

## PDP Resolution of “IT” For Proper Identification of Gender Using Rule Based Approach

Priyanka Gaikwad<sup>1</sup>, Kalpana Khandale<sup>2</sup>, C. Namrata Mahender<sup>3</sup>  
<sup>1,2,3</sup>(Department of CS & IT, Dr. B.A.M.U. Aurangabad, India)

**Abstract:** Resolution of anaphora is most important concept in the linguistic research area. The present research is focused on pronoun ambiguity problem. The major concern in this research is on “it” issue. We cannot identify the antecedent of pronoun “it” is either feminine, masculine or neutral. For resolving “it” issue the rule-based approach is applied. The process of tokenization POS tagging are applied as preprocessing stage for designing the rules. The overall system performance of pleonastic “it” is around 73%.

**Keywords:** Natural Language Processing (NLP), Ambiguity, Anaphora, Pronoun Disambiguation Problem.

### I. Introduction

In the research area of natural language processing there is vast variety of text data available. With this large data so many problems are faced one of them is “pronoun ambiguity”. In this paper we try to address ambiguity related to “it”. The resolution of the ambiguity can be achieved by performing relevant operation on the text and find the reliable ambiguity in the text.

A particular challenge for computational linguistic in all levels of language is ambiguity. There are various types of ambiguity those are lexical ambiguity is the single word ambiguity, syntactic ambiguity which is structural ambiguity, semantic ambiguity these occurs when the meaning of the words themselves can be misinterpreted, discourse ambiguity and pragmatic ambiguity refers to the situation were the context of a phrase gives it multiple interpretation [1]. In the next section we have discussed ambiguity and their types in detail.

#### 1.1 Ambiguity:

Ambiguity refers to the state of having or expressing more than one possible meaning or something open to more than one possible meaning. A meaning of a word is just the thing that a word signifies or applies to in current or possible world. It is a property of linguistic expressions. If an expression (word/phrase/sentence) has more than one interpretation we can refer it as ambiguous. [2]

For e.g.: Consider the sentence,

“The chicken is ready to eat”

The interpretations in the above phrase can be, the chicken (bird) is ready to be fed or the chicken (food) is ready to be eaten. The types of ambiguity are represented in the following figure 1. And it will be discussed in later in section one by one.

PDP Resolution of “IT” For Proper Identification of Gender Using Rule Based Approach

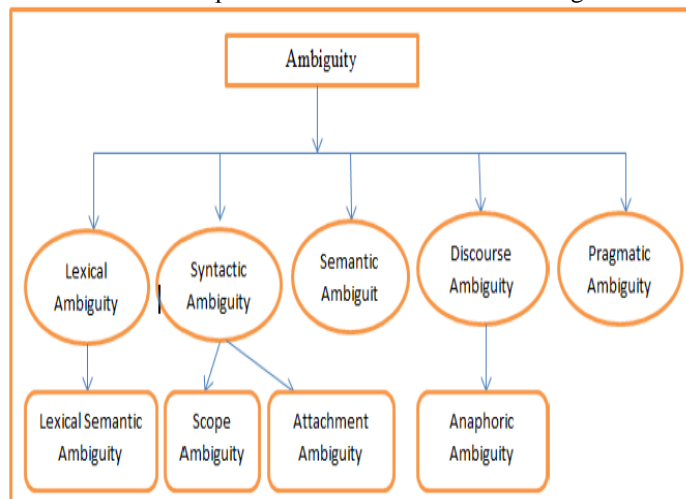


Figure 1. Types of ambiguity

### **1.1.1 types of ambiguity:**

#### **1.1.1.1 lexical ambiguity:**

It is the ambiguity of a single word. A word can be ambiguous with respect to its syntactic class. Eg: book, study.

**For e.g.** The word silver can be used as a noun, an adjective, or a verb.

- i. She bagged two silver medals.
- ii. She made a silver speech.
- iii. His worries had silvered his hair.

Lexical ambiguity can be resolved by Lexical category disambiguation i.e., parts-of-speech tagging.

