# Translation of Pronominal Anaphora from English to Telugu Language

| T. Suryakanthi | Dr. S.V.A.V. Prasad | Dr. T. V. Prasad |
|---|---|---|
| Research Scholar, Dept. of CSE | Dean of R&D | Dean of Computing Sciences |
| Lingaya's University | Lingaya`s University | Visvodaya Technical Academy |
| Faridabad, Haryana, India | Faridabad, Haryana, India | Kavali, Andhra Pradesh, India |

*Abstract*—Discourses are linguistic structures above sentence level. Discourse is nothing but a coherent sequence of sentences. Discourse analysis is concerned with coherent processing of text segments larger than the sentence and this requires something more than just the interpretation of the individual sentences. A phenomenon that operates at discourse level includes cohesion. Text is cohesive if its elements link together. This linking can be either forward or backward. Pronominal referencing is one method for linking sentences. This paper presents the issues in translating pronominal references from English to Telugu language. This work handles resolution and generation of personal pronouns whose antecedents appear before the anaphora. An algorithm is developed for translation of pronominal references.

## I. INTRODUCTION

Bloomfield and Chomsky 1957 have defined that the sentence is the largest grammatical unit for language analysis. Halliday and Hasan (1976) threw light on the concepts like coherence and cohesion. Discourse analysis is concerned with coherent processing of text segments larger than the sentence and this requires something more than just the interpretation of the individual sentences. Machine translation refers to the task of translating text from one natural language to another with minimal human intervention.

The present machine translation system is a rule based machine translation system, where parallel grammar was developed for both source and target languages. Phrase structure grammar framework is used to develop the grammar rules for the languages. This system is able to translate text above sentence level. The text above sentence level also called the discourse text. Translation of discourse includes resolving references used in the sentences. This paper presents the resolution and evaluation of these anaphora problems in translating from English to Telugu language.

## II. PRONOMINAL REFERENCE AND ANAPHORA

A grammatical term for pronoun, which refers back to another word or phrase, is called Anaphora. Halliday and Hasan defined anaphora as the cohesion which points back to some previous item [1]. The item which refers is called anaphor and the item which is referred is called the antecedent.

Ex: 1 Ram went to fruit market. He likes apples very much.

In the above sentence 'He' is a pronoun which refers to Ram in the previous sentence. Here 'He' is an anaphor and 'Ram' is an antecedent. This is the most common type of anaphor called the pronominal anaphora. Anaphora phenomenon has two processes, resolution and generation. 'Resolution' refers to the process of determining the antecedent of an anaphor; 'Generation' is the process of creating references over a discourse entity. This work handles resolution and generation of personal pronouns whose antecedents appear before the anaphora. cataphoric relations are not taken into account in this study. The translation of third person personal pronouns from English to Telugu language has been evaluated on unrestricted corpora. The precision achieved in translating personal pronouns is above 75%. Personal pronouns can also be used as objects to refer to the antecedents which are objects of the previous sentence.

*1) Intra-sentential Anaphora* Intra-sentential anaphora has the two co-referring expressions in the same sentence [2]. The first phrase in co-reference is called the antecedent and the second one is anaphor. Intra-sentential anaphora resolution relies on syntactic, rather than discourse cues [3].
Ex: 2

> *a) When jack arrived at the party, he was drunk*
> *b) When jack arrived at the party she was drunk.*

In the above example 2a) is an ill formed sentence for an obvious constraint, that for two noun phrases to co-refer they must agree in gender, number and person. 1a violates this constraint as jack is female and a pronoun 'he' which is masculine is used to co-refer jack. So the correct usage here is using a pronoun 'she' as in example 2b)

*2) Inter-sentential Anaphora:* Co reference can occur between two different sentences. If a pronoun is used to refer a noun in the previous sentence it is called inter-sentential anaphoric reference [4]. Pronouns are used to replace nouns. Pronouns have all the features that a noun has. Pronouns carry the information called gender, number and person. They are chosen to refer a noun based on the GNP features of a noun they are referring. Personal pronouns in Telugu corresponding to English [5] [9] are shown in table 1.

