or to name a particular one of these (type of Proper noun) as shown in Table 7. When the contiguous repeated words are noun, the combination of both words is called Noun Reduplication.

TABLE 6
NUMERAL REDUPLICATION

S.No.	Reduplicate Word	Transliteration of word –English meaning	Word used in the Sentence
1	ਤਿਨ ਤਿਨ	tin tin (three three)	ਤਿਨ ਤਿਨ ਦੀ ਟੁਕੜੀ ਵਿੱਚ ਹਮਲਾ ਕਰੋ
2	ਦੱਸ ਦੱਸ	das das (ten ten)	ਸਾਰੇ ਨੋਟ ਦੱਸ ਦੱਸ ਰੋਪਏ ਦੇ ਹਨ

TABLE 7
NOUN REDUPLICATION

An IJSTR copyright form must accompany your final submission. You can get a

S.No.	Reduplicate Word	Transliteration of word –English meaning	Word used in the Sentence
1	ਘਰ ਘਰ	ghar ghar (house)	ਘਰ ਘਰ ਪਰਚੇ ਵੰਡੇ ਗਏ
2	ਦਿਨਾਂ ਦਿਨਾਂ	dinan dinan (days)	ਉਹ ਦਿਨਾਂ ਦਿਨਾਂ ਵਿੱਚ ਕਾਮ ਸਿੱਖ ਗਿਆ

1.2.8 Adjective reduplication

A word that highlights the properties of noun or pronoun is called adjective. The work of an adjective is to modify a noun or pronoun. When the contiguous repeated words are adjective the combination of both words is called Adjective Reduplication as shown in Table 8.

TABLE 8
ADJECTIVE REDUPLICATION

S.No.	Reduplicate Word	Transliteration of word –English meaning	Word used in the Sentence
1	ਠੰਡੀ ਠੰਡੀ	thandi thandi (coldness)	ਠੰਡੀ ਠੰਡੀ ਹਵਾ ਚੱਲ ਰਹੀ ਹੈ
2	ਕਾਲੇ ਕਾਲੇ	kale kale (black)	ਕਾਲੇ ਕਾਲੇ ਬੂਤ ਡਰਾਉਣੇ ਸਨ

1.2.9 Discontinue Word Reduplication

When two contiguous repeated words are differentiated by any single different word then the combination of the situation is referred to as discontinue word reduplication as shown in Table 9.

TABLE 9
DISCONTINUE WORD REDUPLICATION

S. No.	Reduplicate Word	Transliteration of word –English meaning	Word used in the Sentence
1	ਸਾਰੇ ਦੇ ਸਾਰੇ	sare de sare (all)	ਉਥੇ ਸਾਰੇ ਦੇ ਸਾਰੇ ਜਾਵਾਂਗੇ
2	ਗੁਲਾਬ ਹੀ ਗੁਲਾਬ	gulab hi gulab (rose)	ਗੁਲਾਬ ਹੀ ਗੁਲਾਬ ਦੇ ਫੁੱਲ ਪਏ ਸਨ