**i. Lexical Semantic Ambiguity:** The type of lexical ambiguity, which occurs when a single word is associated with multiple senses. Eg: bank, pen, fast, bat, cricket etc.

For e.g. The tank was full of water.

I saw a military tank.

The occurrence of tank in both sentences corresponds to the syntactic category noun, but their meanings are different.

#### **1.1.1.2 syntactic ambiguity:**

The structural ambiguities were syntactic ambiguities. Structural ambiguity is of two kinds that are,

**i. Scope Ambiguity:** Scope ambiguity involves operators and quantifiers.

Consider the example:

Old men and women were taken to safe locations [3]

The scope of the adjective (i.e., the amount of text it qualifies) is ambiguous. That is, whether the structure (old men and women) or ((old men) and women)? The scope of quantifiers is often not clear and creates ambiguity.

#### **ii. attachment ambiguity:**

A sentence has attachment ambiguity if a constituent fits more than one position in a parse tree. Attachment ambiguity arises from uncertainty of attaching a phrase or clause to a part of a sentence. [3]

**Consider the example:**

The man saw the girl with the telescope [4]

It is ambiguous whether the man saw a girl carrying a telescope, or he saw her through his telescope. The meaning is dependent on whether the preposition „with“ is attached to the girl or the man. *PDP Resolution of*

#### **1.1.1.3 Semantic Ambiguity:**

This occurs when the meaning of the words themselves can be misinterpreted. Even after the syntax and the meanings of the individual words have been resolved, there are two ways of reading the sentence.

**Consider the example,**

“Seeta loves her mother and Geeta does too.” [4]

The interpretations can be Geeta loves Seeta’s mother or Geeta likes her own mother. Semantic ambiguities born from the fact that generally a computer is not in a position to distinguishing what is logical from what is not.

#### **1.1.1.4 Discourse Ambiguity:**

Discourse level processing needs a shared world or shared knowledge and the interpretation is carried out using this context. Anaphoric ambiguity comes under discourse level.

#### **I. Anaphoric Ambiguity:**

Anaphora are the entities that have been previously introduced into the discourse.

**Consider the example,**

The horse ran up the hill. It was very steep. It soon got tired. The anaphoric reference of „it“ in the two situations cause ambiguity. Steep applies to surface hence „it“ can be hill. Tired applies to animate object hence „it“ can be horse.

#### **1.1.1.5 Pragmatic Ambiguity:**

Pragmatic ambiguity refers to a situation where the context of a phrase gives it multiple interpretations. One of the hardest tasks in NLP. The problem involves processing user intention, sentiment, belief world, and modals etc. all of which are highly complex tasks. [3]

**Consider the example,**

Tourist (checking out of the hotel): Waiter, go upstairs to my room and see if my sandals are there; do not be late; I have to catch the train in 15 minutes.

Waiter (running upstairs and coming back panting): Yes sir, they are there.

Clearly, the waiter is falling short of the expectation of the tourist, since he does not understand the pragmatics of the situation. [3]

For the present work we considered here the anaphoric ambiguity. The anaphoric ambiguity comes under the discourse type of ambiguity. In this work we mainly focused on two major terms i.e. anaphor and the antecedent of that anaphor. Mainly in the process there is a need to examine the problem within the noun and the pronoun which is referred to that noun. Psychologically the human has identified the possibilities in the text, for e.g. "The trophy was fit in to the brown suitcase because it was too big." what was too big? In this example, the human can easily understand the trophy was too big but the system cannot understand either the trophy or the suitcase was too big. The systems are unable to reliably understand what the ambiguous word is or how the pronoun or anaphor referred to its antecedent or the noun. Anaphora is the pronoun and the antecedent is the noun like he, she, they, it etc. are the pronoun and it is too difficult to properly relate the pronoun. This paper discusses the proposed process of resolution of "it" pronoun ambiguity. We mainly concentrate on the feminine, masculine and the pleonastic forms of "it". It means that the "it" are classified by the system as feminine, masculine or the pleonastic. For example,

1. "A wolf was very hungry. It looked for food here and there."

