

Simplification of Punjabi Sentences: Converting Complex Participial Sentences into Simple Sentences

R.M. Jindal^{1,*}, V. Rana¹ and S.K. Sharma²

¹Sant Baba Bhag Singh University, Jalandhar

²DAV University, Jalandhar

Abstract

INTRODUCTION: In this world of internet and artificial intelligence, Natural Language Processing has emerged as most demanding research area. Under Natural Language Processing, sentence simplification is one of the research area that deals with simplification or conversion of complex sentences into simple sentences. **OBJECTIVES:** In this research article, author has proposed a novel approach for conversion (simplification) of complex sentences (participial type) of Punjabi language into easily understandable simple sentences. **METHODS:** Author performed lexical and morphological simplification by using morphological features of the language. Morphological features are used to identify the participial type complex sentences. **RESULTS:** on testing the proposed algorithm, a precision of 96.39%, recall of 91.37% and F-measure as 93.79% was reported. **CONCLUSION:** the developed system can be helpful for Aphasic and Dyslexia readers and can be used as subpart for machine translation system, summarization system and other Natural Language Processing applications.

Keywords: Paraphrasing, sentence simplification, participial sentences, Punjabi language, complex sentences.

Received on 02 December 2019, accepted on 22 February 2020, published on 27 February 2020

Copyright © 2020 R.M. Jindal *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.13-7-2018.163338

*Corresponding author. Email: Rmjindal2002@yahoo.co.in

1. Introduction

Sentence simplification is one of the natural language processing tool that transforms the sentences from complex to simple. Sentence simplification play an important role to help the people suffering from Aphasic and Dyslexia[1]. The people suffering from Aphasia are unable to understanding long and complex sentences and hence they need a tool that simplify the complex sentence (lexical simplification). On the other hand, the people suffering from Dyslexia are unable to understanding complex words[2]. Also the simplification process helps people having limited vocabulary that results in difficulty in learning any new language. Further, long sentences are difficult to parse and hence Text Simplification increase

the throughput of the parser. Also the efficiency of machine translation system will be improved after breaking the long complex sentences into simple small sentences. Also a significance improvement in the accuracy of Text Summarization is observed by simplification of sentences. Further the sentence simplification can be helpful in various other NLP related problems as discussed by [3], [4] and [5]. Even in future, the simplified sentences will be of great useful in disease diagnosis [6], [7] and [8] from medical reports and in rule mining [9].

Punjabi is one of the top ten languages of the world. The “Gurmukhi” and the “Shahmukhi” scripts are used to write Punjabi language. Further in speaking, there are many dialects of Punjabi language [10]. Due to large number of user of Punjabi language, it has significance importance in Natural Language Processing. Many

fundamental and advance tools are being developed for processing of Punjabi language. In this research article Punjabi sentences are transliterated into Roman using Gtrans tool [11].

2. Related Work

As per existing literature, various researchers have used either manual approach [12] or automated approach for simplification of text. The first text simplification system for English was developed by Boeing [13]. After that, phrase based statistical machine translation system were developed by [14-16] for simplification of English sentences. In general, there are similarities between various techniques used to simplify text, to create paraphrase, to summarize text, to generate text from text and machine translation [15, 17]. As per existing literature studied, various authors have done work for simplification of sentences on various languages. The languages on which work has been done are Dutch [18], Brazilian Portuguese [19], French [20], Vietnamese [21], Basque [22], Italian [23], Korean [24], Spanish [25] etc. Further lexical approach is used to simplify the sentences written in English language. This lexical simplification approach is used to develop many tools to assist Aphasic reader. These tools include PSET [14], SIMPLEX [26], KURA [27], HAPPI [28], LexSiS [29] etc. Additionally many researchers have used syntactic approach for simplification. In this approach simplified sentence/sentences are generated from the original large complex sentence. Various methods used for syntactic simplification includes splitting the original sentence into its sub-clauses, converting the sentence into passive voice or by resolving anaphora. Many researchers used it for simplification of text by inducting automatic rules [30], simplifying the text for the applications that seek information [18], during grammatical simplification to keep the discourse unaffected[8], dividing long sentence into smaller parts when creation of explanation [32], Enhancing summaries by simplifying sentences [33], developing text simplification authorising tools [19], by assisting Amphatic readers through simplification of Newspaper text [34], syntactic simplification of French [20], simplification of Vietnamese sentences [21], Bosque sentences [22], Italian sentences [23], Korean sentences [24], for parse tree manipulation [35], eliminating excessive chunks of sentences [36] and for simplification of Spanish sentence [25].

3. Participial Sentences

These are complex sentences containing a dependent clause having non-finite verb phrase which bound the dependent clause (predicate) with independent clause [Brar, B.S.,1995]. Consider the following example:

Punjabi	Transliterated	Translated
---------	----------------	------------

ਰੱਸੀ ਟੱਪਦਿਆਂ ਮੇਰੇ ਪੈਰ ਨੂੰ ਮੋਚ ਆ ਗਈ।	(rassī tappdiām mērē pair nūm mōc ā gāi)	I cricked my foot while hopping the string.
-------------------------------------	--	---

In above example, ਰੱਸੀ ਟੱਪਦਿਆਂ (rassī tappdiām) is the predicate (dependent clause) and the word ਟੱਪਦਿਆਂ (tappdiām) is responsible for binding it to in-dependent clause i.e. ਮੇਰੇ ਪੈਰ ਨੂੰ ਮੋਚ ਆ ਗਈ (mērē pair nūm mōc ā gāi):-

Sr. No	Type of Non-finite verb	Example
1.	Participial: - Contains non-finite verb like ਟੱਪਦਿਆਂ, ਵੇਖਿਆ, ਕੀਤਿਆਂ i.e. contains ਇਆਦੀਆਂ and ਈ as postfix with verb.	ਪੁਲਸ ਨੂੰ ਵੇਖਦਿਆਂ ਹੀ ਚੋਰ ਖਿਸਕ ਗਿਆ । (pulas nūm vēkhdīām hī cōr khisak giā .)

