Conditional Random Fields based Pronominal Resolution in Tamil

A.Akilandeswari
AU-KBC Research Center
Madras Institute of Technology,Chrompet,
Chennai, India
akila@au-kbc.org

Sobha Lalitha Devi
AU-KBC Research Center
Madras Institute of Technology,Chrompet,
Chennai, India
sobha@au-kbc.org

Abstract—This paper deals with Tamil pronominal resolution using Conditional Random Fields a machine learning approach. A detailed linguistic analysis of Tamil pronominals and its antecedence occurring in various syntactic constructs is done, which led to the selection of appropriate features for CRF approach. The syntactic features thus identified made the system learn most frequently occurring pronoun antecedent pattern from the training corpus. The performance of the system is highly encouraging.

Keywords-Anaphora; Antecedent; pronominal;Anaphor; CRF++;

I. INTRODUCTION

Pronominal Resolution is a well studied area for English but for Indian languages not sufficient work has been done. The process of finding the antecedent of a pronoun is pronominal resolution. In this paper we analyze the third person pronominals in Tamil *avan* "he", *aval* "she" and *atu* "it". Consider the following example

Consider the following examples,

Krishnan *avanai*$_i$ maatRRinaan

Krishnan he+acc changed+3sgm

'Krishnan changed him.' *(ex.1)*

In ex.1 Krishnan is the subject and is also a proper noun with nominative case marker. Even though Krishnan is the subject with nominative case marker, it cannot be the antecedent for the anaphor avanai$_i$. Hence the pronoun avanai$_i$ does not refers to Krishnan but it refers somebody in the previous sentences.

Krishnan$_i$ *avanai*$_i$ maatRik koNtaan.

Krishnan he+acc change got+3sgm

'Krishnan changed himself' *(ex.2)*

In ex.2 Krishnan is the subject and is also a proper noun with nominative case marker. Here the pronoun avanai$_i$ with accusative case marker refers Krishnan because of the verb koNtaan (got+3sgm).

I. RELATED WORK

It is observed that approaches to anaphora resolution usually rely on a set of anaphora resolution factors. Factors used frequently in the resolution process include gender,person and number agreement, c-command constraints, semantic consistency, syntactic parallelism, semantic parallelism, salience, proximity etc.

One of the early works in pronominal resolution is by Hobb's naive approach, which relies on semantic information (Hobbs. J, 1978). Carter with Wilkas' common sense inference theory came up with a system (Carter. D, 1987). Carbonell and Brown's introduced an approach of combining the multiple knowledge system (Carbonell. J. G. & Brown. R .D, 1988).

The initial approaches, where broadly classified as knowledge poor and rich approach. Syntax based approach by Hobb (naive approach), centering theory based approaches (Joshi, A. K. & Kuhn. S, 1979; Joshi, A. K. & Weinstein.S, 1981) and factor/indicator based approach such as Lappin and Leass' method of

identifying the antecedent using a set of salience factors and weights associated to it. This approach requires deep syntactic analysis.

Ruslan Mitkov introduced two approaches based on set of indicators, MOA (Mitkov's Original Approach) and MARS (Mitkov's Anaphora Resolution System) (Mitkov. R,1998). These indicators return a value based on certain aspects of the context in which the anaphor and the possible antecedent can occur. The return values range from -1 to 2. MOA does not make use of syntactic analysis, whereas MARS system makes use of shallow dependency analysis.

Several coreference resolution systems are currently publicly available. JavaRap (Qiu et al., 2004) is an implementation of the Lappin and Leass' (1994) Resolution of Anaphora Procedure (RAP). JavaRap resolves only pronouns and, thus, it is not directly comparable to Reconcile. GuiTaR (Poesio and Kabadjov, 2004) and BART (Versley et al., 2008) (which can be considered a successor of GuiTaR) are both modular systems that target the full coreference resolution task.

In addition, the architecture and system components of Reconcile (including a comprehensive set of features that draw on the expertise of state-of-the-art supervised learning approaches, such as Bengtson and Roth (2008)) result in performance closer to the state-of-the-art.

In addition, the architecture and system components of Reconcile (including a comprehensive set of features that draw on the expertise of state-of-the-art supervised learning approaches, such as Bengtson and Roth (2008)) result in performance closer to the state-of-the-art. Johansson, C. (Ed.) Proceedings of the Second Workshop on Anaphora Resolution (2008).

