

IMPORTANCE OF VERB SUFFIX MAPPING IN DISCOURSE TRANSLATION SYSTEM

Suryakanthi Tangirala

Faculty of Business, University of Botswana, Gaborone, Botswana

ABSTRACT

This paper discusses the importance of verb suffix mapping in Discourse translation system. In discourse translation, the crucial step is Anaphora resolution and generation. In Anaphora resolution, cohesion links like pronouns are identified between portions of text. These binders make the text cohesive by referring to nouns appearing in the previous sentences or nouns appearing in sentences after them. In Machine Translation systems, to convert the source language sentences into meaningful target language sentences the verb suffixes should be changed as per the cohesion links identified. This step of translation process is emphasized in the present paper. Specifically, the discussion is on how the verbs change according to the subjects and anaphors. To explain the concept, English is used as the source language (SL) and an Indian language Telugu is used as Target language (TL).

KEYWORDS

MT: Machine Translation, SL-Source language, TL: target Language, POS-Parts of Speech, GNP-Gender Number Person.

1. INTRODUCTION

Language is the medium of communication. Language apart from being a communication medium is a powerful source of information exchange. Different people use different languages. Though English is the globally accepted language, mostly people understand the things better in their native language. In this era of globalization, information sharing is very important. If machine translators are made, languages will not be a barrier any more. Web is becoming multilingual and the need for tools and techniques for automatic processing of languages is evident. NLP is seen as the subject dealing with such problems.

English to Telugu translator can be used in automatic translations of web. Though English is very much adopted in India, less than 5% of the population understands the language and people will understand the things better if they were told in their native language. English to Telugu translator will help people understand the works written in English in a better way. These translators will be very helpful for communication among people and for learning English.

In any translation, whether human or automated, the meaning of a text in the source language including the context must be completely translated to its equivalent meaning in the target language's translation. It is not a straight forward deal as it appears. According to Claude Bedard

“Text has its own organization and is filled with pointers that relate sentences and words into a broader picture, and that a proper translation should respect this fact” [1]. Discourse Translation is never a mere word-for-word substitution. Discourses are texts above sentence level. When a pronoun in the second sentence is referring to a subject in the first sentence it cannot be translated as a separate sentence instead while translation the first and second sentences should be interpreted as a whole and not as individual sentences[2]. The translation process involves identifying the antecedents of the anaphors which is resolution and creating the references over the discourse entity which is termed as generation [3, 4]. After resolving the anaphors the next step is to map the verbs to agree with the GNP features of the anaphors.

Discourse oriented MT makes the translations more natural in MT systems. Paragraph-by-paragraph MT seems to be a complicated task for practical needs. It involves the complete understanding of the paragraph, the determination of discourse topics, goals, intentions, so that the output can be produced in accordance with the respective discourse rules and purposes [5].

A discourse machine translation system performs a series of steps like tokenization, POS (parts of speech) tagging, parsing, reordering, reference resolution and finally, verb suffix mapping to achieve meaningful translations preserving the context. In this paper we discuss the importance of verb suffix mapping.

2. VERB SUFFIXES IN TELUGU AND THEIR VARIATIONS

A verb expresses an action or state of being. 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

Ex: SL: Sita is singing

TL: సీత పాట పడుచున్నది

(Transliteration of Telugu script to English): sIta pATa pADuchunnadi

Verbs in Telugu are inflected for gender, number, person and tense. The structure of a verb is given in Fig 1.

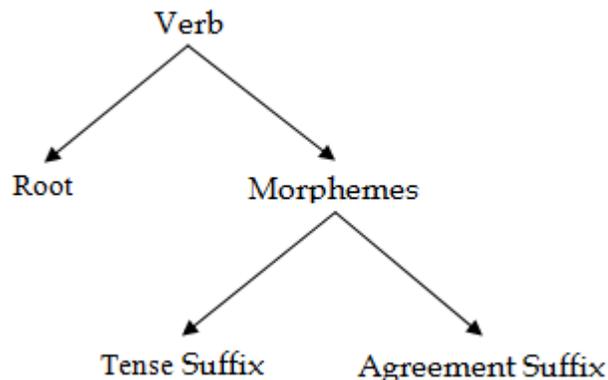


Fig. 1: Structure of Telugu Verb

pADuchunnADu is a verb whose root verb is 'pADu'. 'chunnADu' is a suffix added to the main verb to indicate the tense and feature agreement. 'chunna' indicates the present tense and 'Du' indicates the GNP as Male, singular and third person. The structure of the verb 'pADuchunnADu' is shown in Fig 2. Similarly pADuchunnadi is a verb which inflects for tense and GNP using the suffixes, 'chunna' and 'di', Fig 3.