Some examples showing complex sentence having participial sentences are tabulated in table 1:

Table 1. Some examples of Participial sentences in Punjabi language

Sr. No	Punjabi Sentence	Roman Transliterated sentences
1	ਬਿਨਾਂ ਭਕਾਈ ਕੀਤਿਆਂ ਉਸ ਨੇ ਨਹੀ ਜੇ ਮੰਨਣਾ । (binām bhakāi kītiām ēs nē nahī jē manṇā) .	He didn't do it without being bothered.
2	ਦਿੱਲੀ ਜਾਂਦਿਆਂ ਉਹ ਰਾਹ ਵਿੱਚ ਜਾਖਲ ਉਤਰ ਗਿਆ । (dillī jāndīām uh rāh vicc jākhal utar giā) .	On his way to Delhi, he landed on the road.
3	ਪ੍ਰਹਿਲਾਦ ਦੇ ਅੱਗ ਵਿੱਚ ਜਾਂਦਿਆਂ ਭਾਂਬੜ ਠੰਡਾ ਹੋ ਗਿਆ । (prahlād dē agg vicc jāndīām bhāmbar ṭhāṇḍā hō giā) .	The flame went cold as the Prahlad's fire burst.
4	ਸੂਰਜ ਵੱਲ ਵੇਖਿਆ ਅੱਖਾਂ ਚੁੱਧਿਆ ਜਾਂਦੀਆਂ ਹਨ । (sūraj vall vēkhiā akkhām cundhiā jāndīām han) .	Looking at the sun, eyes are wiped.
5	ਖੋੜੀ ਦੇ ਖਲੋਦਿਆਂ ਹੀ ਉਹ ਉਤਰ ਗਿਆ । (ghōṛī dē khalōndīām hī ēh utar giā) .	He got down just as the horse was down.
6	ਬਿਨਾ ਬੋਲਿਆ ਚੋਰ ਆਦਮੀ ਦੇ	The silent thief followed the men.

	ਮਗਰ ਹੇ ਤੁਰਿਆ । binā bōliā cōr ādmīm dē magar hō turiā .	
7	ਬਿਨਾਂ ਖੰਘਿਆਂ ਹੀ ਬਜ਼ੁਰਗ ਅੰਦਰ ਲੰਘ ਆਇਆ । binām khaṅghīām hī bazurag andar laṅgh āiā .	The elder passed without a potion.
8	ਬਿਨਾਂ ਪੱਗ ਬੰਨ੍ਹਿਆਂ ਉਹ ਦੀਵਾਲੀ ਵਾਲੇ ਦਿਨ ਬਾਹਰ ਫਿਰਦਾ ਰਿਹਾ । binām pagg bannhiām uh dīvālī vālē din bāhar phirdā rihā .	He wandered outside on Diwali day without wearing a turban.
9	ਕਾਲੀ ਕੀਤਿਆਂ ਕੰਮ ਖ਼ਰਾਬ ਹੋ ਜਾਂਦਾ ਹੈ । kālhī kītiām kamm kḥarāb hō jāndā hai .	Hurrying is a waste of work.

4. Proposed algorithm for simplification of participial sentences

As discussed above in literature review section, various algorithms have been used by various researchers for simplification of long sentences. The algorithms used depends upon the type of language and structure of the sentence of that language. The simplification process takes place in the sequence as word level simplification (lexical simplification), syntactic simplification (sentence level) and in the last discourse simplification can also be applied (reduction in the size of sentence). In lexical simplification complex words (Words which are difficult to read and understand) are replace with simple one. In this research work, to identify the complex words frequency based technique has been used. Figure 1 shows the general simplification procedure.



Figure 1. General simplification procedure

4.1 Lexical Simplification (Identifying Complex Words)

This is the first step of sentence simplification in which complex words are replace with simple one. Further the very first step in lexical simplification is to identify complex words (CWs). This is the process of scanning a text and picking out the words which may cause a reader difficulty. Getting this process right is important, as it is at the first stage in the simplification pipeline. Hence, any

errors incurred at this stage will propagate through the pipeline, resulting in user misunderstanding.

Identification of complex words

There are several factors that come together to form lexical complexity. Generally, word frequency is either used by itself, or combined with word length to give a continuous scale on which complexity may be measured. There are a few different methods in the literature for actually identifying CWs. The most common technique unsurprisingly involves attempting to simplify every word and doing so where possible. Lexical complexity can be used to determine which words are complex in a given sentence. To do this, a threshold value must be established, which is used to indicate whether a word is complex. Selecting a lexical complexity measure which discriminates well is very important here. Machine learning may also be used to some extent. Typically, Support Vector Machines (SVMs, a type of statistical classifier) have been employed for this task. Lexical and syntactic features may be combined to give an adequate classifier for this task.

In our research work we developed a synonym database containing more than 10,000 words. Each word is then assigned a frequency based upon its use in the corpus containing more than 70,000 sentences. Now each input sentence is scanned and if a word has more frequency as compare to its synonym then this word is considered as complex and it will be replaced by its synonym having more frequency than the word.

4.2 Syntactic simplification

Here the changes are applied at sentence level. Depending upon the type of sentence different algorithms have been developed for simplification of complex sentences. Now to simplify the complex sentences first dependent and independent clause are identified and then dependent clauses are separated from independent clauses. These dependent clauses are either converted into independent clause or they are removed (if not much affect the meaning of sentence).