A. Anaphora Resolution In Indian Languages

Some of the works done in Anaphora Resolution for the Indian languages are as follows,

VASISTH a rule based system which works with shallow parsing and exploits the rich morphology in Indian languages for identifying the antecedent for anaphors (Sobha.L & Patnaik.B.N, 1999). Pronominal resolution in tamil using machine learning (Murthi.K.N, Sobha.L, 2007) work showed that the task may be feasible, and depend on the reliability of language specific features such as person number, gender and case marking. Also have looked into the anaphora resolution in Tamil, using Machine Learning technique: Linear Regression and compared it with salience factors.

Dhar worked on "A method for pronominal anaphora resolution in Bengali (Dhar.A & Garain.U, 2008). Sobha.L & Pralayankar.P, 2008) worked on "Algorithm for Anaphor Resolution in Sanskrit". Resolving Pronominal Anaphora in Hindi Using Hobb's Algorithm was done by Kamalesh Dutta. Sobha L, (1999) "Anaphora Resolution In Malayalam and Hindi" Doctoral dissertation submitted to Mahatma Gandhi University, Kottayam , Kerala. Dhar worked on "A method for pronominal anaphora resolution in Bengali (Dhar.A & Garain.U, 2008; Sobha.L & Pralayankar.P, 2008) worked on "Algorithm for Anaphor Resolution in Sanskrit". Resolving Pronominal Anaphora in Hindi Using Hobb's Algorithm was done by Kamalesh Dutta. In ICON 2011, NLP tool contest on Anaphora Resolution for Indian Languages was held. The tool contest considered the languages such as Bengali, Hindi, Odiya, Marathi and Tamil. In each language different methods was approached by the participants.

The paper is organized as given below. The introduction is followed by the linguistic analysis of pronominals in Tamil with examples and type of pronominals. In section 5 we discuss in detail the implementation of the system which includes feature selection, CRFs and the pre-processing modules. The last section deals with the evaluation and result followed by conclusion.

II. AN OVERVIEW OF TAMIL LANGUAGE

Tamil belongs to South Dravidian family of languages.Tamil is predominantly a free word order language. Generally Tamil sentence follows the subject, object, and verb pattern. The interchange of subject, object is common. Tamil is morphologically rich and word order free. It has post-positions. It is nominative-accusative language like the other Dravidian languages. The construction of sentences has nominative subjects in Tamil. There are constructions with certain verbs that require dative subjects and possessive subjects. Tamil has PNG (person, number and gender) agreement. First and second person singular and plural are used as deictic, though they are used in anaphoric form in discourse. There are many approaches to solve this problem such as rule based, statistical and machine learning based approaches.

III. TYPES OF ANAPHORA

Pronominal anaphor: Pronominals in Tamil should agree in gender, number and person with its antecedent.

Krishnan$_i$ viittiRkku vantaan. avan$_i$ avanutaiya$_i$ naarkaliyil amarntaan.

Krishnan home+dat came+3sgm. He he+genitive chair+loc sat+3sgm.

'Krishnan came home. He sat on his chair' *(ex.3)*

In ex.3, there are two anaphors avan$_i$ and avanutaiya$_i$. The anaphor avanutaiya$_i$ refers avan$_i$ and avan$_i$ refers Krishnan$_i$.

Quantifier/Ordinal: The anaphor is a quantifier such as one, two etc and has its own suffixes.

Krishnan oru putu Pena$_i$ vankinaan. Ramanum onrai$_i$ vaankinaan.

Krishnan one new pen bought+3sgm. Raman+um one+acc bought+3sgm.

'Krishnan bought a new pen and raman also bought one.' *(ex.4)*

In ex.4 onrai$_i$ is a quantifier which is a pronoun refers Pena$_i$ in the previous sentence.

Pleonastic anaphor: The pronouns 'atu' (it) itself refer to nothing particular in the text.

atu$_i$ oru maalai neram.

It one evening time

'Its an evening' *(ex.5)*

In ex.5 *atu$_i$* is the anaphor which do not refers anything in the text. So it is non-anaphoric.

