

IDENTIFICATION OF COMPOUND SENTENCES IN PUNJABI LANGUAGE

¹Sanjeev Kumar Sharma, ²G S Lehal

¹Assistant Professor, ²Professor

¹B.I.S. Engineering College, Kot Ise Khan, Moga

²Department of Computer Science, Punjabi University, Patiala.

Abstract

Compound sentences constitute major parts of the Punjabi language. All the large sentences are either of compound or of complex type. Detail analysis of compound sentences is helpful in processing the Punjabi language through computer. This study will be helpful in identifying and separating the compound sentences from Punjabi corpus. Also this study will be helpful in developing other NLP applications like converting a compound sentence in simple sentences, grammar checking of compound sentences, summarization and machine translation etc.

1 INTRODUCTION

A **compound sentence** is composed of at least two independent clauses. It does not require a dependent clause. The clauses are joined by a coordinating conjunction (with or without a comma), a correlative conjunction (with or without a comma), a semicolon that functions as a conjunction, a colon instead of a semicolon between two sentences when the second sentence explains or illustrates the first sentence and no coordinating conjunction is being used to connect the sentences, or a conjunctive adverb preceded by a semicolon. A conjunction can be used to make a compound sentence. Conjunctions are words such as for, and, nor, but, or, yet, so. The structure of compound sentences is symmetrical. This structure is composed of two or more than two independent clauses. These independent clauses are composed joined by co-ordinate conjunctions.

ਮੀਂਚ ਪੈ ਰਿਹਾ ਸੀ ਤੇ ਲੋਕ ਭਿੱਜ ਰਹੇ ਸਨ ।

In the above example, ਮੀਂਚ ਪੈ ਰਿਹਾ ਸੀ and ਲੋਕ ਭਿੱਜ ਰਹੇ ਸਨ are two independent clauses and ਤੇ is the co-ordinate conjunction that joins these two independent clause.

2 OVERVIEW OF PUNJABI LANGUAGE

Punjabi language is a member of the Indo-Aryan family of languages, also known as Indic languages. Other members of this family are Hindi, Bengali, Gujarati, and Marathi etc. Indo-Aryan languages form a subgroup of the Indo-Iranian group of languages, which in turn belongs to Indo-European family of languages. Punjabi is spoken in India, Pakistan, USA, Canada, England, and other countries with Punjabi immigrants. It is the official language of the state of Punjab in India. Punjabi is written in „Gurmukhi“ script in eastern Punjab (India), and in „Shahmukhi“ script in western Punjab (Pakistan).

3 PATTERNS OF COMPOUND SENTENCES

On the basis of method for joining the independent clauses using co-ordinate conjunctions following patterns have been designed:-

Pattern 1:-

Independent Clause + / , / +Independent Clause

In this type of compound sentences two independent clauses are joined by using comma(,).

/// ਆਦਮੀਂ ਕੋਈ ਖਿਡਾਉਣਾ ਤਾਂ ਨਹੀਂ, // ਕੁਦਰਤ ਕੋਈ ਮੰਤਕ ਦੀ ਕਿਤਾਬ ਤਾਂ ਨਹੀਂ ।///

Pattern 2:-

Independent Clause + / Conjunction / +Independent Clause

In this type of compound sentences two independent clauses are joined by using co-ordinate conjunction.

ਪਹਾੜਾਂ ਵਿਚ ਮੀਂਹ ਪਹਿਲਾਂ ਵੀ ਤਿੰਨ ਮਹੀਨੇ ਹੀ ਵਰ੍ਹਦਾ ਸੀ // ਪਰ ਪਾਣੀ ਸਾਰਾ ਸਾਲ ਵਗਦਾ ਸੀ ।///

Pattern 3:-

Independent Clause + / , /+ Independent Clause +/ Conjunction / +Independent Clause

These types of compound sentences are composed of three Independent clauses. First two are joined by comma and the third one is joined by using co-ordinate conjunction.