Identification of Dependent Clauses

Dependent clause is always a part of complex sentence. It does not exist independently and always exists with independent clause. Therefore, while marking the dependent clause, independent clause is also marked. In the following section, we have described the identification and marking of various types of dependent clauses as well as independent clauses in complex sentence.

Methodology used

In this research work, the syntactic cue and morphological information have been used for clause boundary identification. The morphological information used includes suffix information of non-finite verb and even part of speech tag at some places. Syntactic cue includes

presence of conjunction or comma. Different morphological and syntactic cues have been used for different type of dependent clauses. For example, suffix information of non-finite verb has been used for marking clause boundaries in complex sentences containing predicate bound type of clauses; subordinate conjunctions are used to mark the clause boundaries in complex sentences containing non-predicate bound type of clauses. In the following section, detailed description about identification of dependent clauses has been provided.

Predicate bound clause

Predicate Bound clauses are those dependent clauses in which non-finite form of the verb (predicate) bounds the dependent clause on independent clause. These clauses are further subdivided in to three categories. Clause boundary identification in all these three categories have been discussed in the following section:

Participial

This type of dependent clause contains present perfect and past perfect non-finite verb. These non-finite verbs are identified from their suffixes i.e. presence of ਦਿਆਂ (Diā), and ਇਆਂ (iā) respectively as suffix in the verb. Consider the following examples:

Sentence 1: ਸਕੂਲ ਜਾਂਦਿਆਂ ਉਹ ਰਾਹ ਵਿਚ ਖੇਡਣ ਲੱਗ ਪਿਆ (sakūl jāndiām uh rāh vic khēḍāṃ lag piā)

While going to school, he started playing on the way.

Sentence 2: ਬਿਨਾਂ ਭਕਾਈ ਕੀਤਿਆਂ ਉਸਨੇ ਨਹੀ ਜੇ ਮੰਨਣਾ। (binām bhakāī kītiām ēsnē nahī jē manṇā.)

He will not agree unless you try hard to convince him.

In the above examples, sentence 1 has present perfect non-finite verb ਜਾਂਦਿਆਂ (jāndiām) with suffix ਦਿਆਂ and sentence 2 has past perfect non-finite verb ਕੀਤਿਆਂ (kītiām) having suffix ਇਆਂ. It can be concluded from these examples that subordinate verbal phrase of predicate bound types of sentences are positioned just before the independent clause of the sentence. So, the starting point of the dependent clauses in participial type of sentence will be the subject of the sentence and the end point will be the non-finite verb of the subordinate verbal phrase or dependent clause. The clause boundary mark of dependent and independent clauses in sentence number 1, and 2 is shown below:

ਰੱਸੀ ਟੱਪਦਿਆਂ ਮੇਰੇ ਪੈਰ ਨੂੰ ਮੋਚ ਆ ਗਈ
Dependent clause *Independent clause*

ਬਿਨਾਂ ਭਕਾਈ ਕੀਤਿਆਂ ਉਸ ਨੇ ਨਹੀ ਜੇ ਮੰਨਣਾ
Dependent clause *Independent clause*

Various features of participial non-finite verb with example are provided in table 2.

Table 2. Features of participial non-finite verb

Features of participial non-finite verb	Feature value(Suffix)	Example
Word related feature (morphological features)	ਦਿਆਂ (Diā), ਇਆਂ (iā)	ਜਾਂਦਿਆਂ, ਸੌਂਦਿਆਂ, ਰੋਂਦਿਆਂ, ਬੋਲਦਿਆਂ, ਖਾਂਦਿਆਂ, ਆਉਂਦਿਆਂ, ਪੀਤਿਆਂ, ਕੀਤਿਆਂ, ਖਲੋਤਿਆਂ, ਜਾਗਿਆਂ, ਸੋਇਆਂ
POS tag related features(suffix used in the tag)	DIAN, NIAN	VBMAXXXXXI NDIAN VBMAXXXXXT NIAN

Algorithm used: Clause boundary identification of non-predicate bound participial type.

Input: Annotated Punjabi sentence

Database used: Features of participial non-finite verb

Output: Punjabi sentence with marked clause boundaries.

```

1 A ← ∅
2 S ← {si}
3 s0 ← pop(S)
4 IDb ← s0
5 i ← 1
6 While S ≠ ∅ do the following
7   si ← pop(S)
8   if si is participial finite verb as listed in table 2
9     si ← IDe
10    si+1 ← IDb
11    A ← AU{ si }
12  else
13    A ← AU{ si }
14  end
15  si ← IDe
16 end
    
```

Flow chart representing above mentioned algorithm is shown in figure 2. An example to illustrate the working of flowchart/algorithm is also provided with this flow chart. Symbol used:

ID_b: Beginning of independent clause

ID_e: End of independent clause.

