JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Odia Verb Sense Disambiguation Model – An Extension to Paninian Grammar Framework

[1]Gayatri Dey, [2]Prof. Hima Bindu Maringanti

[1]Research Scholar, [2]Professor
[1]Deptt. of Computer Applications,
[1]Maharaja Sriram Chandra Bhanja Deo University, Baripada, Mayurbhanj, Odisha, India

Abstract: Odia is a free-word order but an agglutinative language. In this language, the meaning of a sentence does not depend upon the position of the words contented, but depends upon the morphemic behavior of words. Verbs are the important morpheme category as they form the pivot of a sentence which is needed for language understanding. This paper describes the semantic nature of Odia verbs with respect to their syntactic behavior, the basis of which is Paninian Grammar Framework. The syntactico-semantic model proposed here shows the inter- relationship mapping between the syntax and semantics of Odia verbs which helps us to infer the meaning/semantics with verb-specific idiosyncratic properties. The idiosyncratic properties are the tense, aspect, modality (TAM) and role of the verb etc. Tense, aspect and modality could be used to label the verbs with a TAM label. Here the tense refers to the occurrence of an event in present, past or future. The aspect is the time of occurrence of an event in present, past or future time. Modality speaks whether the sentence containing the verb is a general statement or a command. Initially, the verb is classified as transitive or intransitive. Additionally, the role defines whether the verb form is finite, non-finite or used as a noun. Finite verb means it can stand alone as a complete sentence and is the syntactic locus for tense, aspect and modality. Non-finite verbs are the dependent verbs. Lastly, verbs are sometimes used in noun form in sentences. The present work thus focuses on the sense disambiguation of Odia verbs, annotating them with morphologically rich features, helping other Indian Language researchers in word sense disambiguation, machine translation, information extraction and cognitive linguistics, working in any other natural language**.**

IndexTerms **- Paninian Grammar Framework, morpheme, syntactico-semantic, modality, TAM, Word sense disambiguation.**

I. INTRODUCTION

Odia is a free-word order agglutinative Indo-Aryan language [11][1]. In this language the meaning of sentence does not depend upon position of words contended, but depends upon the morpheme behavior of words. Verbs are the important morpheme category as they form pivot of the sentence which is needed for language understanding. So to infer the meaning of a sentence there is to understand semantic nature of Odia verb with respect to its syntactic behavior. The syntactic behavior of verb depends upon its idiosyncratic properties such as tense, aspect, modality and the role of the verb [13]. Here the tense refers to the occurrence of an event in present, past and future. Aspect refers to the time of occurrence of an event in present, past and future. Modality speaks whether the sentence containing the verb is a general statement and a command/request. Initially, verb is of two types: transitive and intransitive. This typological differentiation depends upon the verb's requirement of object to express a complete thought/meaning or not. Transitive verb requires an object to do action on it and intransitive verb work independently without requiring an object. Additionally, the role defines whether the verb form is finite, non-finite or used as noun. Finite verb (khaauchhi, basichhi) means it can stand alone as a complete sentence and is the syntactic locus for tense, aspect and modality. Non-finite verbs (khaai, basi) are the dependent verbs. Lastly, verbs are sometimes used in noun form (khaaibaaku, basibaku) in sentences. This paper describes semantic nature of Odia verb for verb sense disambiguation using Paninian Grammar Framework (PGF) [1].

II. BACKGROUND

The paper [9] describes an Object Oriented model for OriNet System using Object Oriented programming and programming principles of Java to represent and retrieve Odia language related information.

This paper [10] deals with the Design of Odia Morphological Analyzer (OMA) system according to Object Oriented Analysis to increase its reusability, robustness and extensibility. Here some examples the Morphological Analysis(OMA) with the WordNet named OriNet for Indian Language is reflected. Here the development is based on the syntactic approach of Sanskrit. So the technology involved here can be extended to any other Indian language.