In the above example the "it" represent its antecedent i.e. „wolf". Here, we simply understood who is the antecedent and what is the pronoun refer to it. But in many sentences more than one noun are present in discourse. For example,

2. "Once upon a time there lived a lion in forest. One day after a heavy meal, it was sleeping under a tree. "

In this example, there is a lion as well as forest are two nouns. Here, the ambiguity present within the pronoun and antecedent. For these purpose we have to concentrate on the "it" pronoun is to be classified to feminine, masculine or pleonastic in the sentences for properly refereeing what "it" indicates in the above sentence. There is ambiguity among the phrase like whether, „lion is sleeping" *PDP Resolution of* or „forest is sleeping". For these two possibilities we have to find the exact pronoun or anaphor for the possible noun or antecedent. One another example, i.e.

**3. "It is raining"**

In the above example, there is no relevant antecedent for the pronoun „it" because of this it is called as pleonastic "it" (there is no referent for „it") but it refers to abstract noun as either cloud or weather.

In the next section we have discussed PDP (Pronoun Disambiguation Problem) and proposed method which we have applied to identify the ambiguity.

**Ii. Pronoun Disambiguation Problem:**

Accurate disambiguation of relative pronoun is important for any natural language processing system that hopes to process real world texts. It is especially concern for corpora where the sentences tend to be long and information packed. Unfortunately, to understand a sentence containing a relative pronoun, an NLP system must solve two difficult problems: the system has to locate the antecedent of the relative pronoun and then determine the antecedent"s implicit position in the embedded clause.

When the pronoun and its antecedent occur within the same sentence we can determine which antecedent is correct. But when the pronoun and antecedent span separate sentences, two or more antecedent can each produce a grammatical discourse (example 2 as discussed in introduction section). Our approach to investigate proper antecedent for the relevant pronoun. We are concern with anaphoric pronoun and work focuses on the pronoun "it". In some sentences "it" may be feminine, masculine or pleonastic.

The focus of this study is to find the exact or relevant pronoun that is "it". We have to observe that the pronoun ambiguity problem plays an important role in psycholinguistic phenomenon which determines the antecedent for pronoun. There are few types of discourse which are as listed below:

**2.1 Anaphora:-**

Anaphora is a referring expression that refers to some entity which has antecedently been introduced into the discourse. The anaphor resolution is nothing but the process of binding or mapping the „referring expression" to the correct antecedent in the given discourse. For example, "Shyam had to go a meeting so he decided to have a shave". In this example, "he" refers to the "Shyam" and "Shyam" comes before the "he". So "Shyam" is an anaphor. [5]

**2.2 Cataphora:-**

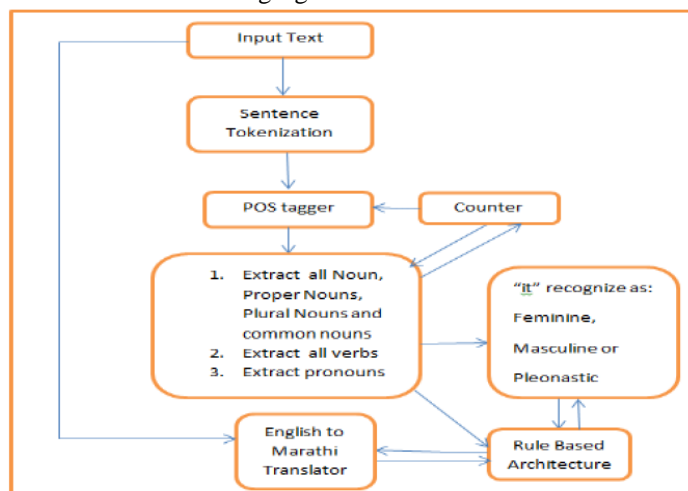
It is used to describe an expression that co-refers with a later expression in the discourse. For example, “When he arrive home. Shyam went to sleep”. In this sentence the pronoun “he” (the anaphor) appears earlier than the noun “Shyam” (the antecedent). [5]

**2.3 Exaphora:-**

If referring expression is missing in discourse or implicit is called as exaphora. For example, “On 21 December 1972, the Basic Treaty was signed by East or West Germany, and relations between the countries started to improve” In these example the Basic Treaty is „exaphoric” referring out of the text to the de facto recognition of East German by the West German government. [6]

**III. Proposed method for resolution of the pdp:**

The flow of the system is shown in the following figure 2.