Whole and the part anaphor: This category of anaphor where the anaphor refers to some real world knowledge which has not been mentioned previously anywhere in the discourse.

Krishnan *matikkaNani$_i$* vaanki ullan. Inta *iyantirattai$_i$* avanal ella itattiRkkum etuttu cella mutikiratu.

Krishnan laptop bought be. This iyantiram+acc he+ins all place+dat take go able

'Krishnan bought a laptop . He is able to take this machine every where' *(ex.6)*

Here *matikkaNani$_i$* is an anaphor and *iyantirattai$_i$* is the antecedent. This type of anaphors need world knowledge to resolve. This work mainly focussed towards pronominal anaphors.

IV. TYPES OF PRONOMINAL ANAPHORA IN TAMIL

The pronominal anaphora in Tamil is further classified into personal anaphors, possessive anaphors, reflexive anaphors, demonstrative anaphors and relative anaphors. Most pronominal anaphora resolution algorithms only account for anaphors referring to individual entities. We have taken the anaphors (avan, aval, atu) for our work and it is given below.

The pronominal anaphora in Tamil are based on personal anaphors such as avan/he, avaL/she, atu/it and avarkaL/they,their.

A. Possessive Anaphora

The possessive anaphors end with morphemes such as *utaiya, atu* and *in.* The anaphors and their inflected forms are given below.

His	avan***utaiya***	Avan***atu***	Avan***in***
	he+gentive	he+genitive	he+gentive
Her	avaL***utaiya***	avaL***atu***	avaLn***in***
	she+gentive	she+gentive	she+gentive
Its	atan/atan***utaiya***	atan**atu**	Atan**in**
	It+genitive	It+genitive	It+genitive
Their/Theirs	avarkaL***utaiya***	avarkaL***atu***	avarkaL***in***

B. Reflexive Anaphora

The reflexive anaphors end with morphemes such as "e" (clitic marker) and *"aaka" (benefactive)* marker. The anaphors are given below.

Himself	avan**e**	avaN**aaka**	-
	he+e(clitic marker)	he+benefactive	
Herself	avaL**e**	avaL**aaka**	-
	she+e(clitic marker)	he+benefactive	
Itself	atuv**e**	at**aaka**	at**e**
	It+e(clitic marker)	It+benefactive	It+e(clitic marker)

C. Non-Possessive Anaphora

The Non-possesive anaphors are pronouns which are not possessive or reflexive. These anaphors have suffixes with case markers such as nominative, sociative, locative etc are non-possessive anaphors.

He	avan	Nominative case marker
She	avaL	
He+loc	avan**itam/itatil**	Locative case marker
she+loc	avaL**itam/itatil**	
he+dat	avan**ukku**	Dative case marker
she+dat	avaL**ukku**	
he+acc	avan**ai**	accusative case marker
she+acc	avaL**ai**	
he+soc	avan**otu/utan**	Sociative case marker
she+soc	avaL**otu/utan**	
he+ins	avan**aal**	Instrumental case marker
she+ins	avaL**aal**	

V. IMPLEMENTATION

A. Annotation of anaphora

Corpus

The Tamil text which is taken from the web on tourism domain and a historical novel are used for this work. This corpus contains 10,000 sentences. And the historical novel consists of simple and complex sentences which is actually differs from the normal web text. The size of this corpus is 16,000 sentences.

Pre-processing

We used the shallow parser technique for pre-processing. The corpus is preprocessed with syntactic information such as POS, noun phrase chunk, verb phrase chunk, morphological information, clause boundary and Named Entity Recognition. After each step of preprocessing, the output of the process is added as a tab-separated column to each token in the file.