TABLE I.     PERSONAL PRONOUN OF TELUGU AND ENGLISH

| Person | Singular | Plural |
|---|---|---|
| 1 | (I) nEnu (M/F/N) | (We) mEmu (M/F/N) |
| 2 | (You) nIvu (M/F/N) | (You) mIru (M/F/N) |
| 3 | (He) ataDu (M)<br>(She) Ame (F)<br>(It) adi (N) | (They) (vAru)<br>(They) (avi) |

### III.     RESOLVING THE ANAPHORA

Anaphoric resolution is of crucial importance in order to translate anaphoric expressions correctly into target language. Resolution refers to the process of identifying the antecedent of an anaphor [3]. If there are more than one noun in a sentence to which an anaphor can refer then ambiguity arises in resolving the antecedent of an anaphor. Understanding the sentence and translating them correctly requires world knowledge. Contextual understanding is required to understand and translate such sentences. Humans use more refined and flexible inference making and problem solving capabilities for interpreting these texts. If we can imbibe these processing capabilities to a machine, the accuracy of the system will be as good as a human translator.

Ex: 3

SL:  Radha and Ravi are good friends of Raju. She is a very naughty girl.

TL:  radha ravi, raju ki ma.nchi snEhitulu. Ame chAla allari pilla.

SL:  Raju is playing guitar. It is an electronic device.

TL:  Raju guitar vayinchutU unnadu. adi oka electronic parikaram

In the above two sentences the antecedents of anaphors are easily identified. In the first example 'she' refers to Radha. In the second example 'it' refers to guitar. Translation is done without any ambiguity as these anaphors have only one interpretation in Target language. In some cases there will be more than one antecedent to which an anaphor can refer.

Ex: 4

SL:     Radha bought bananas and bangles. They were very sweet.

TL:     rAdha araTi paLLu gAjulu techchinadi. avi chAla tiyyaga u.mdinavi

In the above example 'they' can refer to either bananas or bangles. GNP features of both the nouns match with GNP features of 'they'. By applying world and contextual knowledge we can understand that here 'they' refers to bananas as bangles have no taste. Translation does not incur ambiguity as both bangles and bananas are having neuter gender, either of them being the antecedent the anaphor 'they' will be translated as 'avi'.

Ex: 5

SL:  Radha bought bananas and bangles. They were yellow in color.

TL:  rAdha araTi paLLu gAjulu techchinadi. avi pachcha ra.mgu lO u.mdinavi

In the above example 'they' can refer to either bananas or bangles. GNP features of both the nouns match with GNP features of 'they'. By applying world and contextual knowledge we cannot understand whether 'they' refers to bananas as bangles as both of them can be yellow in color. Here either the author wants to express that both bananas and bangles are yellow in color or he should explicitly use the noun instead of anaphor to avoid the ambiguity.

Ex: 6

SL:  Radha came home with her friends and with some fruits to eat. They look quite tired.

TL:  rAdha tana snehitulu mariyu tinuTaku konni paLLu thO vaccinadhi. vAru chAla alisipoyi kanpadutunnaru.

In the above example 'they' can refer to either friends or fruits. Number and person features of both friends and fruits are the same but gender feature differ. The gender of friends CAN be male or female and fruits is neuter. 'They' can refer either friends or fruits. If 'they' refers to a neuter gender noun, then it will be translated as 'avi' else it will be translated as 'vAru/vALLu' and accordingly the verb suffix will change. By applying world and contextual knowledge its understood that human beings tire but fruits don't. Accordingly 'they' will be translated as 'vAru' to refer friends.

Ex: 7:

SL:  Radha came home with her friends and some fruits to eat. They were very fresh.

TL:  rAdha tana snehitulu mariyu tinuTaku konni paLLu thO vaccinadhi. vAru chala tajaga vu.mdinaru

TL:  rAdha tana snehitulu mariyu tinuTaku konni paLLu thO vaccinadhi. avi chala tajaga vu.mdinavi

Taking the same example with slight modification introduces ambiguity. In the example 'they' can refer to either friends or fruits as the adjective 'fresh' can be used for either of them. By applying world and contextual knowledge Its difficult to tell whether 'they' refers to fruits or friends.