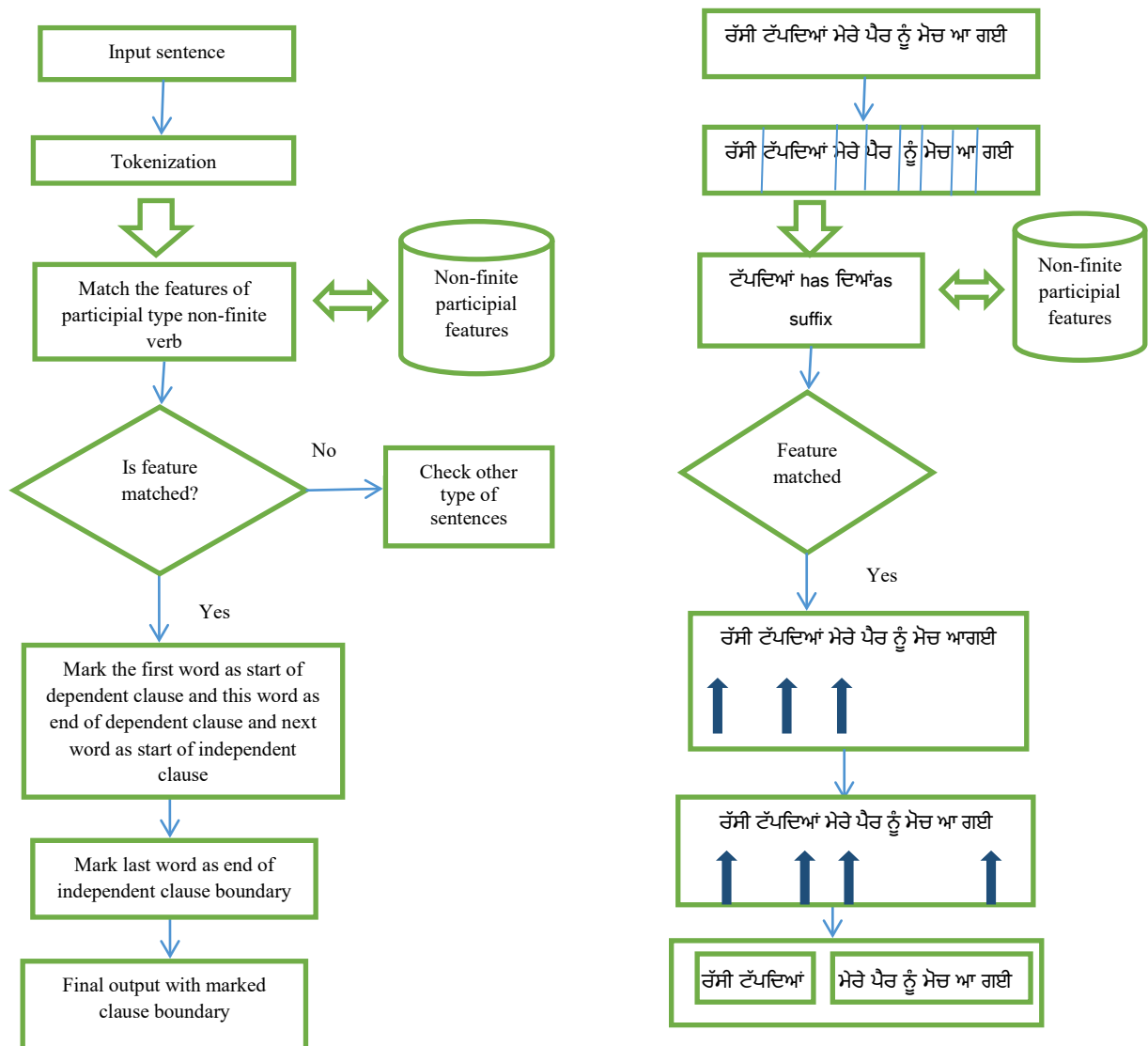


Figure 2. Flow chart and working example to mark clause boundary in participial type complex sentences

Conversion of dependent clause into independent clause

Only the non-finite verb makes the clause as dependent clause. Therefore by converting the non-finite verb into finite verb dependent clause can be converted into independent clause. In the above mentioned example, the dependent clause is: "ਰੱਸੀ ਟੱਪਦਿਆਂ" and in this clause ਟੱਪਦਿਆਂ is the non-finite verb. Following algorithm is implemented to convert the non-finite verb into finite verb:

- Extract non-finite verb from dependent clause.
- Extract the root of the verb by applying stemmer to the non-finite verb and separate the root and suffix portion.

e.g.

ਟੱਪਦਿਆਂ -> ਟੱਪ (Root Word) + ਦਿਆਂ (Suffix part)

- By using suffix part complete the verb phrase (like ਟੱਪਦਿਆਂ will be converted into ਟੱਪ ਰਿਹਾ ਸੀ).
- An optional subject can be added as per the subject of independent clause.

Input Sentence	Output Simplified sentences
ਰੱਸੀ ਟੱਪਦਿਆਂ ਮੇਰੇ ਪੈਰ ਨੂੰ ਮੋਚ ਆ ਗਈ ।	ਰੱਸੀ ਟੱਪ ਰਿਹਾ ਸੀ । ਮੇਰੇ ਪੈਰ ਨੂੰ ਮੋਚ ਆ ਗਈ ।

4.3 Content reduction

In this phase un-necessary content is discarded from the sentence. There are few sentences in which there exist some content that does not contribute in the meaning of sentence and hence is useless. Such content can be discarded. Such situation occurs when the dependent part of participial sentences is not required as this part convey the same information as provided by the previous sentence. Consider the following example:

ਮੁੰਡਾ ਘਰ ਜਾ ਰਿਹਾ ਸੀ । ਉਹ ਘਰ ਜਾਂਦੀਆਂ ਹੀ ਸੋਨ ਚਲਾ ਗਿਆ ।

As in above sentence, the dependent clause ਉਹ ਘਰ ਜਾਂਦੀਆਂ ਹੀ can be discarded as the information provided by this dependent clause is already given by the previous sentence i.e. ਮੁੰਡਾ ਘਰ ਜਾ ਰਿਹਾ ਸੀ ।

5. Datasets used for testing

For the developed system, author developed four datasets from four different online resources. Four different resources are used for development of test corpus. These resources include ILCI (Indian Languages Corpora Initiative) [15]. Total 31 files of seven different domains (Agriculture, entertainment, Health, Literature, religion, Science and Technology and sports) were used. Further details like number of total sentences and number of participial sentences has been provide in table 3. Also the detailed analysis is shown in figure 3.