/// ਆਂਢੀ ਗੁਆਂਢੀ ਹੈਰਾਨ ਸਨ, / ਉਸ ਦੇ ਬੱਚਿਆਂ ਦਾ ਪਿਓ ਤੁਰ ਗਿਆ ਸੀ // ਤੇ ਬੱਚਿਆਂ ਦੀ ਮਾਂ ਨੇ ਕਦੀ ਇਕ ਵਾਰ ਉਸ ਨੂੰ ਯਾਦ ਨਹੀਂ ਸੀ ਕੀਤਾ ।///

Pattern 4:-

Independent Clause + / , /+ Independent Clause + / , /+ Independent Clause+ / , /+ Independent Clause

These types of compound sentences are composed of four Independent clauses. All these clauses are joined by comma.

/// ਇਕ ਦੇਸਤ ਹੀ ਕਾਫੀ ਹੁੰਦਾ ਹੈ, // ਦੋ ਦੇਸਤਾਂ ਜਿਹੀ ਕੋਈ ਰੀਸ ਨਹੀਂ, // ਤਿੰਨ ਦੇਸਤ ਕਰਮਾਂ ਵਾਲਿਆਂ ਦੇ ਹੁੰਦੇ ਹਨ, // ਚਾਰ ਦੇਸਤ ਸੰਭਵ ਨਹੀਂ ।///

Pattern 5:-

Independent Clause + / Conjunction / +Independent Clause+ / Conjunction / +Independent Clause

These types of compound sentences are composed of three Independent clauses. All these clauses are joined by co-ordinate conjunctions.

/// ਪਾਣੀ ਜੀਵਨ ਦਾ ਸ੍ਰੋਤ ਹੈ // ਅਤੇ ਇਸਤਰੀ ਜੀਵਨ ਦਾ ਸਾਧਨ ਹੈ । // ਸੇ ਇਸਤਰੀ ਨਾਲ ਪ੍ਰਕਿਰਤੀ ਦਾ ਸੰਬੰਧ ਮਾਂ ਧੀ ਵਾਲਾ ਹੈ । ///

Pattern 6:-

Independent Clause + / , / +Independent Clause+ / Conjunction / +Independent Clause+ / , / +Independent Clause

These types of compound sentences are composed of four Independent clauses. First two and last two clauses are joined by comma. Middle two clauses are joined by co-ordinate conjunction.

/// ਉਸ ਦੇ ਕਾਫੀ ਚੇਲੇ ਸਨ ਸਜੇ, // ਉਹ ਵੀ ਸ਼ਰਧਾ ਭਰਪੂਰ ਹੋ ਗਏ ਸਨ // ਪਰ ਸਿੰਘ ਨਹੀਂ ਸਨ ਸਜੇ, // ਮੁਖ ਚੇਲੇ ਨੂੰ 'ਮੁਰੀਦ' ਜੀ ਕਹਿੰਦੇ ਸਨ । ///

Pattern 7:-

Independent Clause + / , / +Independent Clause+ / , / +Independent Clause+ / Conjunction / +Independent Clause

These types of compound sentences are composed of four Independent clauses. First three clauses are joined by comma. The last clause is joined by co-ordinate conjunction.

/// ਪਾਣੀ ਦੀ ਬੁੜ੍ਹ ਨਾਲ ਹਿੰਸਾ ਵਧਦੀ ਹੈ, // ਇਕ ਦੂਜੇ ਉਤੇ ਵਿਸ਼ਵਾਸ ਦੀ ਭਾਵਨਾ ਘਟਦੀ ਹੈ, // ਅਤੇ ਸਾਂਝ ਤੇ ਮਿਲਵਰਤਨ ਦੀ ਬਿਰਤੀ ਮਾਂਦ ਪੈ ਜਾਂਦੀ ਹੈ । ///

4 CHARACTERISTICS OF COMPOUND SENTENCES

1. The compound sentences of Punjabi language have the advantage over the complex sentences that the conjunction used in the compound sentences never comes in the start of the sentence.
2. In compound sentence all the clauses except the first clause lacks one element (subject, object etc).
3. There is no limit for the length of compound sentences. The length of these sentences can be increased whenever desired.
4. The compound sentences are composed of only independent clause whereas the complex sentences contain at least one dependent clause along with independent clause.
5. In compound sentences co-ordinate conjunction is used to join two in-dependent clauses where as in complex sentences sub-ordinate conjunction is used for this purpose.