The paper [14] proposes to design a morphological analyser for Odia Verbal forms in a deterministic Finite State Automaton. This model can provide lexical, morphological and syntactic information for each lexical unit in the analyzed verbal form. Here finite state machine accepts valid sequence of morphemes in a verbal form and rejects invalids ones. Here the morphological

analyzer will help us to build a computational lexicon structured as a list of stems and affixes with a representation of the morph tactics and also can be used for designing a morphosyntatic analysis for each word in unrestricted Odia Text.

The paper [6] deals with developing a Morphological Analyzer for Odia language following the paradigm approach. Here the paradigm defines all the words from a given stem and also provides a feature structure associated with every word. The paradigm have been created using XML based morphological dictionary from the Lttoolbox packages. In this work inflectional morphology is only handled by morph analyzer and the work is going on for nouns, verbs, indeclinable and adjectives. In future it can be extended for handling compound words and derivational morphology.

This paper [12] provides detailed semantic-syntactic classification of Hindi Verbs that provides a range of linguistic property information of the verb. It is a base for semantico-syntactic investigation of Hindi verbs and their argument structure.

In this work [5], the Paninian framework is used in Syntactico-Semantic phase of Odia language understanding. Here Lexical and Morphological including Stemming and Tagging functions have been implemented. But in the Syntactico-Semantic phase semantics are not enough because language understanding is much deeper and richer than just surface meanings or word meanings. Further processing would require the building of Lakshan charts for verb sense disambiguation.

This paper [1] describes the Paninian framework applied to modern Indian languages, giving an account of the relation between surface form (vibhakti) and semantic (karaka) roles. The mapping is elegant and compact. This also explains active-passives and complex sentences.

III. METHODOLOGY AND CONTRIBUTORY WORK

In this work, paninian parser of paninian grammar framework (PGF) is used to design Odia verb sense disambiguation model. The paninian parser is given in Fig.1

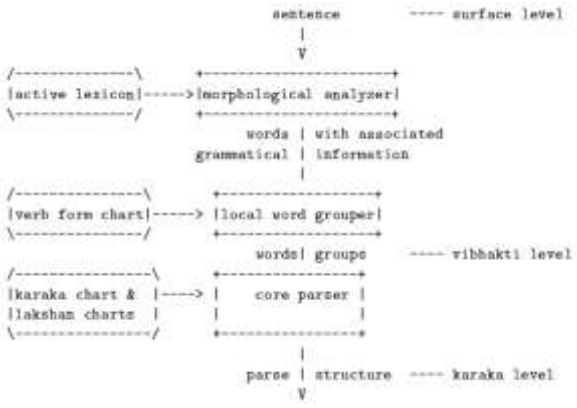

Fig.1 Paninian Parser [1]

Paninian parser is based on information theoretic considerations where at each stage of processing, the right amount of information is extracted. The first part of parser takes care of morphology. Here, a lexicon or a dictionary is needed to be looked up to retrieve associated grammatical information for each word in the input sentences. Then words are grouped together according to their morpheme category. Core parser is used to identify karaka relations among word groups and sense of words [2]. In this paper Odia verb sense disambiguation model given is based on this concept of paninian parser is shown in Fig. 2

In the first part of this model, Lexical analysis or Tokenization is done to extract individual Odia words/terms by removing punctuation marks from an input sentence. This process is followed by morpheme identification phase. Morphological Analysis is used to analyze and generate different types of word forms, which are called Morphemes. In this morpheme identification phase, stemming is the process to separate out each individual word/term from their affixes (prefix, suffix, and infix) and to find out the linguistic root of the word. Here a local word dictionary/lexicon is used. Part-of-Speech-Tagging (POS tagging or POST) is the process of marking up of a word with its corresponding part of speech, such as noun, verb, adjective, adverb, pronoun, preposition etc [8]. In this work we have only considered about the verbs, as verbs play a very important role to find out the actual meaning of the sentence. Is has relation with every other word of the specific sentences. So to identify the actual ‘used sense’ of the verb in the sentence, POST is not enough which we can understand from below examples (Ex1-Ex5)