Table 3. ILCI data set for Punjabi language

File name	Number of files	Total number of sentences	Number of Participial type sentences
Pun_agriculture.txt	11	11000	234
Pun_entertainment.txt	03	3000	101
Pun_health.txt	01	1000	37
Pun_literature.txt	04	4000	112
Pun_religion.txt	05	5000	109
Pun_science_and_technology.txt	02	2000	29
Pun_sports.txt	05	5000	138
Total	31	31000	760

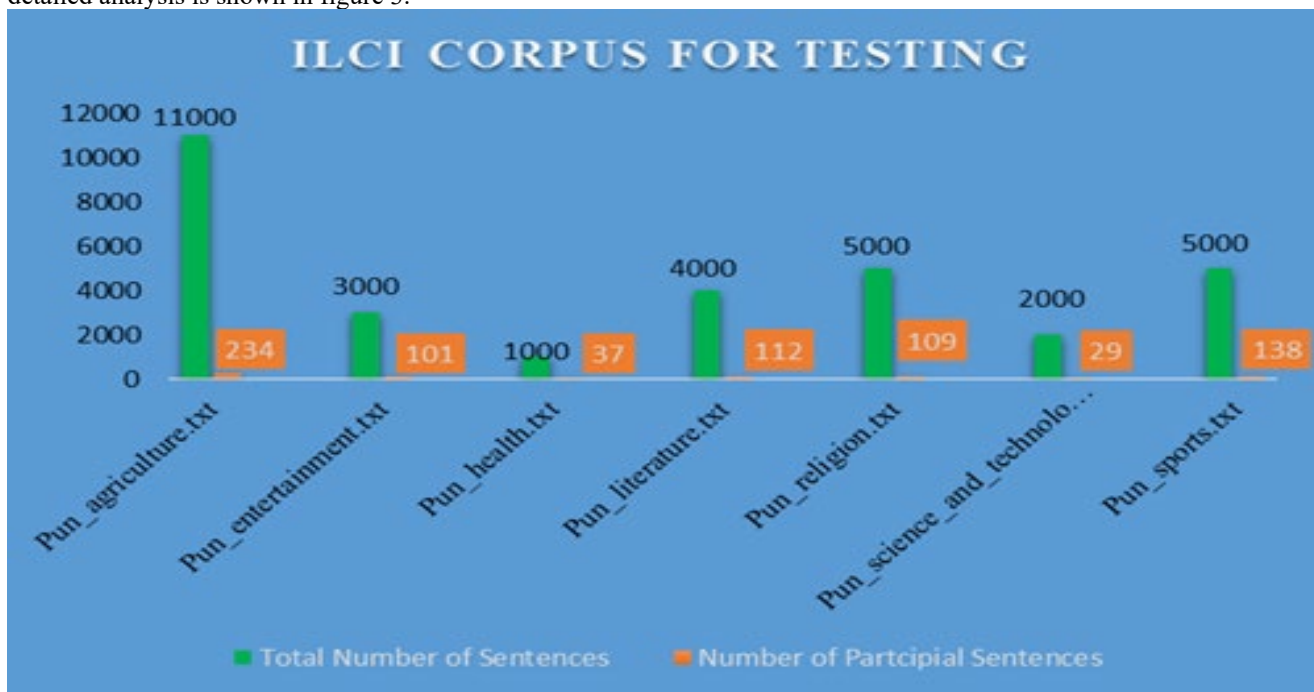


Figure 3. Details of ILCI corpus used for testing

The second resource used for development of test corpus is e-papers. These e-papers are freely available on the internet and sentences can be copied from them. Author used ten Punjabi newspapers (Punjabi Tribune, Punjabi Jagran, Jag bani, Daily Punjab Times, Ajit Jalandhar, Daily Aashiana, Anonymous Newspaper, The Times of Punjab Newspaper, Parvasi Newspaper). Table 4 provides the further details like number of sentences extracted and number of participial sentences present in them. Also the detailed analysis of e-papers is shown in figure 4.

Table 4: Corpus collected from online sources (e-papers)

Sr. No	Name of newspaper	Website Link	Total no. of sentences collected	No. of Participial type sentences
1	Punjabi Tribune	https://epaper.punjabtribuneonline.com	1000	124
2	Punjabi Jagran	https://epaper.punjabijagran.com	1000	78
3	Jag bani	https://jagbani.epapr.in	1000	113
4	Daily Punjab Times	epaper.dailypunjabtimes.com	1000	145
5	Ajit Jalandhar	https://www.ajitjalandhar.com	1000	94
6	Daily Aashiana	https://epaper.dailyaashiana.com	1000	38
7	Anonymous Newspaper	https://www.readwhere.com	1000	64
8	The Times of Punjab	epaper.thetimesofpunjab.com	1000	218
9	Newspaper	https://www.w3newspapers.com	1000	223
10	Parvasi Newspaper	https://parvasinewspaper.com	1000	93
	Total Sentences		10000	1190