Annotation Using PALinkA

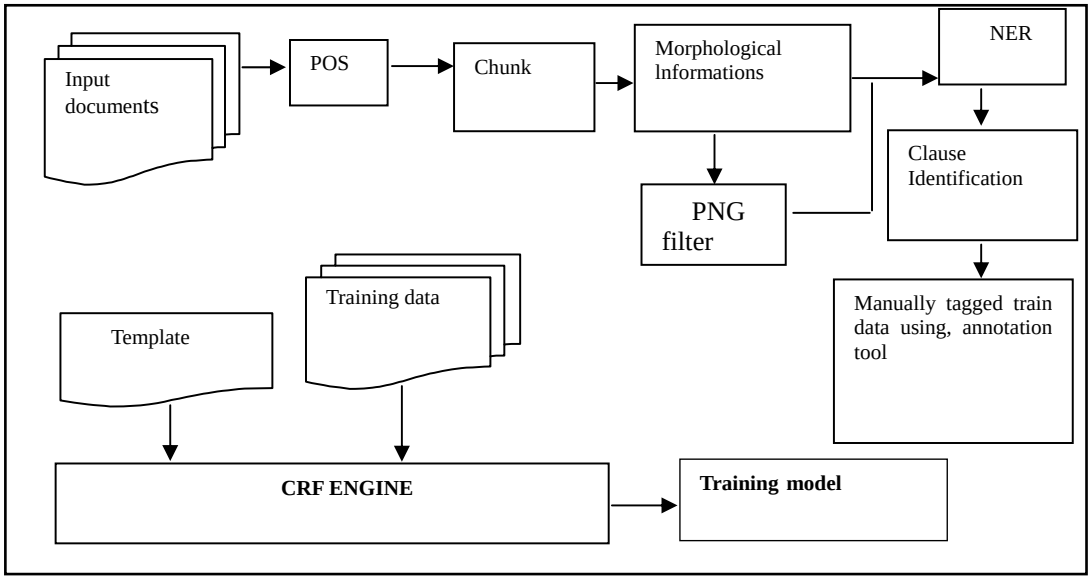
The next task is to annotate the corpus with anaphor and antecedents. PALinkA is an annotation tool which is customized for Indian languages for the tagging purpose. PALinkA is abbreviated as Perspicuous and Adjustable Links Annotator. It's a language independent tool, written in java. The input file to PALinkA has to be a well-formed XML file and the produced output is also a well-formed XML. The preprocessed files which is in column format with all syntactic information is converted to XML format. For annotations, initially, anaphor and antecedent should be marked as markables and if it is anaphoric, link is established between these two markables. Finally all the possible anaphor and antecedents are tagged with index. After annotation, these XML files are converted to column format files which are required for the machine learning system.

B. Anaphora Resolution

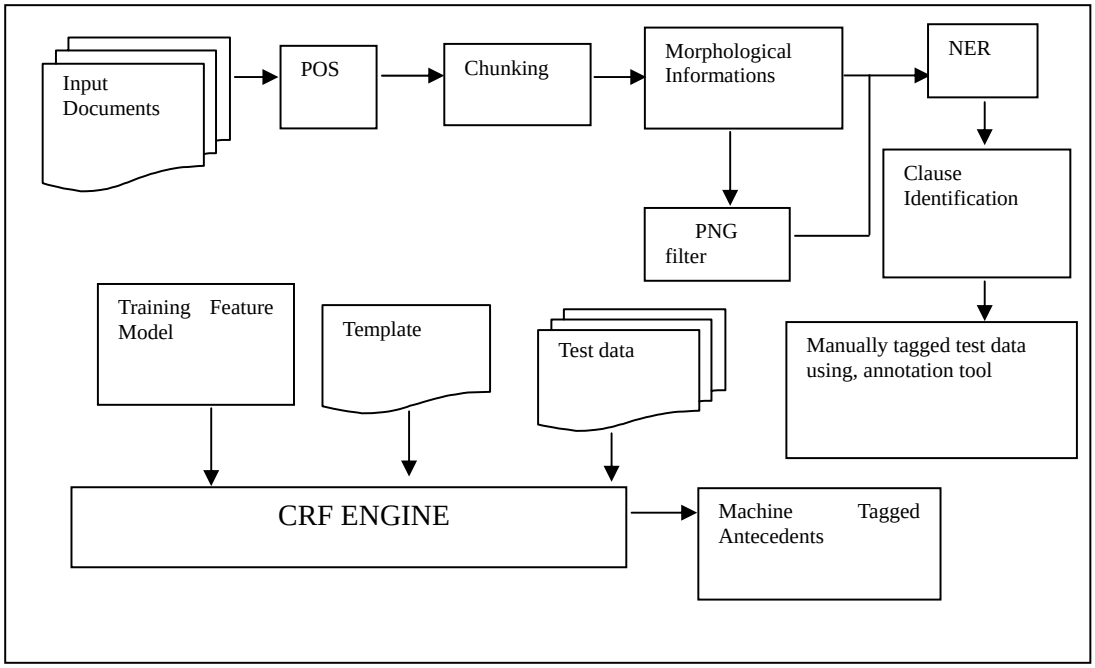
Methodology

We have taken five sentences above the sentence in which anaphor occurs. In identifying the antecedent of the anaphor, we use a machine learning method, CRF where features are syntactic factors.

