

Ambiguity in Chhattisgarhi language – Introduction and Prevention

¹Pragya Bhagat, ²Dr.Chandrakant S. Ragit

¹*Department of information and language Engineering, Mahatma Gandhi Antarrashtriya Hindi Vishwavidyalaya, Wardha , Maharashtra, India*

²*Pro-vice Chancellor and Director, Department of information and language Engineering, Mahatma Gandhi Antarrashtriya Hindi Vishwavidyalaya, Wardha , Maharashtra, India*

Abstract: We use sentences to express our sentiments. Sentences need to be grammatical in order to communicate. These grammatical sentences address the language. The use of words present in sentences to represent different situations is called ambiguity of the word. This research paper highlights the ambiguity and redressal of the words of natural language (Chhattisgarhi). In this research paper, we have compiled and analyzed data from various fields of Chhattisgarh language to understand the ambiguity present in Chhattisgarhi language and then highlighted their redressal. The main objective of this paper is to make aware of the ambiguity present in the Chhattisgarhi language and the rule based method used for its solution. This system is effective only for the prevention of word level ambiguity.

Index Terms – ambiguity, Chhattisgarhi language, contextual rules, knowledge base

I.INTRODUCTION

All organisms express their dialogue through language. But the language used in dialog observation itself reveals many features. One of the different characteristics of the language is ambiguity. It is because of this characteristic of language that sometimes dialogue is also responsible for wrong transmission due to the multi meaning of the word present in the language. This feature of language is found in every natural language.

In general terms, the word represents ambiguity when it reveals more than one meaning. This ambiguity appears both in writing and at the linguistic level.

Under NLP, we study and preserve all natural languages by computer, and the work that takes longer than human beings do is easily done by the computer, one of which is linguistic ambiguity. In linguistic ambiguity, the word represents more than one meaning. We have divided this research paper into 5 parts. First is introduction, under which we have introduced the subject and language. In the second part, the purpose of this research is explained. In the third part, the procedure of this research, data storage and its knowledge-based prevention, algorithm, contextual rules are presented. The result in the fourth part and the conclusion in the fifth part are presented.

1.1 Language Introduction

The State of Chhattisgarh is a tribal dominated area of the country which is transmitted simultaneously by Hindi and Chhattisgarhi in the linguistic scenario. In 1895, George Grierson's survey of India's languages and dialects ranked Chhattisgarhi in section 6 out of 21 volumes. Due to geographical location and caste differences, there is a spatial difference in the nature of Chhattisgarhi as in other languages. Chhattisgarhi has been divided into 5 categories based on the study of scholarly articles and experience gained from visits to different regions of Chhattisgarh.

1. North Chhattisgarh (Sargujia) - This dialect is spoken in raigad jashpur korea and northern part of Sarguja district.
2. East Chhattisgarh (Laria) - It is used on the eastern frontier of Raigad Mahasamund and Raipur districts.
3. South Chhattisgarh (Bastaria) - It is turmeric and gondi dialect spoken in Bastar and Dantewada districts.
4. West Chhattisgarh (Khaltahi) - It is used in the western part of Kawardha and Rajnandgaon districts and in Bilaspur district.
5. Central Chhattisgarh- It is used in Janjgir Durg Dhamtari and Kanker districts. It is free from external effects other than Hindi.

II. OBJECTIVE OF THE STUDY

Chhattisgarhi language is a morphologically rich language. The main objective of this research is to develop a system for the prevention of language ambiguity keeping in view the use of morphological characteristics of language. When

a word displays different positions in different sentences, the word falls into the category of ambiguity that leads to linguistic ambiguity. Linguistic ambiguity represents a state in which it is difficult to understand the true meaning of the word present in a text. The reason for ambiguity is to display more than one meaning of the word. Chhattisgarhi language is also not far from ambiguity. Chhattisgarhi, like other languages, suffers from ambiguity.

आषढ	म	एक	पानी	के	गिरे	ले	सबो	झन	धान	ल	बो	डरीस
NN	PP	QTC	NN_M_S	PP	VM	PP	JJ	NN	NN	PP	VM	VM_S_3_PST
						PT		PT				
						VM						

Example “Due to the fall of water in ashadh, all the people sowed paddy”. In the given example, the words (ले) and (झन) are expressing ambiguity.(ले) (from/take/by) PP(preposition)/VM(main verb)/PT(particles) in the sentence is representing the meaning Similarly, (झन) (don't/people) NN(common noun) PT(particles) represents meaning.

III. METHODOLOGY

Humans use their common linguistic knowledge to avoid ambiguity at the time of use of natural language, but computer carrying out the same task creates complex problems. First of all, we have to address them. Only when we have knowledge of the nature and source of ambiguity can we prevent it. The prevention of ambiguity requires identifying the knowledge of relationships between different characteristics of words. We have used a rule-based method to address this problem in which we prefer context-based rules. It is on the basis of these rules that this problem has been addressed. In which we have first created a large database of Chhattisgarhi language. In this research we have used analytical research methodology in which we manually tagged every word in Chhattisgarhi language in their grammatical category for analysis of large size corpus of Chhattisgarhi language. On that basis, analysis was obtained.