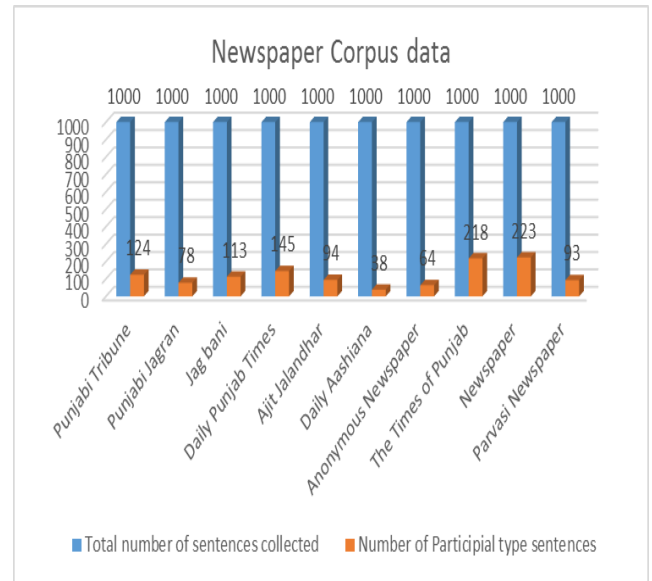


Figure 4. Details of e-paper corpus used for testing

The third source of test corpus creation used by author is other online resources. The various categories used for extraction of sentences are stories, essay, and literature and text books. Further details like the website used and number of sentences etc. has been provided in table 5. Also the details analysis of collected corpus is provided in the figure 5.

Table 5. Other Online Resources

Type of corpus	Website Link	Total number of sentences collected	Number of Participial type sentences
stories	https://punjabistories.com www.punjabikahani.punjabikavita.com www.shurli.com https://www.absolutestudy.com https://sikhville.org	10000	1541
essay	https://www.sbs.com.au www.shurli.com	10000	678

	www.anlc.it https://www.irisbg.com https://brainly.in		
literature	https://www.britannica.com https://www.indianmirror.com https://uddari.wordpress.com	10000	779
Text books	www.pseb.ac.in	10000	431
Total		40000	3429

The last resource used by the author is manually created participial sentences. Five hundred participial sentences were created manually. These include most of the sentences from day today spoken language.

Table 6. Manually created test data

Type of corpus	Number of Participial type sentences created manually
Test corpus	500

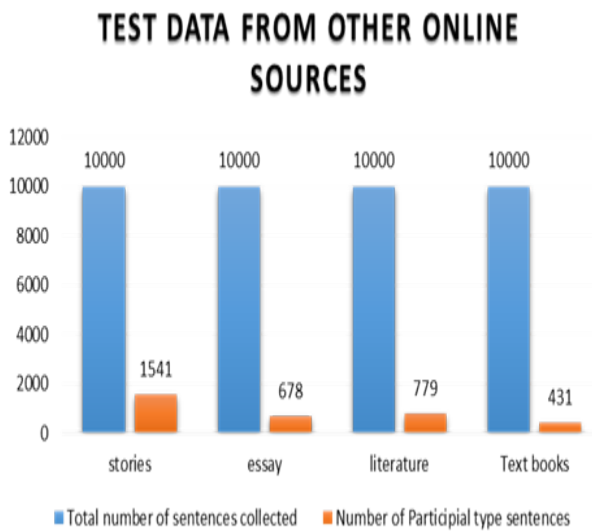


Figure 5. Details of the various online corpus used for testing

6. Result and Discussion:

As discussed in previous section (data set used) and details provided in table 3, table 4, Table 5 and table 6, the test corpus was used to simplify the sentences and the results obtained are shown in table 7. As shown in table 7, the proposed algorithm shows a precision of 96.39%, recall of 91.37% and F-measure as 93.79%. There are two main problems that occurs while converting the participial sentences into simple sentences. One is the Length of participial sentence is generally not very long. Therefore, it becomes difficult to separate this type of complex sentences from simple sentences. Second is that the dependent part of the participial sentence has no subject and hence it becomes difficult to assign the subject automatically. The analysis of the overall datasets used for testing and results obtained is shown in figure 6 and figure 7 respectively. Figure 6 shows that 9% test data used was created manually, 13% data was taken from ILIC resources, 20% from e-papers and rest 58% was taken from the online resources.

Table 7. Results obtained on testing the system

Sr. No.	Corpus Type	Total Number of participial sentences in the corpus	Correctly simplified sentences	In-correctly Simplified sentences	Not Simplified	Precision	Recall	F-measure
1	ILCI Corpus	760	729	22	9	98.81579	95.92105	97.34691
2	e-papers corpus	1190	989	110	91	92.35294	83.10924	87.4876
3	Online corpus	3429	3267	128	34	99.00846	95.27559	97.10616
4	Manually created corpus	500	456	21	23	95.4	91.2	93.25273
	Overall	5879	5441	281	157	96.3943	91.37647	93.79835