Our System



System flow to train the data.



System flow to test the data.

Conditional Random Fields

The task of anaphora resolution is to find out the relations between the references between the sentences and with in the sentence. Hence CRFs is suitable for this problem. CRF++ is a machine learning algorithm designed for generic purpose and can be applied to a various NLP tasks, like Named Entity Recognition, Information Extraction and Text Chunking. CRFs is an undirected graphical model, for labeling and segmenting structured data, such as sequences, trees and lattices, where the conditional probabilities of the output are maximized for a given input sequence (Lafferty et al., 2001). The machine learning method of CRFs was chosen for this problem anaphora resolution because the advantage of CRFs is that it can model not only sequential data, but also non-linear data.

For our task, CRFs works in two stages, namely training and testing. In the training stage, the training data and template file is given as input. The training data is preprocessed and manually tagged. The template file contains the rules to learn the features from the training data. Taking these two files as input CRFs training module generates a model file, which holds the extracted features. In the next phase, testing is done by supplying test data and feature model to the CRFs test engine. The CRFs test module takes these two files as input and generates the antecedent.

Factors Description

Tamil Anaphora Resolution System usually rely on a set of factors. Generally the factors used in anaphora resolution process include gender, number, person agreement. Hence one is eliminating by filtering the set of possible antecedents by constraints and other is to set the preference for possible antecedent. The factors we used as features in CRF are given below.

a. Gender Agreement: The gender male, female and neuter gender is to be differentiated correctly by this agreement.

b. Number Agreement: This agreement distinguishes between singular and plural of anaphor and antecedent is matched.

c. Person Agreement: This filter should correctly distinguish three types of person.

d. Syntactic Constraints: The constraint that there should exist a syntactic relationship between a anaphor and its possible antecedent, refer the ex.1 above.

e. Selection Constraints:

7.1 Raman kaarai$_i$ en vaNtikku arukil niruttinaan.

Raman car+acc my vehicle+dat near+loc parked+3sgm

'Raman parked the car near my vehicle.' (ex.7)

7.0 avan $atai_i$ irantu naatkaLukku mun vankinaan.

he it+acc two days+dat before bought+3sgm

'He bought it two days before.' (ex.7)

Here in ex.7 sentence 7.0 is the base sentence and sentence 7.1 is the previous sentence for the base sentence. $atai_i$ is the anaphor referring $kaarai_i$, a noun phrase with accusative case marker and not vaNtikku which is a dative noun phrase in the previous sentence.

f. Sentence recency: The pronominals avan, avaL, atu are the pronoun which are always traverse backward from the current sentence and to other previous sentence to find the antecedent. During the traverse it may also refers the same person as pronoun like below

NNP <= NN(PRP) <= anaphor

Sita <= avaLukku <= avaL

We have considered the anaphor atu which is having noun phrase antecedents only, atu also refers events. In Indian languages, a nominative noun phrase, a possessive noun phrase with a nominative head and a dative noun phrase could be a subject of a sentence. The sentence recency is one of the important factor and hence the current sentence in which the pronoun occurs gets a score of 50 and this gets reduced by 10 for each preceding sentence.

The sentence recency is one of the techniques that are used by the system to implement the above constraints precisely. In the example below there are three relative sentences, which shows the importance of this factor. For example in the sentence

8.2 Raman kaiPeci vaankinaan.

Raman mobile bought+3sgm.

'Raman bought mobile.' *(ex.8)*

8.1 $Krishnano_i$ $matikaNani_j$ vaanki uLLaan.

Krishnan laptop bought be+3sgm .

