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

R.M. Jindal¹,*, 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 (participle 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 participle 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.

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 sub-clauses, converting the sentence into passive voice or by resolving anaphora. Many researchers used it 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 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], Basque 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:

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Punjabi Sentence</th>
<th>Roman Transliterated sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ਫਿਲ੍ਹ ਕਰ ਦੇਣ ਦੀ ਵਿਚਾਰ ਦੀ ਨੱਕੀ ਨੋਂ ਮੈਨੁੰਤ।   (binām bhakāī kītāī ēs nē na hi jē māmnā)</td>
<td>He didn't do it without being bothered.</td>
</tr>
<tr>
<td>2</td>
<td>ਫਿਲ੍ਹ ਕਰ ਦੇਣ ਦੀ ਵਿਚਾਰ ਦੀ ਨੱਕੀ ਨੋਂ ਮੈਨੁੰਤ।   (dillī jāndīām uh rāh vīc jākhal utar ē)</td>
<td>On his way to Delhi, he landed on the road.</td>
</tr>
<tr>
<td>3</td>
<td>ਪੁਲਸ ਲੈ ਵੇਦੀਂ ਮੌਚ ਵੇਦੀਂ ਹੋ ਜਾਂਦਾ ਹਾਲ ਮਨੂਕ।   (pulas  ਲੇ ਪਚਿਟਾਂ ਦੀ ਦੇਵੇ ਦਿੱਪ ਹੋ ਜਾਂਦਾ)</td>
<td>The flame went cold as the Prahlad's fire burst.</td>
</tr>
<tr>
<td>4</td>
<td>ਮੁੰਰੂ ਲੈ ਦੀਲਣ ਅੱਠ ਚਾਰਦਾਨ ਨਚਣਾ ਉੱਤਰ ਮਨੂਕ।   (sūraj vall vēkhiā akkhāṃ cundhīā jāndīām hā)</td>
<td>Looking at the sun, eyes are wiped.</td>
</tr>
<tr>
<td>5</td>
<td>ਏਕੜੀ ਦੇ ਕਲੀਫਾਕਾ ਦੇ ਉੱਤਰ ਨੋਂ ਮੈਨੁੰਤ।   (ghōṛī dē khalōndiāṃ ēh utar ē)</td>
<td>He got down just as the horse was down.</td>
</tr>
</tbody>
</table>

In above example, ਫਿਲ੍ਹ ਕਰ ਦੇਣ (rassī ṭappdiām) is the predicate (dependent clause) and the word ਵਿਚਾਰ (ṭappdiām) is responsible for binding it to in-dependent clause i.e. ਮੈਨੁੰਤ। (mērē pair núm môc ā ē).
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](image)

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 ਜਦਆ (Diā), and ਇਆ (iā) respectively as suffix in the verb. Consider the following examples:

**Sentence 1:** ਸਕੂਲ ਜਦਆਂ ਉਹ ਰਾਹ ਵਿੱਚ ਖੇਡਣ ਲਗਿਆ (sakūl jāndiāṁ uh rāh vic khēḍāṁ lag piā)

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

**Sentence 2:** ਬਿਨਾਮ ਭਕਾਈ ਕੀਤਆਂ ਉੱਤੇ ਨਹੀ ਜੇ ਮੰਣਾ। (bināṃ bhakāī kītiāṁ ēsnē nahī jē mannṇā.) He will not agree unless you try hard to convince him.

In the above examples, sentence 1 has present perfect non-finite verb ਜਦਆਂ (jāndiā) with suffix ਜਦਆ and sentence 2 has past perfect non-finite verb ਕੀਤਆਂ (kītiā) having suffix ਕੀਤਆ. 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:

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

<table>
<thead>
<tr>
<th>Table 2. Features of participial non-finite verb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Features of participial non-finite verb</strong></td>
</tr>
<tr>
<td><strong>Word related feature</strong></td>
</tr>
<tr>
<td><strong>POS tag related features(suffix used in the tag)</strong></td>
</tr>
</tbody>
</table>

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←Ø
2 S←{s₁}
3 s₀←pop(S)
4 IDb←s₀
5 1←i
6 While S≠ Ø do the following
7   sᵢ←pop(S)
8   if sᵢ is participial finite verb as listed in table 2
9     sᵢ← IDb
10    sᵢ⁺₁← IDb
11    A←AU{ sᵢ }
12   else
13     A←AU{ sᵢ }
14  end
15  sᵢ← IDb
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:**

- IDb: Beginning of independent clause
- IDb: 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: "ਟੱਪਿਦਆਂ ਮੇਰੇ ਪੈਰ ਨੂੰ ਮੋਚ ਆਗਈ।" and in this clause ਟੱਪਿਦਆਂ 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.