### IV.     TRANSLATION OF ANAPHORS

Translation of anaphors involves three major steps. First step is identification of the antecedent of the anaphor by matching the features of the anaphor with the nouns of the previous in the nominative form. Second step is identification of anaphor of the target language. While translating anaphor from SL to TL the features of anaphors of SL are mapped to the anaphors of target language. If more than one entry is available for the anaphor in the bilingual lexicon then match the GNP features of the antecedent to which the anaphor is referring and anaphor of TL. Third step is verb suffix change according to the subject verb agreement rules of the target language.

- *Verb dependency on Anaphors*

English verbs are not strongly inflected. The only inflected forms are third person singular simple present in –s, a simple

past form, a past participle form, a present participle and gerund form in -ing. Most verbs inflect in a simple regular fashion. There are some irregular verbs with irregular past and past particle forms [1]. If pronoun is the subject then the auxiliary verb should agree with the number and person features of the subject.

Telugu verbs are formed by combining roots with other grammatical information. Simple verbs in their finite forms are inflected for tense followed by GNP endings or states. In order to indicate aspect and modality of verbs various auxiliaries are employed

The structure of the verb will be like *Verb stem+ Tense*

*Suffix+ GNP Suffix.* When a pronoun is the subject of a sentence, the verbs agrees in person, number, and when using third person agrees with gender also [7] [8].

The verb inflections should agree with gender and number features of the subject, noun. Though Telugu nouns have three genders and two numbers the verb suffixes change in a different way.

In singular number, feminine and neuter nouns have the same verb suffixes but masculine nouns have different verb suffixes. In plural numbers masculine and feminine nouns have same GNP endings, but for neuter nouns they differ. The suffixes for the verb 'go' are shown in the table 2.

TABLE II. SUFFIXES OF VERB 'GO' FOR DIFFERENT GNP FEATURES

| Person | Singular | | Plural | |
|---|---|---|---|---|
| | Pronoun | Verb (go/goes) | Pronoun | Verb (go) |
| 1 | I (nEnu) (M/F/N) | veLLanu | We (mEmu) (M/F/N) | veLLamu |
| 2 | You (nIvu) (M/F/N) | veLLavu | You (mIru) (M/F/N) | veLLaru |
| 3 | He (ataDu)(M) She (Ame) (F) It (adi) (N) | veLLaDu veLLi.mdi veLLi.mdi | They (vAru) They (avi) | veLLaru veLLayi |