5 STRUCTURE OF PUNJABI SENTENCE

Punjabi sentence follow SOV (Subject-Object-Verb) order. In Punjabi sentences, the subject occurs first followed by the object and then the verb. Punjabi sentences can be categorized in to three types. These are simple sentence, compound sentence and complex sentences. A sentence is further composed of clauses which can be further classified as in-dependent clause and dependent clause. An independent clause can constitute a simple sentence on its own. Every sentence contains independent clause as a basic element. The independent clause contains a finite verb phrase as an essential element

6 CONJUNCTIOS USED IN COMPOUND SENTENCES: Mainly coordinate conjunctions are used for the construction of compound sentences. These includes:- ਤੇ, ਪਰ, ਅਤੇ, ਫੇਰ,

7 STRUCTURE OF COMPOUND SENTENCE

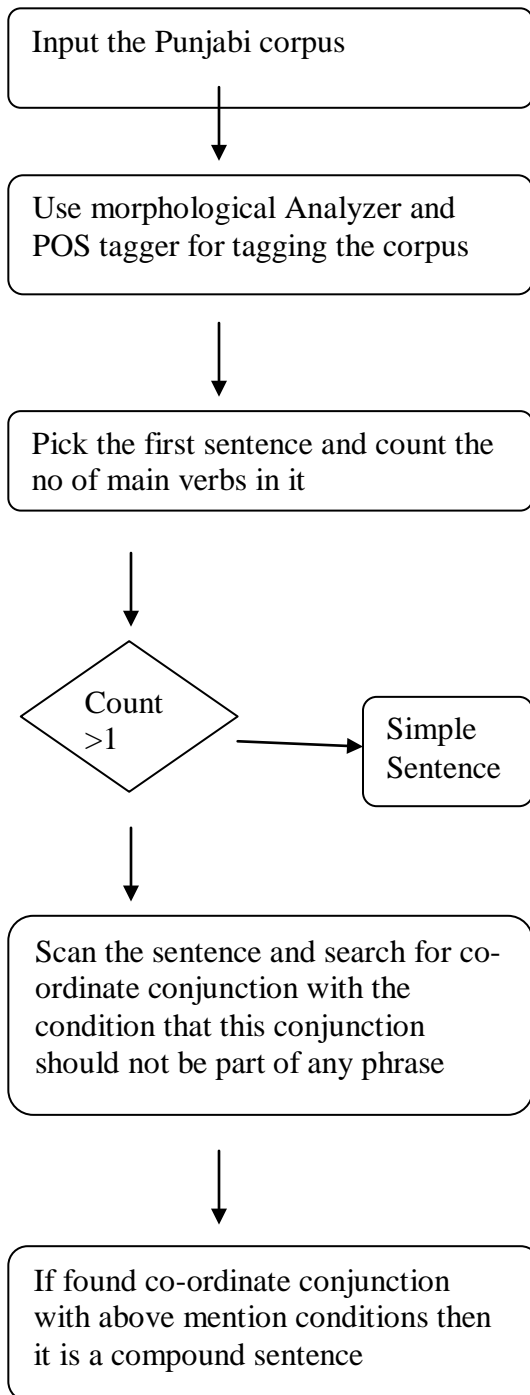
In Punjabi language compound sentences have simple structure. These are composed of independent clauses and conjunctions. Two or more than two independent clauses are joined by using either conjunction or comma. The conjunctions used in the construction of compound sentences are mainly coordinate conjunctions. These conjunctions join two symmetric parts (independent clauses) of the compound sentences.