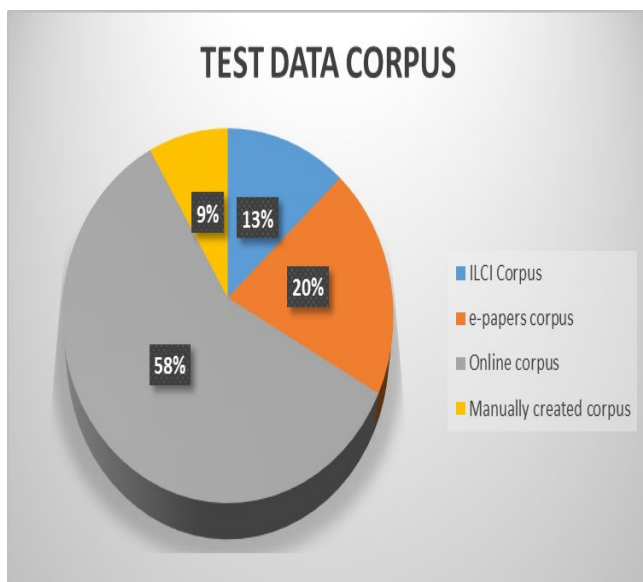


Figure 6. Details of test corpus used for testing

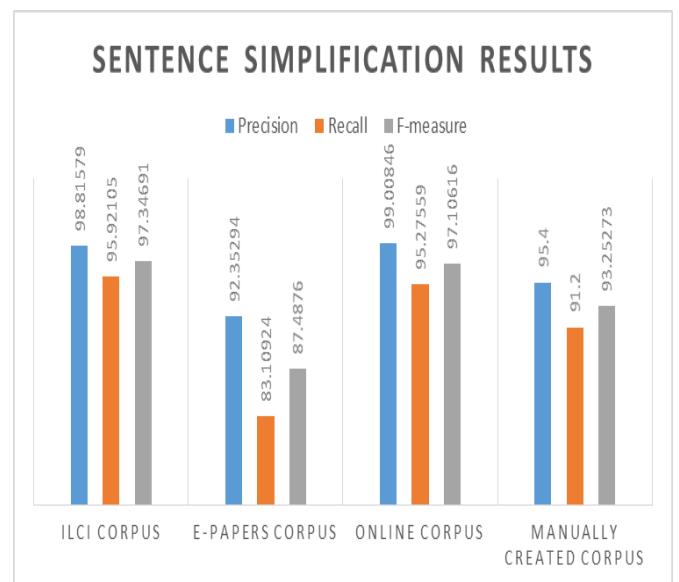


Figure 7. Sentence Simplification result

7. Comparison with the existing systems

As per the literature reviewed no such system has been developed for Punjabi language. Although some work has been done for foreign languages. Therefore the developed system can be compared only partially with the existing systems. [37] used numerical information present in the large complex sentences to simplify the text and on testing, system gave a precision of 0.94, a recall of 0.93 and an F-measure of 0.93. [38] used Corpus-based Sentence Deletion and Split Decisions for Spanish Text

Simplification and showed overall F-measure up to 0.92. [39] described an ongoing research project on text simplification for Japanese language using SVMbased classifier and obtained a precision 95% and recall 89%. [40] used parsing technology for syntactic simplification of English sentences with precision at 0.92 and recall at 0.95.

8. Conclusion and future scope

In this research, author proposed a novel approach for simplification of participial type sentences in Punjabi language. Author performed lexical and syntactical simplification techniques. For syntactic simplification, author used morphological features along with part of speech (POS) tagging. Further clause boundary information is used to identify dependent and independent clauses and then dependent clause is converted into independent clauses in all possible cases. On testing the system, author claimed a precision of 96.39%, recall of 91.37% and F-measure as 93.79%. During this research work large corpus of Punjabi sentences has been created for testing the developed system, therefore the same corpus can be used to develop and test other similar applications i.e. to simplify other categories of complex sentences. In future the work can be extended for other variants of complex sentences like conjunctival type and infinitival type complex sentences. Also in future, statistical and machine learning approaches can be implemented for this task. Further the technique proposed in this paper can be further implemented in other Indo-Aryan languages having same sentence structure as that of Punjabi language.