TABLE III. VERB PATTERNS OF ENGLISH AND TELUGU

| English Pattern | Telugu Pattern | Example Englsh | Telugu Translation |
|---|---|---|---|
| **Single word verbs** | | | |
| VBD | VBD | He goes | ataDu veLLenu |
| VBZ | VBZ | We see | mEmu chUsamu |
| VBP | VBP | I left | nEnu veLLitini |
| **Two word verb Phrases** | | | |
| MD+VB | VB+MD | I will stay | nEnu u.mDa galanu |
| have/has/had+VBN | VBN+ have/has/had | I have gone. She has gone. We had gone | nEnu veLLi unnanu Ame veLLi unnadi mEmu veLLi u.mDagalamu |
| am +VBG | VBG + am | I am going | nEnu veLLuchU unnanu |
| is/are +VBG | VBG + is/are | She is going They are going | Ame veLLuchU unnadi vAru veLLuchU unnaru |
| was/were +VBG | VBG+ was/were | He was going. They were going | ataDu veLLuchU u.mDinADu vAru veLLuchU u.mDiri |
| am+ VBN | VBN + am | I am done | nEnu chEsinAnu |
| is/are +VBN | VBN+ is/are | He is released It is taken They are forgiven | ataDu viDudala chEyabaDi u.mnnaDu. Adi tIsukObaDi unnadi vAru kshami.mcha baDi unnAru |
| was/were +VBN | VBN + was/were | She was forgiven They were forgiven | Ame kshami.mcha baDi u.mDinadi vAru kshami.mchabaDi u.mDiri |
| **Three word Verb Phrases** | | | |
| MD+have+VBN | VBN + have + MD | I could have danced | nEnu Adi u.mda galanu |
| MD+be+VBG | VBG+ be+ MD | She should be arriving | Ame vachuchU u.mDa valenu |
| MD+be+VBN | VBN+ be + MD | He must be stopped | ataDu Agi u.mDa valenu |
| have+been+VBG | VBG + been + have | We have been travelling | mEmu prayANamu chEyuchU u.mDi unnamu |
| has+been+VBG | VBG + been + has | She has been travelling | Ame prayANamu chEyuchU u.mDi unnadi |
| had+been+VBG | VBG + been + had | It had been raining | ikkaDa varshi.mchuchU u.mDi u.mDagaladu |
| have+been+VBN | VBN+ been + have | I have been waited | nEnu nirIkshistU u.mDi unnanu |
| has+been+VBN | VBN+ been + has | She has been tortured | Ame vEdhi.mchabaDi u.mDi unnadi |
| had+been+VBN | VBN+ been + had | He had been tortured | ataDu vEdhi.mchabaDi u.mDi u.mDagalaDu |

| English Pattern | Telugu Pattern | Example Englsh | Telugu Translation |
|---|---|---|---|
| am+being+VBG | VBN+ being+ am | I am being groomed | nEnu lali.mchabaDi u.nTU unnanu |
| is/are+being+VBG | VBN+ being+ is/are | It is being discussed | adi tarki.mchabaDi u.nTu unnadi |
| was/were+being+VBG | VBN+being+was/were | They were being interrogated | vAru prasni.mchabaDi unTu u.mDiri |
| **Four word verb phrases** | | | |
| MD+have+been+VBG | VBG+ been+have+MD | It should have been raining | ikkDa varshi.mchuchU u.mDi u.mda valenu |
| MD+have+been+VBN | VBN+been+have+MD | It should have been rained | ikkaDa varshi.mchabaDi u.mDi u.mDa valenu |
| MD+be+being+VBN | VBN+being+be+MD | It may be being discussed. | adi tarki.mchabaDi unTu u.mDa galadu |

Basic verb phrase patterns in English and their corresponding Telugu translations are shown in below table. From the table below it can be noticed that any verb phrase in Telugu will end with VBD/ VBZ/ VBP/ MD/ have/ has/ had/ am/ is/ are/ was/ were. Depending on the GNP features of the anaphor the last word of a verb phrase should change its suffix [8] [9].

## V. ANAPHORA RESOLUTION ALGORITHM

The algorithm identifies noun phrase (antecedents) of personal pronouns in English. This work is mainly concentrated on identifying inter-sentential antecedents and is applied to syntactic analysis. Certain constraints and heuristic rules are applied to get a possible solution for an anaphor. Constraints are the GNP agreements of anaphor and the antecedent [3].

*Algorithm*

*Step 1: Get the anaphor from a sentence, i.e a word with POS tag as PP. let it be P1*
*Step 2: Extract the GNP features of P1 from lexical DB*
*Step 3: Search for words with NN, NNS, NNP, and NNPS as their POS tags. Let them be N1, N2...*
*Step 4: Extract the GNP features of N1, N2...*
*Step 5: If NP features of P1=3S then*
  *direct translation*
  *else if NP features of P1=3P then*
  *Match the GNP features of P1with N1, N2..*
    *If exactly one match found, N1*
      *then go to Translation module*
    *else if more than one match*
      *Apply world and contextual*
*knowledge to disambiguate.*
*Step 5: Translation module:*
  *Get the TL anaphor corresponding to SL anaphor*
   *from the bilingual lexicon. Let t hem be*
   *TP1, TP2...*
  *Match gender feature of TP1, TP2... with N1*
    *If Exact match found, TP1*
     *Successful translation*
    *else*
    *no match found for the anaphor.*
*Step 6: Change the verb suffix according to the GNP features of TP1.*