**Figure2** architecture of pleonastic “it”

**3.1 Input Text:**

In first module we have consider 50 sentences which has the word or the pronoun “it”.

**3.2 Tokenization:**

The first module is responsible for text segmentation, and it divides the text into tokens. In which we have tokenize the text into the sentences. When the full-stop (.) is occur in the sentence then tokenize it. [1]

**3.3 Pos Tagging:**

Using NLTK we have tag tokenized sentences. The tag is labeled such as noun, pronoun, verb, adjective etc. Here we use the counter for count the number of times all the part-of speech occurred in the sentence and used it further.

**3.4 Rule Based System:**

The work observes the difficulties in finding/resolving ambiguity in the pronoun “it”. For this we designed rule based architecture for the finding/resolving the ambiguity and state whether it is the anaphor used for the antecedent. For the purpose of better understanding of the pronoun ambiguity we took support of a language which classifies “it” to a gender form too, here used language is Marathi. In Marathi following forms are used to represent “it”. Thus before framing the rules firstly translation of English statement into Marathi is done then accordingly rules are framed by differentiating among the pronoun and noun or antecedent and its anaphor. The following table shows the differences with the bilingual results in which “it” represents feminine, masculine or pleonastic. The below table shows the result for pronoun “it”. For the resolution of the anaphora or the pronoun we have work on the 50 random sentences, in which basically they are feminine, masculine or pleonastic. Most of them as the pleonastic in nature. PDP Resolution of “IT” For Proper Identification of Gender Using Rule Based Approach

*Pdp Resolution Of "It" For Proper Identificationof Gender Using Rule Based Approach*

Sr. No.	English Sentence	Noun				Verb						Pronoun		Marathi Meaning	What Found	
		NN	NNP	NN S	NNP S	VB	VB G	VB P	VB Z	VB D	VB N	PR P	PRP S			
1	It is raining.						raini ng		is				it		पाऊस पडत आहे.	Pleonas tic it.
2	It is Sunny.								is				it		चांगला सुयेपकाश आहे.	Pleonas tic it.
3	It is sleeping						sleep ing		is				it		तो झोपेत आहे.	Pleonas tic it.
4	It is running.						runn ing		is				it		कायैरत.	Pleonas tic it.
5	Dam it!		Dam										it		रफू करणे.	Pleonas tic it.
6	It seems important that I see him.								see	seem s			It, I, him		मला त्याला पाहणे महत्वाचे वाटते.	Masculi ne it.
7	The session is opened, it was announced.								is	Onc e	Ope n, Ann ounc ed		it		सत्र उघडले आहे हे जाहिर करण्यात आले.	Masculi ne it.
8	But it doesn't take much to get burned.						Take, get		does			burn ed	it		परंतु त्याने जळव्यासाठी जास्त घेतल नाही .	Masculi ne it.
9	It is important to observe it is raining.						obs erv ed	raini ng	Is				It, it		पाऊस पडत आहे हे पाहणे महत्वाचे आहे.	Pleonas tic it.
10	Its just a veterinary clinic.												Its		तो फक्त एक पशुवैद्यकीय दवाखाना आहे.	Masculi ne it.
11	It appears John is late.		John						App ears, is				It		अस दिसतय जॉनला उशीर झाला.	Pleonas tic it.
12	It turned out the money was		Mo ney							Turn ed, was			It		पैसे बनावट असल्याचे बाहेर आले.	Pleonas tic it.
	counterfeit.	Co unt erf eit														
13	Priyanka enjoys playing basketball; she knows that it is her favorite sport.	Ba ske tba ll, Sp ort	Priya nka	Enjo ys			Playi ng		Kno ws, is				She, it	her	प्रियन्का आनंदाने बास्केटबॉल खेळते, ती तिचा आवडता खेळ आहे हे तिला माहित आहे.	Femini ne it.
14	it was quite humble and behaved well.									Was, Beha ved			It		तो खूप नम्र आणि चांगली वागणूक असणारा होता.	Masculi ne it.
15	Its half past one.	Hal f, Pas t											Its		तो अर्धे भूतकाल होता .	Masculi ne it.
16	Its sunny today.	Su nn y, tod ay											Its		आज उन् आहे.	Pleonas tic it
17	It was a good film.	fil m							Was				it		तो चांगला चित्रपट होता.	Masculi ne it
18	It will be February 1 next week.	we ek					be						it		१ फेब्रुवारी पुढच्या आठवड्यात असेल.	Pleonas tic it
19	One cold windy night the lamp was blown off, it started crying.	Wi nd y, nig ht, la mp					cryin g		Cold was, start ed	blow n			it		एका वादळी राती दिवा विझला, तो रडायला लागला.	Masculi ne it
20	The lamp felt very proud of itself that it was the brightest object in the universe.	La mp , obj ect, uni verse							Felt, was	prou d			it	itself	दिव्याला स्वतःचा खूप अभिमान आहे की ती विश्वातील प्रतिभाशाली वस्तू आहे.	Femini ne it