2 LITERATURE REVIEW

There are some study material found related to my research work. These are given below:- Gupta and Sharma [4] describes that the most common procedure of information exchange is Speech in this paper. There are two processes involved in the speech interfacing, one is speech synthesis and other is speech recognition. Speech recognition is the process through which computer recognizes the words that a person speak on the telephone or a microphone. The two main methods involved in the speech recognition are signal processing and pattern matching. Signal processing method is used in front-end and pattern matching at the backend. In this paper, they make a setup of speech recognition of partially reduplicated words spoken in Hindi language. Mel Frequency Cepstral Coefficient (MFCC) is used in front-end and artificial neural network is used at backend to evaluate the experimental result. In this paper, authors experimentally find highest average recognition rate through feature extraction methods such as MFCC. They have also proposed that the recognition rate may be further increased that is beyond 98.2% be using another feature extraction techniques. Khan [5] discussed the reduplication and its types in Arabic and Urdu language. The focus of the paper is on the spoken form and reduplication in literature of two languages. This paper explained the three types of reduplication as complete, partial and discontinuous word reduplication. In this paper, the author made a comparison between two languages on the level of reduplication. After review and research of reduplication, he concluded that both languages are common in literature as well as at spoken level. He found that in full reduplication, copied part is a better prosodic constituent as compared to the kind of syllable. He also found that in partial reduplication, the reduplicate is the best prosodic constituent. Dutta and Jindal [6] developed a system which is focused on extracting reduplication words from the entered text and also identified the category of extracting words based upon syntactic and semantic analysis. Reduplication words belong to a class of multi-word expressions. Classification of reduplication words used in parsing and dictionary based applications like cross lingual information retrieval and machine translation. The approach is used in identification and classification of reduplication words is that the two words are separated by hyphens is translated into English language and then they are compared from the backend. Depending according to the degree of similarity they are analyzed into different categories of reduplication. They have used the rule based theory for identification and classification of reduplication words in different categories which can be used to design a lexicon. Singh, Ojha and Jha [7] discussed that reduplication is an important sub-class of multi-word expressions and they are highly used in Hindi language. Disambiguating Multi-Word Expressions (MWEs) is a very difficult task in Natural Language Processing (NLP) applications. The main challenge in classification and identification of reduplicated multi-word expressions (RMWEs) in Hindi is linguistic. The paper focused on the linguistic challenges and their formalization aspects using the concept of conditional random field. According to the authors for annotation of multi word expressions in Hindi, there are no guidelines available. Therefore, they have presented first guideline for annotation, multi-word expressions in Hindi language. In this paper authors reduced the overall accuracy of RMWEs extractors by using an experimental setup of RMWEs and dealing with the challenges occurred during the experiment. Gupta, Dutta and Rana [8] presented two different techniques for identification of reduplication words in Hindi language. There are three categories of reduplicated words based on the patterns. First one is the repetition of the same syllable twice. The second is a repetition of syllable partially same and in the third both syllable is totally different. Authors combined corpus based approach with a rule based approach to identify reduplication words in Hindi language. The rule based approach focused on semantic characteristic of the language and the corpus is created from Hindi books and

Hindi literature. From the rule based technique author found that whether the rule satisfied the constraints reduplicated words or not and also identified whether the reduplicated words found in the corpus or not. Dolatian and Heinz [9] described an efficient method for the computational modeling of reduplication. Reduplication is a linguistic phenomenon and it is treated as a stumbling block of finite state treatment for morphology. In this paper they worked on the strategy that most finite state implementations neither count the productivity of unbounded copying in reduplication nor the bounded copying in reduplication. An understudied type of finite state machine and two ways finite state transducer catch all reduplicated processes. In this paper a small but representative typology of reduplication process is explained. Finite state technology catches productive reduplication as used in natural language. This paper concluded that understudied type of finite state machines-two way finite state transducer can capture model reduplication and its typology. Al-Asbahi [10] compared two languages, English and Arabic in terms of the semantics of reduplication. This paper explained that in two languages found more similarity than differences in terms of reduplication. Arabic reduplication is semantically more productive than English language. Both languages have the same classification of reduplication like free/bound, full/partial and continuous/discontinuous. Both languages share the sense of reduplication like contempt, repetition, intensity, affection, emphasis, plurality, onomatopoeia, non-uniformity, scatter, completion, movement, continuity, lack of control and spread out. Ambiguity was common between both languages when the semantic connection was developed between most of these concepts. Both the languages explained reduplication in prayers, second language teaching, nursery rhymes, lyrics, slogans, newspaper headlines, advertisement and children phonics cartoons. One difference found by the author in the semantics of reduplication that the features of pluractionality, diminution, and augmentation are only available in Arabic language. These features are not available in English language. In this paper the author concluded that Arabic reduplication was semantically more productive than English reduplication. Noor et al. [11] classified reduplicated into different categories and explained different semantic functions. They also presented various examples to support the observations of research. The classifications of reduplication defined by the examples are partial and full reduplication. In this paper some examples also explained rhyming reduplication and echoic reduplication. The paper concluded that the reduplication is the most important morpho-semantic phenomenon which plays a large role in the word formation process of any language.