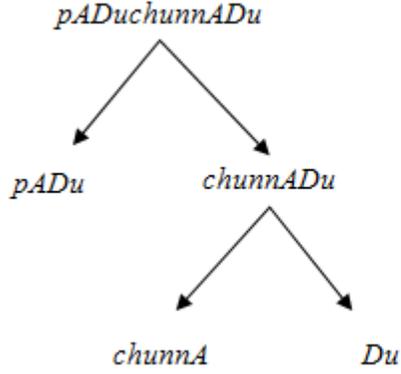


Fig. 2: pADuchunnADu

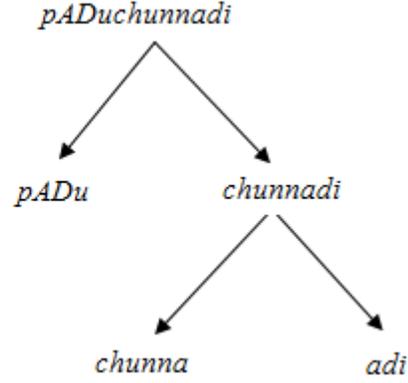


Fig. 3: pADuchunnadi

2.1 Subject Verb Agreement in English and Telugu

Subject-verb agreement is found in many languages, yet the degree of agreement varies considerably. In English, correspondence of a verb with its subject lies in person and number features. In Telugu language, the verb agrees with subject in gender, number and person i.e. when the subject is a third person personal pronoun, the verb agrees with subject in gender number person for remaining pronouns as subjects the verb agrees with them in person and number features only. Telugu language has got its own set of agreement rules.

- In Telugu language a finite verb exhibits agreement with nominative form of a noun in gender, number and person

Ex: SL: Boy is singing

TL: అమ్మాయి పాడుచున్నది.

Transliteration: ammAyi pADuchunnadi.

SL: Girl is singing

TL: అబ్బాయి పాడుచున్నాడు.

Transliteration: abbAyi pADuchunnADu

- All non-pronominal subjects are considered to be in 3rd person. When subject is in third person and singular in number, both feminine and neuter genders have same verb suffixes

and for masculine gender verb suffix would be different. When subject is in third person and plural in number, both masculine and feminine genders have same verb suffixes but for neuter gender the suffixes of verbs differ as shown in Table 1.

Table 1: Change of Verb suffixes with GNP features of Subject

English	Telugu	Gender	Number	Verb Suffix
Girl is singing	అమ్మాయి పాడుతున్నది	Girl, F	S	అది
Radio is singing	రేడియో పాడుతున్నది	Radio, N	S	అది
Boy is singing	అబ్బాయి పాడుతున్నాడు	Boy, M	S	ఆడు
Boys are eating	అబ్బాయిలు పాడుతున్నారు	Boys, M	P	ఆరు
Girls are eating	అమ్మాయిలు తింటున్నారు.	Girls, F	P	ఆరు
Dogs are eating	కుక్కలు తింటున్నాయి	Dogs, N	P	ఆయి

3. VERB SUFFIX MAPPING

Agreement of gender number person (GNP) is realized in two cases in subject verb agreement and agreement of anaphoric pronoun with its antecedent [6]. After Anaphora generation next step is verb suffix mapping. If the anaphora is at subject position, the verb of that sentence should agree with the anaphora. In Telugu language verb inflects for gender, number and person. Subject-Verb agreement rules of Telugu are discussed in detail in section 1. After anaphora generation, grammatical gender and number information of the pronoun are required for verb suffix change.

Table 2: Third person Pronouns marking gender in English and Telugu

Pronouns in English	Mark gender	Pronouns in Telugu	Mark gender
He	Yes	అతడు	Yes
She	Yes	ఆమె	Yes
It	Yes	అది	Yes
They	No	వారు, అవి	Yes

3.1 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 participle forms. 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 [7].

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 [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 3.