**Table2. Overall result of the system**

SENT. DATA	PLEONASTIC "IT" SENTENCE	MASCULINE "IT" SENTENCE	FEMININE "IT" SENTENCE
DATA TAKEN	30	15	05
IDENTIFIED SENTENCES BY SYSTEM	22	13	05

This paper work on the total 50 sentences of the represent "it" pronoun. In which we have taken the 30 sentences of the pleonastic, 15 sentences of the masculine and the 5 sentences of the feminine "it". The system has identified the 22 sentences from the 30 sentences as pleonastic "it", 13 sentences as masculine from the 15 sentences, and 5 sentences as feminine from the 5 sentences of "it".

#### **IV. Conclusion**

The ambiguity is the major problem in the natural language processing. Ambiguity refers to the state of having or expressing more than one possible meaning or something open to more than one possible meaning. There are few types of ambiguity that are, lexical ambiguity, Syntactic ambiguity, semantic ambiguity, discourse ambiguity etc. Our work is based on the discourse type of ambiguity. In which the anaphoric ambiguity we have to focus in our work. In this problem the ambiguity comes under the sentences in which antecedent is not present for the pronoun or the anaphor. Here, focuses on the pronoun "it". In this "it" may be feminine, masculine or pleonastic. The sentences is based on the "it" type of pronoun and we have find the whether the sentence is pleonastic or not. The overall system performance of pleonastic "it" is around 73% and overall system gender identification is 80%. This system performance is for the small database further the performance of the rule base system can be tested on a large database.

#### **References**

- [1]. Bird, S., Klein, E., & Loper, E. Natural language processing with Python. "O'Reilly Media, Inc." (2009).
- [2]. Wilkinson, D. J. (2006). The Nature of Ambiguity. In *The Ambiguity Advantage* (pp. 9-41). Palgrave Macmillan UK.
- [3]. Bhattacharyya, P. (2012). Natural language processing: A perspective from computation in presence of ambiguity, resource constraint and multilinguality. *CSI journal of computing*, 1(2), 1-13.
- [4]. Tiwary, U. S., & Siddiqui, T. (2008). *Natural language processing and information retrieval*. Oxford University Press, Inc...
- [5]. Pardeep Singh, Kamlesh Dutta (July 2015) Analysis of Anaphora, Cataphora and Exaphora for Demonstrative Pronoun of Hindi *International Journal of Advanced Research in Computer Science and Software Engineering*
- [6]. Anjali M K, Babu Anto P *Ambiguities in Natural Language Processing Vol.2, Special Issue 5, October 2014*