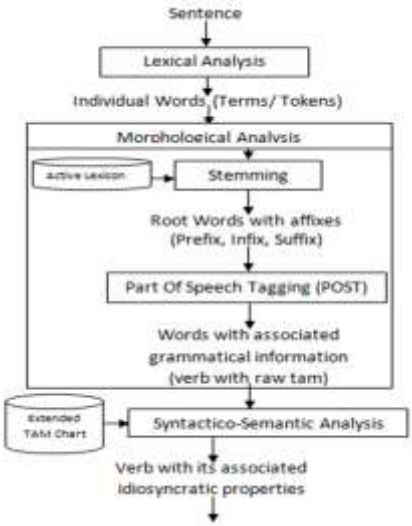

Fig 2: Odia Verb Sense Disambiguation Model

Ex1. ରାମ ଭାତ ଖାଉଛି (ଖା-ଉଛି)
Raama bhaata khaauchhi (khaa-uchhi)
Raama-Rice-is Eating
Raama is eating rice

Ex2. ରାମ ଭାତ ଖାଇଲା(ଖା-ଇଲା)
Raama bhaata khaailaa (khaa-ilaa)
Raama-Rice-ate
Raama ate rice

Ex3. ରାମ ଭାତ ଖା (ଖା-Null)
Raama bhaata khaa (khaa-Null)
Raama-Rice-Eat
Raama eat rice

Ex4. ରାମ ଭାତ ଖାଇ(ଖା-ଇ) ଟିଭି ଦେଖୁଛି
Raama bhaata khaai (khaa-i) TV dekhuchhi (dekh-uchhi)
Raama-Rice-after eating-TV-is watching
Raama is watching TV after eating rice.

Ex5. ରାମ ଭାତ ଖାଇବାକୁ(ଖା-ଇବାକୁ) ବସିଛି
Raama bhaata khaaibaaku (khaa-ibaku) bacichhi (bas-ichhi)
Raama –Rice-to eat-is sitting
Raama is sitting to eat rice

In the above examples ‘khaa’ is the transitive verb, which is used with its different inflected/influenced forms (khaauchhi, khaailaa, khaai, khaaibaku) and stand alone form (khaa) which we can know from POST. But POST cannot help us to identify the semantic behavior of the verb in the sentences. So to know the semantic behavior of the verb we have to identify the syntax or architecture of each influenced verb form with its idiosyncratic properties. The idiosyncratic properties influence the verb to show it’s meaning/semantic with respect to is syntactic or architectural behavior. The syntactical or architectural representations of each inflected verb form in the above examples are given below:

ଖାଉଛି (Khaauchhi)-------- ଖା + ଉଛି (khaa + uchhi)

ଖାଇଲା (Khaaila)---------- ଖା + ଇଲା (khaa + ila)

ଖା (khaa)----------ଖା + Null (khaa + Null)

ଖାଇ (khaai)---------- ଖା + ଇ (khaa + i)

ଖାଇବାକୁ (khaaibaaku)------ଖା + ଇବାକୁ (khaaa + ibaaku)

b has two parts. First part is the root verb and second suffix part is the TAM label which represents different idiosyncratic properties of the verbal form such as tense, aspect, modality and role of the verb in that sentence . This TAM maps with the TAM labels of extended TAM chart given in Table I to identify tense, aspect, modality and role of respective verb. Here extended term is used before TAM chart because it contains little extra features to represent the role of the verb.

Table I: Extended TAM Chart

TAM Label	Tense	Aspect	Modality
ଏ, ଓ, ଅ, ଆଁ	Present	Simple	Statement
ଉଛି, ଉଛୁ, ଉଛନ୍ତି	Present	Continuous	Statement
ଇଛି, ଇଛ, ଇଛୁ	Present	Perfect	Statement
ଇଲି, ଇଲୁ, ଇଲେ	Past	Simple	Statement
ଉଥିଲି, ଉଥିଲୁ, ଉଥିଲେ	Past	Continuous	Statement
ଇଥିଲି, ଇଥିଲୁ, ଇଥିଲେ	Past	Perfect	Statement
ଇବି, ଇବା, ଇବ, ଇବେ	Future	Simple	Statement
ଉଥିବି, ଉଥିବୁ, ଉଥିବେ	Future	Continuous	Statement
ଇଥିବି, ଇଥିବେ, ଇଥିବୁ	Future	Perfect	Statement
ଠ, ଆନ୍ତୁ, ନ୍ତୁ	Present	Simple	Command/Request
ଇବାକୁ	Null	Null	Noun
ଇ	Null	Null	Nonfinite
	Null	Null	Nonfinite