References

- [1]. Johannes C. Ziegler, Conrad Perry, Anna Ma-Wyatt, Diana Ladner, and Gerd Schulte-Körne. 2003. Developmental dyslexia in different languages: Language-specific or universal? *Journal of experimental child psychology*, 86(3):169–193.
- [2]. Núria Gala and Johannes Ziegler. 2016. Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC), COLING 2014*, pages 59–66.
- [3]. Singh, Ravinder, et al. "A Framework for Early Detection of Antisocial Behavior on Twitter Using Natural Language Processing." *Conference on Complex, Intelligent, and Software Intensive Systems*. Springer, Cham, 2019.
- [4]. Subramani, Sudha, et al. "Deep Learning for Multi-Class Identification From Domestic Violence Online Posts." *IEEE Access* 7 (2019): 46210-46224.
- [5]. Sharma, Manik, Gurvinder Singh, and Rajinder Singh. "Design of GA and Ontology based NLP Frameworks for Online Opinion Mining." *Recent Patents on Engineering* 13.2 (2019): 159-165.
- [6]. Sharma, Manik, Gurvinder Singh, and Rajinder Singh. 2018. "An Advanced Conceptual Diagnostic Healthcare Framework for Diabetes and Cardiovascular Disorders." *EAI Endorsed Transactions on Scalable Information Systems*. 5.18:1-11.
- [7]. Kaur, Prableen, and Manik Sharma. "A survey on using nature inspired computing for fatal disease diagnosis." *International Journal of Information System Modeling and Design (IJISMD)*8(2) (2017): pp. 70-91.
- [8]. Kaur, Prableen, and Manik Sharma. "Diagnosis of Human Psychological Disorders using Supervised Learning and Nature-Inspired Computing Techniques: A Meta-Analysis." *Journal of medical systems* 43.7 (2019): 204.
- [9]. Zheng, Hui, et al. "Dynamic optimisation based fuzzy association rule mining method." *International Journal of Machine Learning and Cybernetics* 10.8 (2019): 2187-2198.
- [10]. https://en.wikipedia.org/wiki/Punjabi_dialects (Retrieved on 1.11.2019)
- [11]. <http://www.learnpunjabi.org/gtrans/index.asp>
- [12]. Blum, S., Levenston, E.A.: *Universals of lexical simplification*. *Lang. Learn.* 28(2), 399–415 (1978)
- [13]. Hoard, J.E., Wojcik, R., Holzhauser, K.: *An automated grammar and style checker for writers of simplified English*. In: O'Brian Holt, P., William, N. (eds.) *Computers and Writing*, pp. 278–296. Springer, Dordrecht (1992).
- [14]. Coster, W., Kauchak, D.: *Learning to simplify sentences using Wikipedia*. In: *Proceedings of the Workshop on Monolingual Text-To-Text Generation*. Association for Computational Linguistics, Portland, June 2011, pp. 1–9 (2011)
- [15]. Zhu, Z., Bernhard, D., Gurevych, I.: *A monolingual tree-based translation model for sentence simplification*. In: *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 1353–1361 (2010)
- [16]. Wubben, S., van den Bosch, A., Krahmer, E.: *Sentence simplification by monolingual machine translation*. In: *Proceedings of the 50th Annual*
- [17]. Woodsend, K., Lapata, M.: *Learning to simplify sentences with quasi-synchronous grammar and integer programming*. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 409–420. ACL, Stroudsburg (2011)
- [18]. Klebanov, B.B., Knight, K., Marcu, D.: *Text simplification for information-seeking applications*. *LNCS*, vol. 3290, pp. 735–747. Springer, Heidelberg (2004).
- [19]. Scarton, C., de Oliveira, M., Candido Jr., A., Gasperin, C., Aluísio, S.M.: *Simplifica: a tool for authoring simplified texts in Brazilian Portuguese guided by readability assessments*. In: *Proceedings of the NAACL HLT 2010 Demonstration Session*, pp. 41–44. ACL, Stroudsburg (2010)
- [20]. Seretan, V.: *Acquisition of syntactic simplification rules for French*. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. ELRA, Istanbul, May 2012
- [21]. Hung, B.T., Minh, N.L., Shimazu, A.: *Sentence splitting for Vietnamese-English machine translation*. In: *2012 Fourth International Conference on Knowledge and Systems Engineering (KSE)*, pp. 156–160, August 2012
- [22]. Aranzabe, M.J., de Ilarraza, A.D., Gonzalez-Dios, I.: *Transforming complex sentences using dependency trees for automatic text simplification in Basque*. , pp. 61–68 (2012)
- [23]. Barlacchi, G., Tonelli, S.: *ERNESTA: a sentence simplification tool for children's stories in Italian*. In: Gelbukh, A. (ed.) *CICLing 2013*. *LNCS*, vol. 7817, pp. 476–487. Springer, Heidelberg (2013).
- [24]. Chung, J.-W., Min, H.-J., Kim, J., Park, J. C.: *Enhancing readability of web documents by text augmentation for deaf people*. In: *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics*, pp. 30:1–30:10. ACM, New York (2013)
- [25]. Štajner, S., Drndarević, B., Saggion, H.: *Corpus-based sentence deletion and split decisions for Spanish text simplification*. *Revista Computación y Sistemas* 17(2) (2013)
- [26]. Biran, O., Brody, S., Elhadad, N.: *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. ACL, Stroudsburg, pp. 496–501 (2011)
- [27]. <http://sanskrit.jnu.ac.in/ilci/index.jsp> (Retrieved on 04.11.2019)

- [28]. Inui, K., Fujita, A., Takahashi, T., Iida, R., Iwakura, T.: Text simplification for reading assistance: a project note. In: Proceedings of the Second International Workshop on Paraphrasing, pp. 9–16. ACL, Sapporo, July 2003
- [29]. Bott, S., Rello, L., Drndarević, B., Saggion, H.: Can Spanish be simpler? LexSiS: lexical simplification for Spanish. In: Proceedings of COLING 2012. Mumbai, India, pp. 357–374, December 2012
- [30]. Chandrasekar, R., Srinivas, B.: Automatic induction of rules for text simplification. Knowledge Based System 10(3), pp. 183–190 (1997)
- [31]. Siddharthan, A.: Syntactic simplification and text cohesion. Res. Lang. Comput. 4, pp. 77–109 (2006)
- [32]. Kandula, S., Curtis, D., Zeng-Treitler, Q.: A semantic and syntactic text simplification tool for health content. In: AMIA Annual Symposium Proceedings, pp. 366–370. (2010)
- [33]. Silveira, S., Branco, A.: Combining a double clustering approach with sentence simplification to produce highly informative multi- document summaries. In: 2012 IEEE 13th International Conference on Information Reuse and Integration (IRI), pp. 482–489, August 2012
- [34]. Carroll, J., Minnen, G., Canning, Y., Devlin, S., Tait, J.: Practical simplification of English newspaper text to assist aphasic readers. In: Proceedings of AAI-1998 Workshop on Integrating Artificial Intelligence and Assistive Technology, pp. 7–10 (1998)
- [35]. Feblowitz, D., Kauchak, D.: Sentence simplification as tree transduction. In: Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations, pp. 1–10. ACL, Sofia, August 2013
- [36]. Klerke, S., Søgaard, A.: Simple, readable sub-sentences. In: 51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop, pp. 142–149. ACL, Sofia, August 2013
- [37]. Bautista, Susana, et al. "An approach to treat numerical information in the text simplification process." Universal Access in the Information Society 16.1 (2017): 85-102.
- [38]. Štajner, Sanja, Biljana Drndarevic, and Horacio Saggion. "Corpus-based sentence deletion and split decisions for Spanish text simplification." Computación y Sistemas 17.2 (2013).
- [39]. Inui, Kentaro, et al. "Text simplification for reading assistance: a project note." Proceedings of the second international workshop on Paraphrasing-Volume 16. Association for Computational Linguistics, 2003.
- [40]. Lee, John, and J. Buddhika K. Pathirage Don. "Splitting Complex English Sentences." Proceedings of the 15th International Conference on Parsing Technologies. 2017.