Table 3: Suffixes of verb ‘go’ for different GNP features

Person	Singular		Plural	
	Pronoun	Verb (go/goes)	Pronoun	Verb (go)
1	I (నేను) (M/F/N)	వెళ్ళాను	We (మేము) (M/F/N)	వెళ్ళాము
2	You (నీవు) (M/F/N)	వెళ్ళవు	You (మీరు) (M/F/N)	వెళ్ళారు
3	He (అతడు)(M)	వెళ్ళాడు	They (వారు)(M/F)	వెళ్ళారు
	She (ఆమె) (F)	వెళ్ళింది		
	It (అది) (N)	వెళ్ళింది	They (అవి) (N)	వెళ్ళాయి

3.2 Verb Patterns

Basic verb phrase patterns in English and their corresponding Telugu translations were shown in table 4. It can be noticed that any verb phrase in Telugu will end with VBD (Past tense)/ VBZ (3rd person singular)/ VBP (non 3rd person singular)/ MD (Modal)/ have/has/ had/ am/ is/ are/ was/ were, Table 4. Depending on the GNP features of the anaphor the last word of a verb phrase should change its suffix.

Table 4.9: Verb Patterns in English and Telugu

English Pattern	Telugu Pattern	Example English	Telugu Translation
Single word verbs			
VBZ	VBZ	He goes.	అతడు వెళ్ళాడు.
VBP	VBP	We see.	మేము చూశాము.
VBD	VBD	I left.	నేను వెళ్ళితిని.
Two word verb Phrases			
MD+VB	VB+MD	I will stay.	నేను ఉండ గలను.
have/has/had+VBN	VBN+ have/has/had	I have gone. She has gone.	నేను వెళ్ళి ఉన్నాను. ఆమె వెళ్ళి ఉన్నది.
am +VBG	VBG + am	I am going	నేను వెళ్ళుచూ ఉన్నాను.
is/are +VBG	VBG + is/are	They are going	వారు వెళ్ళుచూ ఉన్నారు.
was/were +VBG	VBG+ was/were	He was going.	అతడు వెళ్ళుచూ ఉండినాడు.
am+ VBN	VBN + am	I am done	నేను చేసినాను.
is/are +VBN	VBN+ is/are	He is released	అతడు విడుదల చేయబడి ఉన్నాడు.
was/were +VBN	VBN + was/were	She was forgiven.	ఆమె క్షమించబడి ఉండినది.
Three word Verb Phrases			
MD+have+VBN	VBN + have + MD	I could have danced	నేను ఆడి ఉండ గలను.
MD+be+VBG	VBG+ be+ MD	She should be arriving	ఆమె వచ్చుచూ ఉండ వలెను.
MD+be+VBN	VBN+ be + MD	He must be stopped	అతడు ఆగి ఉండ వలెను.

English Pattern	Telugu Pattern	Example English	Telugu Translation
have+been+VBG	VBG + been + have	We have been travelling	మేము ప్రయాణము చేయుచూ ఉండి ఉన్నాము.
has+been+VBG	VBG + been + has	She has been travelling	ఆమె ప్రయాణము చేయుచూ ఉండి ఉన్నది.
had+been+VBG	VBG + been + had	It had been raining	ఇక్కడ వర్షించుచూ ఉండి ఉండగలదు.
have+been+VBN	VBN+ been + have	I have been waited	నేను నిరీక్షిస్తూ ఉండి ఉన్నాను.
has+been+VBN	VBN+ been + has	She has been tortured	ఆమె వేధించబడి ఉండిఉన్నది.
had+been+VBN	VBN+ been + had	He had been tortured	అతడు వేధించబడి ఉండి ఉండ గలడు.
am+being+VBG	VBN+ being+ am	I am being groomed	నేను లాలించబడి ఉంటూ ఉన్నాను.
is/are+being+VBG	VBN+ being+ is/are	It is being discussed	అది తర్కించబడి ఉంటూ ఉన్నది.
was/were+being+V BG	VBN+being+was/w ere	They were being interrogated	వారు ప్రశ్నించబడి ఉంటూ ఉండిరి.
Four word verb phrases			
MD+have+been+V BG	VBG+ been+have+MD	It should have been raining	అక్కడ వర్షించుచూ ఉండి ఉండ వలెను.
MD+have+been+V BN	VBN+been+have+ MD	It should have been rained	అక్కడ వర్షించబడి ఉండి ఉండ వలెను.
MD+be+being+VBN	VBN+being+be+M D	It may be being discussed.	అది తర్కించబడి ఉండి ఉండ గలదు.

4. VERB SUFFIX DEPENDENCY ON GNP FEATURES OF NOUN

Change the verb suffix according to the GNP features of a Noun and corresponding Pronoun.

Ex:1 SL: Students came to the zoo. They are watching birds.