'Krishnan bought laptop.' *(ex.8)*

8.0 $avan_i$ vakuppil $atai_j$ parri Peci koNtiruntaanar

He class+loc it+acc about talk got+3plmf

'In his class every body talked about it.' *(ex.8)*

In the ex.8, the sentence 8.0 is the base sentence and sentence 8.1 and sentence 8.2 are the two previous sentences for the base sentence.

Features

The features are identified after analysis of the tamil text and it is given below.

1. Current sentence
2. Previous sentences upto 5^{th} sentence
3. Current clause
4. Immediate clause
5. Non-immediate clause
6. Possessive pronoun
7. Non-possessive pronoun
8. Reflexive pronoun
9. Person, Number and Gender Match
10. Noun phrase nominative
11. Noun phrase possessive
12. Noun phrase dative
13. Noun phrase accusative
14. Noun phrase others
15. Named Entity Recognition.

VI. RESULTS AND DISCUSSION

The experiment is done using 26000 sentences in which there are 1000 pronouns. We used 80% of the data for training and 20 % for testing.

TABLE I. RESULTS OF ANAPHORA RESOLUTION

S.no	No of Anaphors in Testing	No of Anaphors system tagged	No of Anaphors system tagged correctly	Precision	Recall
1	158	162	138	85.18%	87.34%

The system tagged incorrectly in the case of when more than one nominative noun phrase occurs in the same sentence, which makes the system ambiguity and the system tagging it wrongly.

VII. ERROR ANALYSIS

9.3 {Raaman$_j$ aluvalakattiliruntu viitirku vantu,}NF {Siitaavai$_i$ paartan.} MCL

Raman office+loc+ abl home+dat come sita+acc saw+3sgm

'Raman came from the office to home and saw Sita'.

9.2 {*avaL$_i$* tanakuu uTampu cariyillai enRu connataal}CON {*avaLai$_i$*

She her+dat health not good told she+acc

maruthuvamanaikku azhaittu cenraan.}MCL

hospital+dat taken went+3sgm

' Because she said that she is not feeling well he took her to hospital'.

9.1 {pookum vazhiyil}RP { *Siitaavin$_i$* toozhi, Giita etiril vanthaaL}MCL

Going way+loc sita+gen friend Gita direction came+3sgf

'On the way Sita's friend Gita came in direction.'

9.0 {pinnar viiTiRku vantavuTan, *avan$_j$ avaLukku$_i$ roTTiyum paalum tayaar*

After home+dat came he she+dat roti+um milk made

ceytu koTuttan.} MCL

prepared give+3sgm

'After coming home, he prepared bread and milk for her.' *(ex.9)*

In this example, sentence 9.0 have two pronouns *avan* and *avaLukku.* Eventhough *avaLukku* refers *Siitaavin$_i$* which is possessive/genitive case marker , Gita is the nominative proper noun which could be the most probable antecedent according to salience factor.

In sentence 9.2 *Raman* is a subject, which is dropped. Hence *avan$_i$* in the sentence 9.0 which refers to *Raman* which is in the sentence 9.3.

10.2. {neRRu vanta paiyan *Raaman$_i$*} RP {anku ninRu koNTiruntaan.}MCL

Yesterday came boy Raman there stand be+3sgm

'The boy who came yesterday is standing there'.

10.1. *{avan$_i$* neRRu mazhaiyil nanaintu koNTu}NF {veelai ceytataaka kumaar

He yesterday rain+loc wet be work done+ben Kumar

connaan.}MCL

told +3sgm

'Kumar told that yesterday he was working in the rain'.

10.0. {*avane$_i$* tan kuTumpattai kaappaaRRi varum vicayam}RP {makhizhciyaaka

He+e his family+acc taken care come information happpy+aaka

iruntatu enRum connaan.}MCL .

be+pst also told+3sgm.

'He is taking care of his family by himself, he added'. *(ex.10)*

Generally, Tamil allows subject drop. In the above example the first pronoun *avane* is reflexive pronoun in Tamil. As per theory, reflexive pronoun always refers subject as its antecedent. In the above example the subject *Raman* is dropped. Hence the antecedent cannot be identified.

According salience factor, Kumar is the nominative proper noun in the previous sentence is the probable antecedent. But actually avane$_i$ refers Raman.