In natural language, there are many sources of ambiguity, such as the problem of syntax, the multi-oral problem, so we have difficulty in solving the problem of ambiguity.

3.1 Data Collection

In order to address the problem of ambiguity, we must first have knowledge of that language| In which the grammatical knowledge of that language, knowledge of syntax is first. For which we collected data from different fields of the language in which we produced corpus from Chhattisgarhi language stories (परसार के गरुदन, गोबरहिन डोकरी, मनटोरा, चुगलाहा, मइके के सुखा, मया अउ माया, सइताहा, सियान मन के सीख), Chhattisgarhi language novels (चंद्रकला, दिया के अंजोर).

3.2 Knowledge Base Solution

For a knowledge-based solution to the ambiguity of language, we have based the available literal source of Chhattisgarhi language (grammar written by Chandrakumar Chandrakar). Large size training data is not required to use knowledge based method, so this method comes under supervisor method. In the formulation of knowledge-based method, we first tagged the words of Chhattisgarhi language in their grammatical category so that each word can be identified by its grammatical category. So that words that display more than one grammatical category can be differentiate. It is on that basis that the rules have been framed.

IV. CONTEXTUAL RULES

When the word displays more than one tag in the text, the contextual rules are formed by basing the grammatical meaning of that word and the position in the sentence of that word. Contextual rules mean the special position and meaning of the word present in a text. That is, the next and previous word of the word in a sentence present in a sentence is based on the contextual rules. To remove the lexical and semantic ambiguity, it is necessary to have the grammatical knowledge of that word and the knowledge of the next and previous words of that word. for example...

$$\Sigma S = \text{Sentence}$$

$$\Sigma S_T = \text{Tag. Sentence}$$

$$S_T = w_1 t_1, w_2 t_2, w_3 t_3 \dots \dots \dots w_n t_n$$

$$S \frac{w_1 t_1, w_2 t_2, w_3 t_3 \dots \dots \dots w_n t_n}{tnwn - \text{change tag } wn - 1tn - 1}$$

$$[W_n t_n - W_{n-1} t_{n-1} / W_{n+1} t_{n+1}] \quad [\text{change the tag of } t_{n-1}]$$

Explanation...

If ΣS is a sentence. And ΣS_T is a tagged sentence then we can divide the whole sentence into $w_1 t_1, w_2 t_2, w_3 t_3, \dots, w_n t_n$. If there is ambiguity in the tag of $w_3 t_3$ due to the effect of $w_n t_n$ then it is By rule, t_3 will be changed to the correct tag and will give its output.

Table 4.1 :Example Of Ambiguous Words

AMBIGUOUS WORD	MULTIPLE TAG	MEANING OF WORD	EXAMPLE
फेर	NN,CC	चक्कर,फिर	1.मंगलू ह बड़ आसरा लेके सहर रइपुर के एक बनिया इहां नउकरी के फेर मां आय रिहिसा 2.फेर जग ह बड़ निरदई हवया।
डहर	NN,A	रास्ता,पथ,ओर,तरफ	1.गोधुली बेला म गांव डहर लहुटत बरदी। 2.हमर डहर के डउकी मन तुम्हर घर के डउकी मन ले ऊंचहा हवंया।
झिन	NN,PT	लोग,निषेध सूचक शब्द (मत)	1.सब झिन कहे लागिन – भगवान बड़े निर्दयी हे। 2.फेर रिस झिन करा।
बखत	NN,A	समय,बार	1.आज मेंहर तोर ले आखरी बखत एक ठन भीख मांगत हौं। 2.ओ बखत तुंहर ये जलकुकड़ा समाज हर कहां चल देय रिहिसा।
बने	JJ,VM,A	अच्छा,बनना,बहुत	1.मालूराम जी बने आदमी रहिस । 2.हर समय भीड़ बने रहय । 3.पुस्पादेवी अउ रमौतीन कथे, बेटी पीरा ह बाढ़गे बने जोर से धक्का दे।

In table no4.1 where NN (common noun),CC (conjunction),JJ(adjective),A(adverb),PT(particle),VM(main verb).