TL: పిల్లలు జంతు ప్రదర్శన శాల కి వచ్చిరి. వారు పక్షులను చూచు చున్నారు.

pillalu ja.mtu pradarshana shAla ki vachchiri. vAru pakshulanu chUchu chunnAru.

Ex:2 SL: Monkeys are in the zoo. They are doing mischief

TL: కోతులు జంతు ప్రదర్శన శాల లో ఉన్నవి. అవి అల్లరి చేయు చున్నవి.

kOtulu ja.mtu pradarshana shAla lo unnavi. avi allari cheyu chunnavi.

Ex:3 SL: AC is not working properly. It is making loud noise.

TL: ఏసీ పని చెయ్యట లేదు. అది గట్టిగా చప్పుడు చేయు చున్నది.

EsI pani cheyyuTa lEdu. adi gaTTigA chappuDu chEyu chunnadi.

Ex:4 SL: AC is not working properly. Can the engineer repair it?

TL: ఏసీ పని చెయ్యడము లేదు. ఇంజనీరు దాన్ని బాగు చేయ గలడా?

EsI pani cheyyaDamu lEdu. i.mjanIru dAnni bAgu chEya galaDA?

In example 1 'they' refers to 'students'. The GNP features of students being (M/F, P, 3), 'they' is translated as 'వారు' and accordingly verb 'are' is translated as 'చున్నారు'. In example 2 'they' refers to monkeys. The GNP features of monkeys being (N, P, 3) 'they' is translated as 'అవి' and accordingly verb 'are' is translated as 'చున్నవి'.

In examples 3 and 4, 'it' is the anaphor referring to a third person, singular pronoun of neuter gender, 'AC'. The grammatical role of 'it' in both examples differ. In example 3 the anaphor is at subjective position and in example 4 the anaphor is at objective position. In English language same pronoun 'it' will be used at both subjective and objective positions. But in Telugu language two different pronouns are used for different grammatical roles. 'అది' is used for subjective and 'దాన్ని' is used for objective role. Consequently the verbs in the two sentences are చేయు చున్నది, చేయ గలడా respectively.

5. CONCLUSION

Translating texts may not be a new concept but translating texts preserving the context is an area of research which has been explored very little. Generally, the translations lack the flair of SL because of the lexical and syntactical differences of the language pairs involved in the translation. Discourse oriented MT makes the translations more natural in MT systems. The present paper discusses the importance of verb suffix mapping in the Anaphora resolutions and generation from English to Telugu language. The concept can be applied to many of the foreign languages.

REFERENCES

- [1] Claude Bedard (2008), "Suddenly its Discourse Analysis", Language Technology 13, May-June 1989
- [2] T. Suryakanthi, Kamlesh Sharma (2015) "Discourse Translation from English to Telugu" In Proceedings of the Third International Symposium on Women in Computing and Informatics (WCI-15), ACM publishers
- [3] Jes'us Peral, Antonio Ferr'andez (2003) "Translation of Pronominal Anaphora between English and Spanish: Discrepancies and Evaluation" Journal of Artificial Intelligence Research Vol.18 pp. 117-147
- [4] T. Suryakanthi, Dr. S.V.A.V Prasad, Dr. T.V Prasad, "Translation of Pronominal Anaphora from English to Telugu Language", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 4, 2013
- [5] Hauenschild C., (1988) Discourse structure - some implications for Machine Translation, Proc. of Conf. on New Directions in Machine Translation, Budapest, August 18-19 Dodrecht-Holland
- [6] Abdel-Aal Attia Mohammed, (2002) "Implications of the Agreement Features in Machine Translation", M.A Thesis, Faculty of Languages and Translation, Al-Azhar Univ.
- [7] Krishnamurti, Bh., (1985) A Grammar of Modern Telugu, Oxford Univ. Press, New York.
- [8] Henry Arden Albert, (1905) A Progressive Grammar of the Telugu Language With Copious Examples And Exercises, S.P.C.K Press, India.

AUTHORS

Dr. S. Tangirala earned her master's degree in computer applications in 2006 from Andhra University, Visakhapatnam, India and doctoral degree in 2014 from Lingaya's University, Faridabad, India. She has worked for around 2 years in software industry and has been teaching for 5 years at University Level. She was Assistant Professor of Computer Applications at Lingaya's University and worked as Fellow at Botho University, Gaborone, Botswana. Currently she is working with University of Botswana. She has 17 research papers to her credit in various international conferences and journals. Her current research interests include Artificial Intelligence, Natural Language Processing, Machine Translation, Big data analytics and Theory of automata.