And finally if we take the pronoun in sentence 10.1, *Raman* is the subject of the previous sentence 10.2 and it is the antecedent of *avan.*

A. Issues of atu

11.1 {viratam irukkum naaTkaLil}RP {*manjal allathu civappu tuNikaLai*

Fasting be+fut days+loc yellow or red dressess

aNintukoLLa$_i$ veeNTum.}MCL

wear be+aux

' Have to wear yellow or red dressess during fasting days'.

11.0 *{atu$_i$* muTiyaatavarkaL kazhuttil tuNtu aNintukoLLa veeNTum.} MCL

It unable+3pl+h neck+loc towel wear be+aux

'Those who are unable to wear it can wear towel around neck.' *(ex.11)*

In the above example the anaphor atu refers to the antecedent which is a combination of noun phrase and verb phrase. So it is difficult to distinguishing "*atu"* whether it refers the event or noun phrase. And each of them should be resolved separately.

12.0 {piRaku poonkuzhali, *"cakravarttikku uTampu cukamillai enRu collkiRaarkale$_i$,}COM*

After poonkuzhali king+dat health not good be told+3pl

{*atu$_i$* unmaitaanee?" enraaL.}MCL

it true? said+3sgf

' Poonkuzhali said after , "Everybody is telling king is not well", is it true?.' *(ex.12)*

Here the anaphor atu$_i$, refers a*ntecedent which is Immediate complement clause.*

13.1 {avarkaL iraNTu peerum iraNTu peNmaNikaL!}MCL

They two psp two women.

'They both are two women.'

13.0 {atu$_i$ ennake terintu viTTatu; jothiTare! naan kuruTan illai.}MCL

It myself known ; astrologer+e I blind not

'I myself known this, Astrologer! I am not blind'. *(ex.13)*

Here the anaphor atu$_i$, refers the whole previous sentence as antecedents.

14.1 munpoolavee antac camayatil appeNkaL *kalakalavenRu cirikkum cattam$_i$*

As before that time+loc that women Laughing sound

keTTatu.

heard+3sgn

'As like before, that time Vanthiyathevan heard laughing sound of women'.

14.0 vanthiyateevan kaatukku *atu$_i$* naaraacaramaayiruntatu.

vanthiyathevan ear+dat it horrible

'For his ear it is very horrible.' *(ex.14)*

Here the anaphor is *atu*$_i$ and the antecedent is *kalakalavenRu cirikkum cattam*$_i$ *(verb phrase+noun phrase).* Here the anaphor atu$_i$, refers the verb phrase and the noun phrase as antecedents.

15.1 {naalu maathathiRkku munnal apacakunam maatiri toonRak kooTiya}NF

four month+dat before bad sign like think possibility

{oru kaariyam naTantatu.}MCL

One event held

'Four months back something happened which appears as bad sign.'

15.0 *{etoo onRu tavaRi vizhuntatu*$_i$ aanaal}CON {*atu*$_i$ unmaiyiL apacakuNam iLLai.}

something one slipped fell down; But it truth+loc bad sign not

'Something fell down but really it is not a bad sign.' *(ex.15)*

Here the anaphor is *atu*$_i$ and the antecedent is *etoo onRu tavaRi vizhuntatu*$_i$ *(adjective+qc+verb phrase+verb phrase).*

A. NOMINATIVE - OTHER CASE MARKERS OF PRONOUN:

KrishNan avan*ukku*$_i$ pena vaankinaan.

Krishnan he+dat pen bought+3sgm

'Krishnan bought pen for him.' *(ex.16)*

KrishNan$_i$ avan*ukku*$_i$ pena vaankik koNTaan.

Krishnan he+dat pen bought be+3sgm

'Krishnan bought pen for himself.' *(ex.17)*

In the ex(17) the pronoun avan*ukku*$_i$ with dative case marker refers Krishnan because of the verb koNTaan. This verb changed the meaning of the sentence. This pattern is in NOMINATIVE – DATIVE form.

B. DATIVE - OTHER CASE MARKERS OF PRONOUN

The possible occurence of sentence is given below.

KrishNan***ukku*** **avan$_i$** aNNan.

Krishnan+dat he elder brother.

'He is the elder brother of Krishnan.' *(ex.18)*

Here **avan$_i$** is the pronoun with nominative case marker which does not refer to KrishNan***ukku***$_i$. The antecedent lies outside of the sentence. This pattern is in DATIVE-NOMINATIVE form.