- *Explanation of Algorithm with examples*

Ex: 8

SL:  Students came to the zoo. *They are* watching birds

TL: pillalu ja.mtu pradarshana shAla ki vachiri. *vAru* pakshulanu chUchu *chunnaru*

Ex: 9

SL:  Monkeys are in the zoo. *They are* doing mischief

TL: kotulu ja.mtu pradarshana shAla lo vunnavi. *avi* allari cheyu *chunnavi*

In example 1 'they' refers to 'students'. The GNP features of students being (M/F, P, 3), 'they' is translated as 'vAru' and accordingly 'are' is translated as 'chunnaru'. In example 2 'they' refers to monkeys. The GNP features of monkeys being (N,P, 3) 'they' is translated as 'avi' and accordingly 'are' is translated as 'chunnavi'.

For readability purpose the Telugu script in example sentences is transliterated into Roman English using the schema given in Appendix 1.

## VI. CONCLUSION

The algorithm was implemented to translate anaphoric expressions. Personal pronouns whose antecedents precede them were translated successfully.

This work can be extended to deal with cataphoric expressions. This work can be extended to deal with anaphoric expressions having indefinite and reflexive pronouns.

## APPENDIX – 1

Schema for transliterating Telugu as English:

Vowels : అ (a) ఆ (A,aa) ఇ (i) ఈ (I,ii,ee) ఉ (u) ఊ (U,oo) ఋ (RRi) ౠ (RRI) ఎ (e) ఏ (E) ఐ (ai) ఒ (o) ఓ (O) ఔ (au,ou) ;

Mathras : ా (A,aa) ి (i) ీ (I,ii,ee) ు (u) ూ (U,oo) ృ (RRi) ౄ (RRI) ె (e) ే (E) ై (ai) ొ (o) ో (O) ౌ (au,ou) ; Anusvara, Visarga and Bindus : ం (.n,.m) ః (H)

Consonants: క (k,q) ఖ (kh,K) గ (g) ఘ (gh,G) ఙ (~N) చ (ch) ఛ (chh,Ch,CH) జ (j) ఝ (jh,Jh,JH) ఞ (~n) ట (T) ఠ (Th,TH) డ (D) ఢ (Dh) ణ (N) త (t) థ (th) ద (d) ధ (dh) న (n) ప (p) ఫ (ph,f) బ (b) భ (bh) మ (m) య (y) ర (r) ల (l) ళ (L) వ (v,w) శ (sh,S) ష (Sh,SH) స (s) హ (h) ;

Extended Consonants: ఙ (J) క్ష (x) జ్ఞ (GY) ఱ (R)

REFERENCES

[1] Halliday, M.A.K and Hasan, R.1976, Cohesion in English. London: Longman

[2] Webber, B and Reiter, R, Anaphora and Locial form: On Formal Meaning Representations of Natural Language In Proceedings of the fifth IJCAI, Pages 121-131. Cambridge, MA 1977

[3] Shalom Lappin and Herbert J. Leass An Algorithm for Pronominal Anaphora Resolution Journal of Computational Linguistics, Vol. 20, Number 4, 1994

[4] Sidner, C.L, Focusing for Interpretation of Pronouns. Journal of Computational Linguistics 7: 217-231, 1981

[5] English Verbs, Wikipedia, available at http://en.wikipedia.org/wiki/English_verbs

[6] Dr. Divakarla Venkatavadhani, Telugu in Thirty Days, Dakshina Bharat Press, 1976

[7] Albert Henry Arden, A Progressive Grammar of the Telugu Language With Copious Examples And Exercises. India: S.P.C.K Press, 1905

[8] Krishnamurti, B., A Grammar of Modern Telugu. 1985, Delhi; New York: Oxford University Press.

[9] Brown, C.P, The Grammar of the Telugu Language.1991, New Delhi: Laurier Books Ltd.