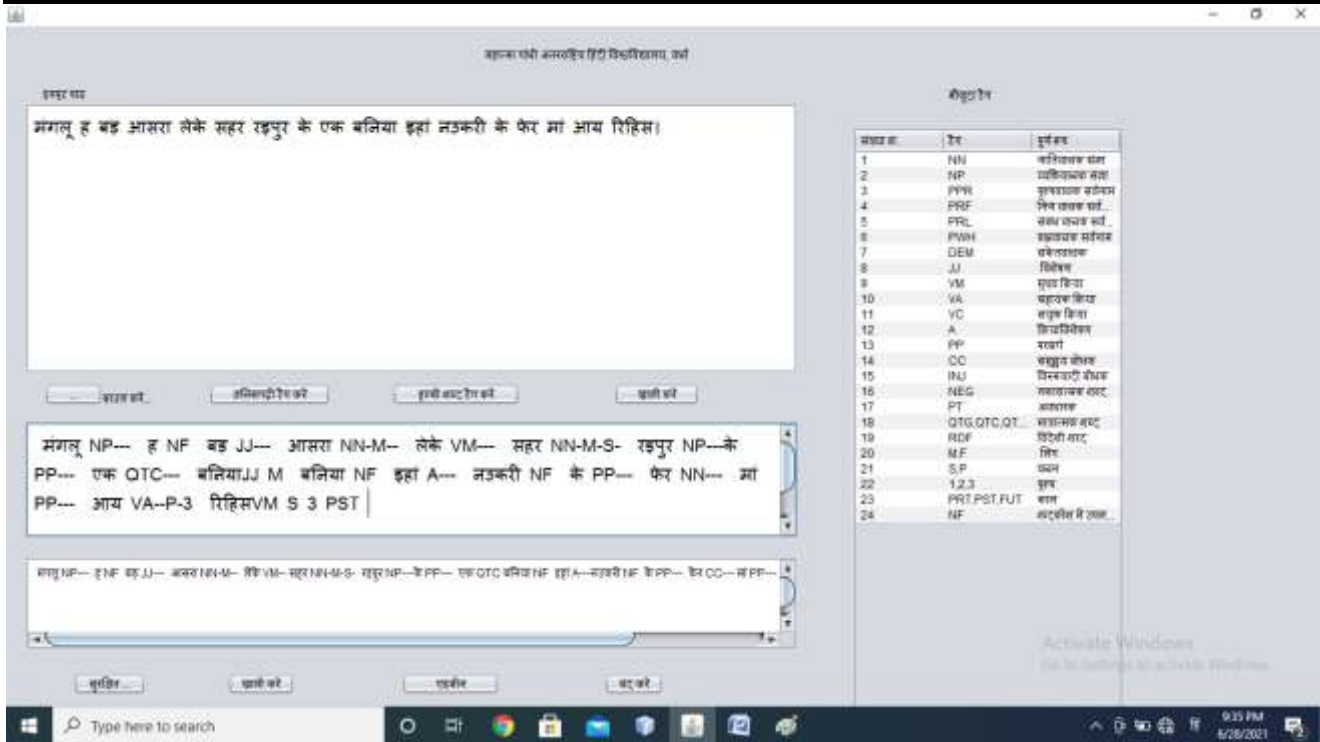
Algorithm steps.

Our algorithms perform disambiguation through the following steps.

1. Input the text of Chhattisgarhi language.
2. Input text is tokenized.
3. Tagged tokened input text.
4. If tagged words receive more than one tag.
5. To find out the ambiguity based on the analysis of a word with more than one tag.
6. To find out the correct tag of the word based on contextual rules.
7. The correct tag of the word will be obtained as output.

V . RESULT AND CONCLUSION

In this research, we have described a knowledge-based method to remove the ambiguity of Chhattisgarhi language. Under which we have built contextual rules. What is the role of Context in ambiguity? We have described it in this research paper. The context in ambiguity refers to the determination and limitation of the meaning of a word in a particular position in a text on the basis of which ambiguity is addressed. This system only prevents word - level ambiguity. This system can disambiguate noun, adjective, preposition, conjunction, verb, adverb present in a Chhattisgarhi language text. The research concludes that Chhattisgarhi, like other languages, suffers from ambiguity and that a knowledge - based method can be adopted based on the context of the word in that text to prevent it.



Screen shot of the system

REFERENCES

1. chandrakar chandrakumar 2006 “mank Chhattisgarhi vyakarn” satakshi prakashan Raipur.
2. chandrakar chandrakumar 2012 “vruhat Chhattisgarhi shabdkosh” Chhattisgarh rajya hindi granth akadami Raipur .
3. D.M. berry “ambiguity in natural language requirement documents” in innovations for requirement analysis from stakeholders needs to former designs springer 2008 pp1-7.
4. Dhopavkar Gauri , Kshirsagar Manali, Malik Latesh 2015 “Exploiting Rules For Resolving Ambiguity In Marathi Language Text” International Journal of Research in Engineering and Technology Volume: 04 Issue: 12 .
5. Jha Sanjay Kumar 2017 “EXPLORING AMBIGUOUS STRUCTURES IN ENGLISH” global journal of interdisciplinary social sciences Vol.6(3):1-4.
6. Dr. Narendra Dev Varma, “Chhattisgarhi Bhasha Ka Udvikas”, Chhattisgarh Rajya Hindi Granth Academy.