<table>
<thead>
<tr>
<th>Input Sentence</th>
<th>Output Simplified sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>ਟੱਪਿਦਆਂ ਮੇਰੇ ਪੈਰ ਨੂੰ ਮੋਚ ਆਗਈ।</td>
<td>ਟੱਪਿਦਆਂ has ਦਿਅਨਾ as suffix</td>
</tr>
<tr>
<td>ਟੱਪਿਦਆਂ has ਦਿਅਨਾ as suffix</td>
<td>ਟੱਪਿਦਆਂ has ਦਿਅਨਾ as suffix</td>
</tr>
</tbody>
</table>

*Example: ਟੱਪਿਦਆਂ -> ਟੱਪ (Root Word) + ਦਿਅਨਾ (Suffix part)*

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

```plaintext
Input Sentence          | Output Simplified sentences          |
------------------------|--------------------------------------|
ਟੱਪਿਦਆਂ ਮੇਰੇ ਪੈਰ ਨੂੰ ਮੋਚ ਆਗਈ। | ਟੱਪ ਦਿਅਨਾ ਆਗਈ।                   
```
4.3 Content reduction

In this phase unnecessary 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 ਿਰਹਾ ਸੀ। ਉਹ ਘਰ ਜਦੀਆਂ ਹੀ ਸੋਨ ਚਲਾ ਿਗਆ।

As in above sentence, the dependent clause ਉਹ ਘਰ ਜਦੀਆਂ ਹੀ can be discarded as the information provided by this dependent clause is already given by the previous sentence i.e. ਮੁੰਡਾ ਘਰ ਜਾ ਿਰਹਾ 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>
<thead>
<tr>
<th>File name</th>
<th>Number of files</th>
<th>Total number of sentences</th>
<th>Number of Participial type sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pun_agriculture.txt</td>
<td>11</td>
<td>11000</td>
<td>234</td>
</tr>
<tr>
<td>Pun_entertainment.txt</td>
<td>03</td>
<td>3000</td>
<td>101</td>
</tr>
<tr>
<td>Pun_health.txt</td>
<td>01</td>
<td>1000</td>
<td>37</td>
</tr>
<tr>
<td>Pun_literature.txt</td>
<td>04</td>
<td>4000</td>
<td>112</td>
</tr>
<tr>
<td>Pun_religion.txt</td>
<td>05</td>
<td>5000</td>
<td>109</td>
</tr>
<tr>
<td>Pun_science_and_technology.txt</td>
<td>02</td>
<td>2000</td>
<td>29</td>
</tr>
<tr>
<td>Pun_sports.txt</td>
<td>05</td>
<td>5000</td>
<td>138</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>31000</td>
<td>760</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Name of newspaper</th>
<th>Website Link</th>
<th>Total no. of sentences collected</th>
<th>No. of Participial type sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Punjabi Tribune</td>
<td><a href="https://epaper.punjabitribuneonline.com">https://epaper.punjabitribuneonline.com</a></td>
<td>1000</td>
<td>124</td>
</tr>
<tr>
<td>2</td>
<td>Punjabi Jagran</td>
<td><a href="https://epaper.punjabijagran.com">https://epaper.punjabijagran.com</a></td>
<td>1000</td>
<td>78</td>
</tr>
<tr>
<td>3</td>
<td>Jag bani</td>
<td><a href="https://jagbani.epapr.in">https://jagbani.epapr.in</a></td>
<td>1000</td>
<td>113</td>
</tr>
<tr>
<td>4</td>
<td>Daily Punjab Times</td>
<td>epaper.dailypunjabtimes.com</td>
<td>1000</td>
<td>145</td>
</tr>
<tr>
<td>5</td>
<td>Ajit Jalandhar</td>
<td><a href="https://www.ajitjalandhar.com">https://www.ajitjalandhar.com</a></td>
<td>1000</td>
<td>94</td>
</tr>
<tr>
<td>6</td>
<td>Daily Aashiana</td>
<td><a href="https://epaper.dailyaashiana.com">https://epaper.dailyaashiana.com</a></td>
<td>1000</td>
<td>38</td>
</tr>
<tr>
<td>7</td>
<td>Anonymous Newspaper</td>
<td><a href="https://www.readwhere.com">https://www.readwhere.com</a></td>
<td>1000</td>
<td>64</td>
</tr>
<tr>
<td>8</td>
<td>The Times of Punjab</td>
<td>epaper.thetimeofpunjab.com</td>
<td>1000</td>
<td>218</td>
</tr>
<tr>
<td>9</td>
<td>Newspaper</td>
<td><a href="https://www.3newspapers.com">https://www.3newspapers.com</a></td>
<td>1000</td>
<td>223</td>
</tr>
<tr>
<td>10</td>
<td>Parvasi Newspaper</td>
<td><a href="https://parvasi">https://parvasi</a> newspaper.com</td>
<td>1000</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Total Sentences</td>
<td></td>
<td>10000</td>
<td>1190</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Type of corpus</th>
<th>Website Link</th>
<th>Total number of sentences collected</th>
<th>Number of Participial type sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>stories</td>
<td><a href="https://punjabistories.com">https://punjabistories.com</a></td>
<td>10000</td>
<td>1541</td>
</tr>
<tr>
<td><a href="http://www.punjabikahani.punjabi-kavita.com">www.punjabikahani.punjabi-kavita.com</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.shurli.com">www.shurli.com</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="https://www.absolutestudy.com">https://www.absolutestudy.com</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="https://sikhville.org">https://sikhville.org</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>essay</td>
<td><a href="https://www.sbs.com.au">https://www.sbs.com.au</a></td>
<td>10000</td>
<td>678</td>
</tr>
<tr>
<td><a href="http://www.shurli.com">www.shurli.com</a></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Type of corpus</th>
<th>Number of Participial type sentences created manually</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test corpus</td>
<td>500</td>
</tr>
</tbody>
</table>

Table 6. Manually created test data

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.
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 SVM-based 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