3 EXPERIMENTAL EVALUATION

The proposed system is tested to evaluate the effectiveness of developed algorithm. The application is implemented by using Microsoft visual studio 2010 as front end and Microsoft SQL server as backend. The code of the application is written in C# language. The proposed system is tested on 50 paragraphs. Each paragraph contains more than 40 words.

3.1 Techniques Used

3.1.1 Rule based technique

Rule based is an interesting top down approach. In this approach we have used the condition like if, then, and, or, not and various ranges of near. We have divided the sentence into tokens. These tokens are stored in the array. Each token contains a word which is compared with the next word. Through comparison, we can find the difference between two consecutive words. Let suppose we take the sentence"ਬਾਰਿਸ਼ ਦੀਆ ਬੂੰਦਾਂ ਟਿਪ ਟਿਪ ਬਰਸ ਰਹੀਆਂ ਹਨ". In this sentence when the algorithm compares two words 'ਬੂੰਦਾਂ' and'ਟਿਪ', there is no matching. Same as when the algorithm compares next pair of words and, all the letters of the word completely matched and our algorithm identifies that this is the case of full reduplication because the total length of the word is 3 and the total number of letters is also 3. The difference between the total length of the word and total successful match count is 0 which identifies that the words in comparison are type of full reduplication. If the difference between the total length of the word and total matching count is non-zero, then it is the case of partial reduplication. For example'ਦੀਵੇ ਜਗ ਮਗ ਕਰ ਰਹੇ ਹਨ', in this sentence when two consecutive words'ਜਗ'and 'ਮਗ' compared, total length of the first word is 2 and the successful match is 1. The difference between total length and total successful match count of letters is 1 which identifies that these words are a type of partial reduplication. As all the words stored in the array, the designed algorithm compares two consecutive words as well as compare two words by skipping the centered word. In the example 'ਉਥੇ ਅਸੀਂ ਸਾਰੇ ਜਾਵਾਗੇ ਸਾਰੇ', the algorithm firstly compares the two consecutive words i.e.'ਸਾਰੇ &'ਜਾਵਾਗੇ' and then compare the two words by skipping the centered word i.e. 'ਸਾਰੇ'and 'ਸਾਰੇ'. When the algorithm finds such cases in which a single word successfully matched with second next words, then this is the case of discontinuous word reduplication.

3.1.2 A corpus based approach

The corpus based approach has been developed to support investigation of language variation and use by generating a database of linguistic words. The main strength of the corpus based approach is that test corpus contains large database of patterns of language which combined with a rule based approach to make an efficient research tool on natural languages. In this research, we make a corpus of 1000 semantic reduplicated words of Punjabi language. As we discussed earlier in rule based approach, the related continuous words are identified by our algorithm and also classified whether it is a partial or full reduplication. There are some words which are semantically reduplicated like'ਬਾੜ੍ਹਾ ਬਗੀਚਾ'and'ਅੱਜ ਕੱਲ', which are not identified by the rule based approach. There are two words 'ਬਾੜ੍ਹਾ' and 'ਬਗੀਚਾ'which are not recognized by rule based approach, but the meaning of both words is same i.e. 'garden'. Similarly'ਅੱਜ'and'ਕੱਲ'are two different words, but the concern of both the words is with the 'present situation'. We make a corpus of these semantic reduplicated words which is matched with the each word of the entered paragraph by the user.

3.2 Results and Discussion

In this paper reduplication words are identified and these words are classified into different types. An example is given below which shows Punjabi text paragraph entered in the developed application.

ਮੋਹਨ ਨੇ ਕਿਹਾ ਅਸੀਂ ਸਾਰੇ ਜਾਵਾਗੇ ਸਾਰੇ। ਤਾਂ ਮੋਹਨ ਦੀ ਮਾਂ ਨੇ ਕਿਹਾ ਰੋਟੀ ਰੁਟੀ ਖਾ ਕੇ ਜਾਣਾ। ਬਾਹਰ ਪਾਣੀ ਦੀਆ ਬੂੰਦਾਂ ਟਿਪ ਟਿਪ ਕਰ ਕੇ ਬਰਸ ਰਹੀਆਂ ਹਨ। ਆਸ ਪਾਸ ਕੋਈ ਬੱਸ ਬੁੱਸ ਵੀ ਨਹੀਂ ਮਿਲਣੀ। ਮੋਹਨ ਅਤੇ ਉਸਦੇ ਦੋਸਤ ਰੋਟੀ ਖਾ ਕੇ ਜਲਦੀ ਜਲਦੀ ਨਿਕਲ ਗਏ ।