For the examples Ex1 ‘uchhi’ from verbal form ‘khaauchhi’ maps with the extended TAM chart to assign its tense , aspect, modality and its role to its verb. Here present tense and continuous aspect for ‘uchhi’ represent the verbal activity based on the verb ‘khaa’ is continuing in present time. Means the process of eating is going on. In modality the “statement’ represents the verb is a finite verb and the sentence containing the verb is a complete statement or a complete sentence. Similarly the sentence containing the verb ‘khaailaa’ in Ex2 is a complete sentence and the action eating is completed in past time. In Ex3 the verb ‘khaa’ is used in present time with command/request mode, it is basically shown in spoken text. In Ex.4 raw tam ‘i’ represents ‘khaa’ as nonfinite verb with no TAM and it cannot stand alone without the help of second verb ’basichhi’. And in the Ex5 ‘ibaaku’ represents the verb as noun. The software representation of this Odia Verb Sense Disambiguation model is given in Fig. 3.

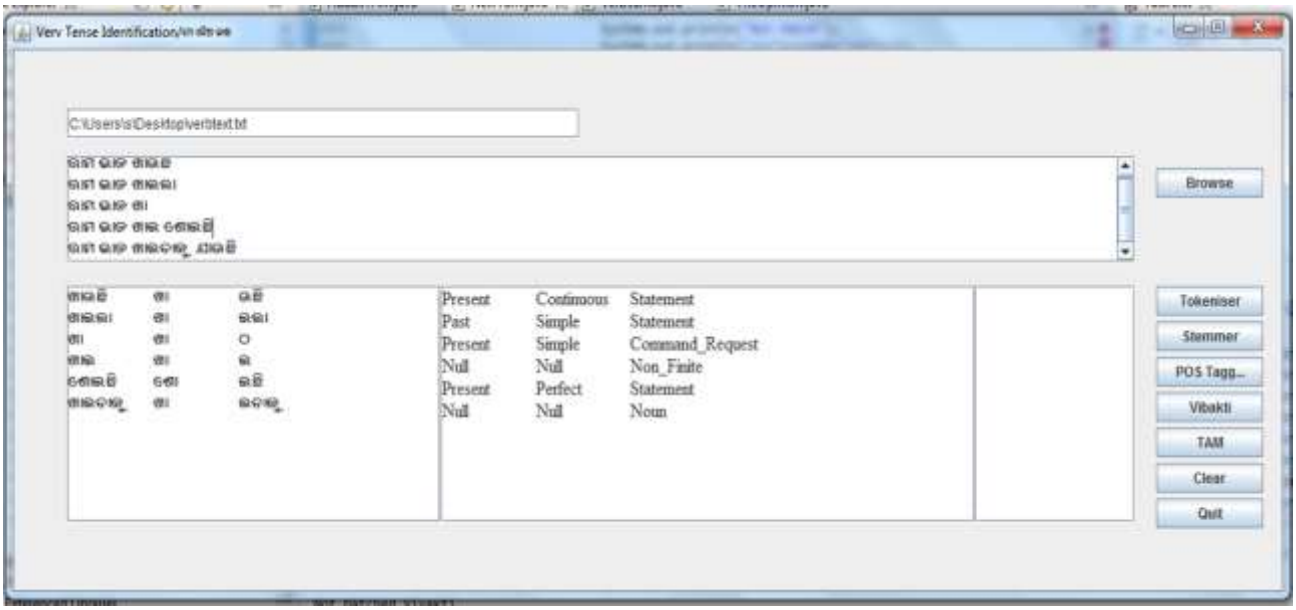

Fig. 3: Odia Verb Sense Disambiguation Model

IV. CONCLUSION

In this paper we have designed Odia verb sense disambiguation model using Paninian Grammar Framework. Here, we have focused on verb’s tense, aspect, modality and nonfinite nature of dependency. It is also focused on how verb changes its morpheme behavior depending on its suffixes. Further processing is required to build Lakshan chart for extended verb sense disambiguation and left as future work.