Independent Clause: - The independent clause can be defined as an independent grammatical unit that is congruent with the sentence. A finite verbal phrase is further of two types. First one is affirmative verbal phrase that is composed of one to five verbal forms (Main verb + primary operator + progressive operator + modal operator + auxiliary verb). The second form is negative and emphatic verbal phrase which contains one to seven forms. This is essential element of the independent clause. This finite verbal phrase occurs finally in the clause. On the basis of structure an independent clause can have more than one noun phrase, adjective phrase, adverb phrase, prepositional phrase etc but it cannot have more than one verb phrase. These independent clauses can occur at any position in the compound and complex sentences.

8 ALGOIRITHM USED FOR IDENTIFICATION OF COMPOUND SENTENCES

The compound sentences can be identified by identifying the type of conjunctions present in the sentence. In Punjabi mainly co-ordinate conjunctions are used for the construction of compound sentences. Now in general the co-ordinate conjunction are used for joining two symmetric parts of the sentence so while checking for the compound sentences care

has to be taken that the conjunction present in the sentence should not be the part of the phrase.



9 RESULTS AND DISCUSSION

We tested our module on Punjabi corpus randomly picked from the internet. We take two samples from different sites. One sample is given name set A and the second Sample given name set B.

Test set	Size (No of sentences)	Accuracy
A	2400	85%
B	3100	88%

10 CONCLUSIONS AND FUTURE WORK

In this study, we made a detailed analysis of compound sentences and observed that compound sentences have fixed pattern that is composed of independent clauses separated by conjunctions or commas. This study could be helpful in computational linguistic for identification of compound sentences in Punjabi language. Further this could be helpful for the grammar checking of compound sentences. Another utilization of such study can be used to differentiate the compound sentences from complex sentences.

References

- [1].N. UzZaman and J. F. Allen, "TRIPS and TRIOS System for TempEval-2: Extracting Temporal Information from Text," International Workshop on Semantic Evaluations, ACL 2010.
- [2].J. Pustejovsky and M. Verhagen, "SemEval-2010 task 13: evaluating events, time expressions, and temporal relations (TempEval-2)," Workshop on Semantic Evaluations:Recent Achievements and Future Directions, 2010.
- [3].Poornima C, Dhanalakshmi V, Anand Kumar M and Soman K P (2011)' Rule based Sentence Simplification for English to Tamil Machine Translation System', International Journal of Computer Applications (0975 – 8887)Volume 25– No.8
- [4].Katsuhito Sudoh et al. 2010. "Divide and Translate:Improving Long Distance Reordering in Statistical Machine translation".

- [5].Zheming Zhu, Delphine Bernhard and Iryna Gurevych 2010. “A Monolingual Tree-based Translation Model for Sentence Simplification”, Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010).
- [6].Akshar Bharati, Dipti Misra Sharma, Sukhada (2009) Adapting Link Grammar Parser (LGP) to Paninian Framework Mapping of Parser Relations for Indian Languages’, *National Seminar on Computer Science and its Applications in Traditional Shastras (CSATS'09)*.
- [7]. ਡ.ਬਲਦੇਵ ਸੰਘ ਚੇ ਮ, 2005. ਪੰਜ ਬੀ ਵਕ ਪ੍ਰਬੰਧ (ਥਣਤਰ ਅਤੇ ਕ ਰਜ) , ਪਬਲੀਕੇਸ਼ਨ ਬੰਊਰੋ ਪੰਜ ਬੀ ਯੂਨੀ ਵਰਸਟੀ ਪਟਿਆਲਾ ।
- [8].ਬੂਟ ਸੰਘ ਬਰ ਰ,2008. ਪੰਜ ਬੀ ਵੀਆਕਰਨ (ਸਥਿ ਤ ਅਤੇ ਵੀਹ ਰ) , ਪਬਲੀਕੇਸ਼ਨ ਬੰਊਰੋ ਪੰਜ ਬੀ ਯੂਨੀ ਵਰਸਟੀ ਪਟਿਆਲਾ ।