From the above paragraph, experimental results of developed application are given below:
Total no. of words in the paragraph- 50
Actual no. of reduplicated words in the paragraph- 6
No. of reduplicated words identified by the proposed system-6
No. of reduplicated words whose reduplication type is identified- 6
Accuracy of the proposed system- 100%
Table 10 shows the identified reduplicated word from the entered paragraph and their classification in three different types.

TABLE 10
IDENTIFICATION AND CLASSIFICATION OF REDUPLICATED WORDS

S. No.	Actual reduplicated words	Reduplicated words identified by the proposed algorithm	Type of reduplication
1.	ਸਾਰੇ ਜਾਵਾਗੇ ਸਾਰੇ	ਸਾਰੇ ਜਾਵਾਗੇ ਸਾਰੇ	Discontinuous
2.	ਰੋਟੀ ਰੁਟੀ	ਰੋਟੀ ਰੁਟੀ	Partial
3.	ਟਿਪ ਟਿਪ	ਟਿਪ ਟਿਪ	Full
4.	ਆਸ ਪਾਸ	ਆਸ ਪਾਸ	Partial
5.	ਬੱਸ ਬੁੱਸ	ਬੱਸ ਬੁੱਸ	Partial
6.	ਜਲਦੀ ਜਲਦੀ	ਜਲਦੀ ਜਲਦੀ	Full

The statistics of the overall results of our research are represented in Table 11. Fig. 2 shows the graphical representation of statistical results. In which vertical axis represents the actual no. of reduplicated words in the paragraph, no. of reduplicated words identified by the proposed system, no. of reduplicated words whose reduplication type is identified and the horizontal axis represents the total no. of words in the paragraph.

TABLE 11
STATISTICAL RESULTS OF PROPOSED ALGORITHM

S. No.	Total no. of words in the paragraph	Actual no. of reduplicated words in the paragraph	No. of reduplicated words identified by the proposed system	No. of reduplicated words whose reduplication type is identified	Accuracy of the proposed system (%)
1'	75	10	9	9	90
2	100	12	11	11	92
3	110	12	11	11	92
4	120	15	14	13	87
5	150	18	17	16	89
	Avg.	**13.4**	**12.4**	**12**	**90**

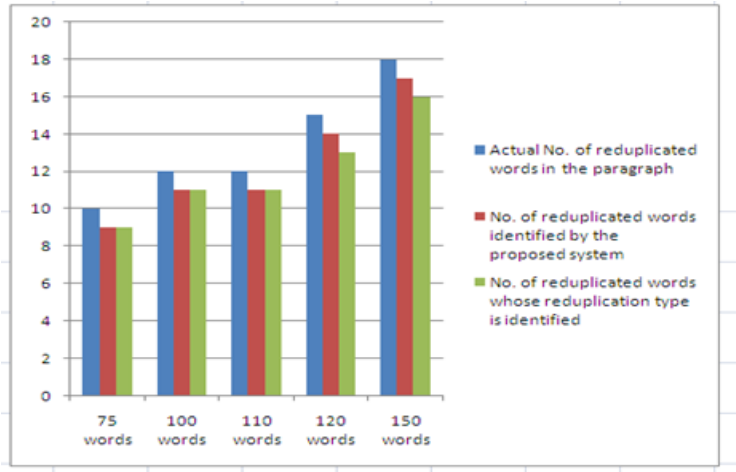

Fig. 2*. Graphical representation of statistical results*

Fig. 3 shows the graphical representation of the accuracy of the proposed algorithm. In which vertical axis represents the accuracy percentage of the proposed system with respect to the total no. of words in the paragraph.