V. ACKNOWLEDGMENT

I want to thank our Maharaja Sriram Chandra Bhanja Deo University for providing technical and moral support to develop the project.

REFERENCES

[1] Bharati and Sangal, R., 1993, “Parsing Free Word Order Languages in the Paninian Framework”, ACL '93 Proceedings of the 31st annual meeting on Association for Computational Linguistics, Pages 105-111, (1993).

[2] Bharati, A., Chaitanya, V. and Sangal, R, 1996, "Natural Language Processing: A Paninian Perspective", Page 85-87, Prentice-Hall of India, New Delhi.

[3] Bharati, Bhatia, M., Chaitanya, V. and Sangal, R., 1996, February “Paninian Grammar Framework Applied to English”, Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, (1996).

[4] Dey, G. and Maringanti H. B., 2014, “Thematic Analysis for Odia Language”, Proceedings of National Conference on Current Trends in Computing (NCCTC-2014) 23rd -24th March 2014, pp.49-54, ISBN No. 978-3-642-24819-6.

[5] Dey,G. and Maringanti, H. B., 2014, "Paninian Framework for Odia Language Processing", National Conference on Indian Language Computing (NCILC 2014), 1st and 2nd February (2014).
[6] Jena, Choudhery, S., Choudhry, H. and Sharma, D. M., 2011 "Developing Oriya Morphological Analyzer Using Lt-toolbox", ICISIL, Springer,Communications in Computer and Information Science 139, Page 124-129, (2011)
[7] Kumar, B., Maringanti, H. B., 2010, and Asawa, K., "Adaptive Pragmatic Analysis of Natural Language", Proceedings of International Conference on Intelligent Interactive Technolo-gies and Multimedia, IITM, Archived in the ACM Digital Library, (2010).
[8] Mohanty, S. and Balabantaray, R. C., 2010, "Computational Model for Oriya Morphological Analyser", CENTRAL INSTITUTE OF INDIAN LANGUAGES, Department of Higher Education, Ministry of Human Resource Development, Government of India, Manasagangotri, Mysore - 570 006, Linguistic Data Consortium for Indian Languages (LDC-IL), Morphological Analysers and Generators, 22nd -23rd March 2010.
[9] Mohanty, S. and Santi, P. K., 2002, "Object Oriented design Approach to OriNet System: Online Lexical Database for Oriya Language", IEEE Proceedings for LEC-2002, University of Hyderabad, Hyderabad India. (2002)
[10] Mohanty, S., Santi, P. K. and Das Adhikari, K. P., 2005, "Analysis and Design of Oriya Morphological Analyzer (OMA): Some Tests with OriNet", Proceedings of symposium on Indian Morphology, Phonology and Language Engineering, IIT Kharagpur, India, (2005).
[11] Ojha, A. K., Behera, P., Singh, S. and Nath, G., 2015, "Trainning and Evaluation of POS Tagger in Indo-Aryan Languages: A Case of Hindi, Odia and Bhojpuri", Proceedings of 7th Language ans Technology Conference:Human Language Technologies as a Challenge for Computer Science and Linguistics, ISBN No-978-83-932640-8-7,2015.
[12] Richa, "Verb in Hindi: A Semantico-Syntactic Classification", Journal of Modern Language, Vol. 21, (2011)
[13] Sanjeev, K., Begum, R. and Mamidi, R., 2005, "A COMPARATIVE STUDY OF ENGLISH AND HINDI VERB SEMANTICS", LSI Platinum Jubilee Conference, HCU, 6th-8th December (2005)
[14] Shabadi, K. R. 2003, "Finite State Morphological processing of Oriya Verbal Forms", In Proceeding of EACL2003 workshop on computational Linguistics, Pages 49-56, Association for computational linguistics, (2003).
[15] Tiwary, U. S. and Siddique, T., 2010 Natural Language Processing & Information Retrieval, Oxford University Press, Chapter 9, Page 271, (2010).