VIII. CONCLUSION AND FUTURE WORK

This work considered pronominal in Tamil for resolution. The features which are identified work efficiently. There are some challenges at the discourse level which are discussed in results and evaluation. These issues will be taken up as future work.

REFERENCES

[1] Hirst, Graeme, "Anaphora in natural language understanding." Berlin Springer Verlag, 1981.
[2] D. Carter, "Interpreting anaphors in natural language texts" Chisester: Ellis Horwood ltd, (1987)
[3] Hobbs, J. (1978) Resolving pronoun references. Lingua 44, 339—352.
[4] R. Mitkov, "Robust pronoun resolution with limited knowledge" In: 17th International Conference on Computational Linguistics (COLING' 98/ACL'98), Montreal, Canada, pp. 869—875.
[5] G.N.Jha, L. Sobha, D. Mishra, S.K Singh, P. Pralayankar, "Anaphors in Sanskrit" In: Proc. Second Workshop on Anaphora Resolution Johansson, C.(Ed.). (2008)
[6] A. Dhar, and U. Garain, "A method for pronominal anaphora resolution" in Bengali In: proc. 6th Int. Conf. on Natural Language Processing (ICON) at Pune, India, December. (2008)
[7] S. Lappin, and H.J Leass, "An algorithm for pronominal anaphora resolution." Computational Linguistics 20 (4), 535—561.(1994)
[8] A.K Joshi, and S. Kuhn, "Centered logic: The role of entity centered sentence representation in natural language inferencing." In: International Joint Conference on Artificial Intelligence. (1979)

[9] A.K Joshi, and S. Weinstein, “Control of inference: Role of some aspects of discourse structure - centering.” In: International Joint Conference on Artificial Intelligence, pp. 385—387.(1981)
[10] J. Lafferty, A. McCallum, and F. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: 18th International Conference on Machine Learning, pp .282--289. Morgan Kaufmann, San Francisco, USA.(2001)
[11] N.K.N Murthi, L. Sobha, B.Muthukumari,”Pronominal Resolution in Tamil Using Machine Learning Approach” The First Workshop on Anaphora Resolution (WAR I), Ed Christer Johansson, Cambridge Scholars Publishing, 15 Angerton Gardens, Newcastle, NE5 2JA, UK pp.39-50. (2007)
[12] C.Orasan, ”PALinkA: a highly customizable tool for discourse annotation.” In: proc. 4Th SIGdial Workshop on Discourse and Dialog, Sapporo, Japan, 5 – 6 July, pp. 39 – 43.(2003)
[13] L. Sobha, B.N Patnaik, “VASISTH- An Anaphora Resolution System” Unpublished Doctoral dissertation. Mahatma Gandhi University,Kottayam,Kerala. (1999)
[14] L. Sobha, P. Pralayankar, “Algorithm for Anaphor Resolution in Sanskrit” In: Proc. 2nd Sanskrit Computational Linguistics Symposium, Brown University, USA (2008)
[15] L. Sobha, “Resolution of Pronominals in Tamil”, Computing Theory and Application, The IEEE Computer Society Press, Los Alamitos, CA, pp. 475-79 .(2007)
[16] J.G. Carbonell, and R.D Brown, “Anaphora resolution: A multi-strategy approach.” In:12th International Conference on Computational Linguistics, 96—101. (1988)
[17] Hobbs, R. Jerry “Pronoun resolution” Research Report 76-1. New York: Department of Computer Science,City University of New York. 1976
[18] Hobbs, R. Jerry "Resolving pronoun references". Lingua, 44, 339-352.1978

AUTHORS PROFILE

A.Akilandeswari has received her Bachelor’s Degree in Computer Science and Master of computer applications from Bharathidasan University, Trichy, Tamil Nadu. She is a Ph.D research scholar in the CLRG group and working as research engineer at AU-KBC Research Center, Madras Institute of technology.

Sobha Lalitha Devi, is a Scientist at AU-KBC Research Center, Anna University, Chennai, India. She is heading the Computation Linguistics Research Group(CLRG). Her area of specialization is on Discourse Analysis and Computation. More about CLRG is available at http://nlp.au-kbc.org