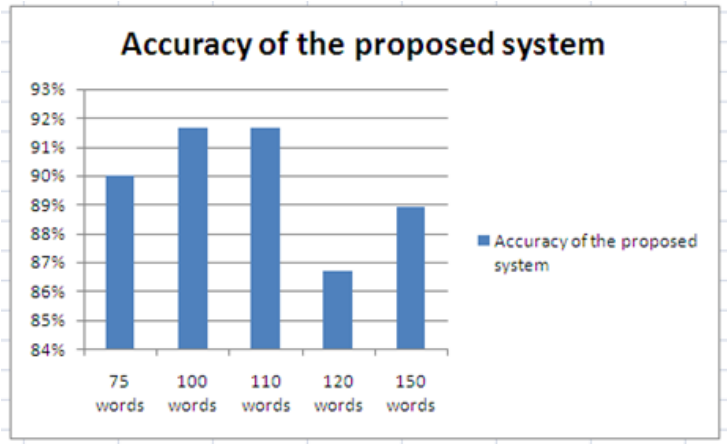

Fig. 3*. Graphical representation of the accuracy of proposed algorithm*

4 CONCLUSION AND FUTURE SCOPE

Large amount of work in the identification and classification of Reduplication words has been done in English language, as English is a global language. But there is a need to perform research on reduplication words in regional languages also. Very less research work is done on reduplication words in Punjabi language. In this paper, we have designed an algorithm and develops Graphical User Interface which accepts input as a Punjabi text and gives output by highlighting reduplicative words and also classified the types of identifying reduplicative words. The proposed framework yields satisfactory results on more than 50 paragraphs of Punjabi text with the over-all accuracy of more than 85% for reduplication identification and classification. Future work includes incorporating more classification to improve the existing results. As the designed algorithm classified reduplication words into three types Full, Partial and Discontinuous. In future researcher may improve our algorithm and classify reduplication words in more types. In Punjab state most of the official work is done in Punjabi language, so this

research will help the applications developed in Punjabi linguistic.

REFERENCES

[1] M.C. Surabhi, "Natural Language Processing Future," Proc. IEEE Optical Imaging Sensor and Security, 2-3 July 2013, doi: 10.1109/ICOISS.2013.6678407.

[2] D. Kaur and N. Kaur, "Implementation of DJ Rule Based Algorithm for Dhuni Vishleshan of Compound Punjabi Words," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 7, July 2013.

[3] O. Iheanetu and M. Adeyeye, "Finite State Representation of Reduplication Processes in Igbo," IEEE, 2013, doi: 10.1109/AFRCON.2013.6757772.

[4] V. Gupta and A. Sharma, "Classification of the Spoken Hindi Partially Reduplicated Words using Artificial Neural Network," International Journal of Computer Applications, vol. 93, no. 10, May 2014.

[5] M.I. Khan, "Reduplication in Arabic and Urdu," International Journal of English and Education, vol. 5, no. 4, pp. 336-344, Oct. 2016.

[6] K. Dutta and A. Jindal, "System for Identification and Analysis of Reduplication Words in Hindi Corpus," International Journal of New Technology and Research, vol. 2, no. 4, pp. 18-21, April 2016.

[7] R. Singh, A.K. Ojha and G.N. Jha, "Classification and Identification of Reduplicated Multi-Word Expressions in Hindi," Proc. 10th LREC2016, pp-18-22, May 2016.

[8] S.K. Gupta, K. Dutta and P. Rana, "Issues and Challenges of Reduplication in Hindi," International Journal of Engineering Science, vol. 27, pp. 162-172, March 2018.

[9] H. Dolatian and J. Heinz, "Modeling Reduplication with 2-Way Finite-State Transducers," Proc. 15th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, pp. 66–77, Oct. 2018.

[10] K.M.M. Al-Asbahi, "Insights into the Semantics of Reduplication in English and Arabic," International Journal of English Linguistics, vol. 10, no. 1, pp. 384-394, Jan 2020.

[11] M. Noor, R.A. Mangrio, F. Muhabat and M. Iqbal, "Reduplication in Punjabi: A Morpho-Semantic Phenomenon," Journal for Studies in Management and Planning, vol. 1, no. 3, April 2015.