Two-Word and Three -Word Disambiguation Rules for Telugu Language Sentences: A Practical Approach

J. Sreedhar[1]

[1] JNTUK, Kakinada

Received: 12 December 2013 Accepted: 2 January 2014 Published: 15 January 2014

Abstract

This paper describes Two-word and Three-word Disambiguation Rules for Telugu language sentences, which are written in WX-notation. Generally in real life good number of words, which are having many meanings. If a word has many meanings, then we can call it as a word ambiguity. To resolve a word ambiguity, Natural Language Processing (NLP) system having lot of Word Sense Disambiguation (WSD)[1] methods. Among many methods, here we are proposing rule based method for Word Sense Disambiguation.

Index terms— natural language processing, word sense disambiguation, rules, parts-of-speech.

1 Introduction

atural Language Processing(NLP) is a theoretically motivated multiple methods and techniques from which are selected for the accomplishment of particular type of language in analyzing and representing a human communicable at one or more level of linguistic analysis in the purpose of achieving human like languages processing for a range of tasks or applications.

Word Sense Disambiguation (WSD) [2] is the process of differentiating among the senses of words. The process of selecting most appropriate meaning of the word based on the context in which they occur. Computational identification of meaning for words in context is called Word Sense Disambiguation.

WSD [3] process to remove the ambiguity of word in a given context is an important for NLP applications such as Information Retrieval, Machine

2 Approach for Two Word Disambiguation Two Word Disambiguation Rules

Morphological analysis [10], [13] of a word gives detailed information about a word. Morphologically [11] every word carries information with reference to its lexemic form, morpho syntactic [12] category, and inflection. The detailed information may include among many other features, such as root/stem i.e. the lexemic shape listed in the dictionary the lexical category like noun/verb/adjective/adverb/pronoun /number /indeclinable as the case may be.

The following are some of the POS tag [4], **??**5] [6] disambiguation rules [7], [8], [9] used in the task: W1 :: W2 => W1 :: W2

Where W1 and W2 a sequence of words in that order. Where n is noun, v is verb, pn is pronoun, adj is adjective and adv is adverb.

Here from rule 2 when a word carries tags (n,pn) and followed by another word carrying the tag n then the tag pn retained eliminating the n from (n,pn). From rule 10 a word carrying the tag such as (n,pn) followed by avy then most the times pn will be retained and v will be eliminated. Depending on the context linguist will decide which tag will be retained and which one has to be eliminated. These are mostly contextually based syntactic rules. If two word sequences is unable to resolved unique tags then three words, four words sequence rules may be used for disambiguation.

3 III.

Theoritical Explanation with Example for Two Word Ambiguity Here in the above sentence the word carries tags (n,adj) and followed by another word carrying the tag n then the tag adj retained eliminating the n from (n,adj).so from the above sentence adj is eliminated and n is retained.

4 c) After Applying Disambiguation Rule

Adaxi a Nacivewaku alavAtu padipoyiMxi . n n n v punc Where punc is punctuation.

5 d) Analysis of Two Word Disambiguation

Here the below figures 1 and 2 explores the analysis of the Accuracy. Where X-axis indicates the number of test sessions and Y-axis indicates the Accuracy. As the result, we found that the proposal method can disambiguate nearly 98%. :: w2 :: w3 => w1 :: w2 :: w3 n,v,pn :: n :: pn,v => v :: n :: pn In the above sentence the first word carries tags (n,v,pn) and followed by second word carrying the tag n and followed by third word carrying the tags (pn,v) then the tag v retained from the first word and pn retained from the third word eliminating the (n,pn) from (n,v,pn) and eliminating v from (pn,v). iv. Analysis Of Three Word Disambiguation Here the above figures 3 and 4 explores the analysis of the Accuracy. Where X-axis indicates the number of test sessions and Y-axis indicates the Accuracy. As the result, we found that the proposal method can disambiguate nearly 96%.

We are very thankful to all the authors in a reference list, to make this research article in a better shape and right direction.

6 Conclusion and Future Research Direction

This research article explores the impact of twoword disambiguation and three-word disambiguation.

, [1]

Figure 1: N

n,adj :: n => n :: n -----------------(2)

n,pn :: n => pn :: n---------------(3)

n :: n,pn,v => n :: v ----------------(4)

n :: v,pn => n :: pn---------------(5)

avy :: v,pn => avy :: v----------------(6)

v ,pn :: avy => v :: avy----------------(7)

v,n :: n => n :: n------------------(8)

n :: n,v => n :: v -----------------(9)

v,pn :: avy => pn :: avy-------------(10)

n :: v,n,pn => n :: pn---------------(11)

n :: v,pn => n :: v ---------------(12)

n :: v,pn => n :: pn----------------(13)

n :: v,n => n :: n-----------------(14)

pn :: v,pn => pn :: pn---------------(15)

avy :: v,pn => avy :: v---------------(16)

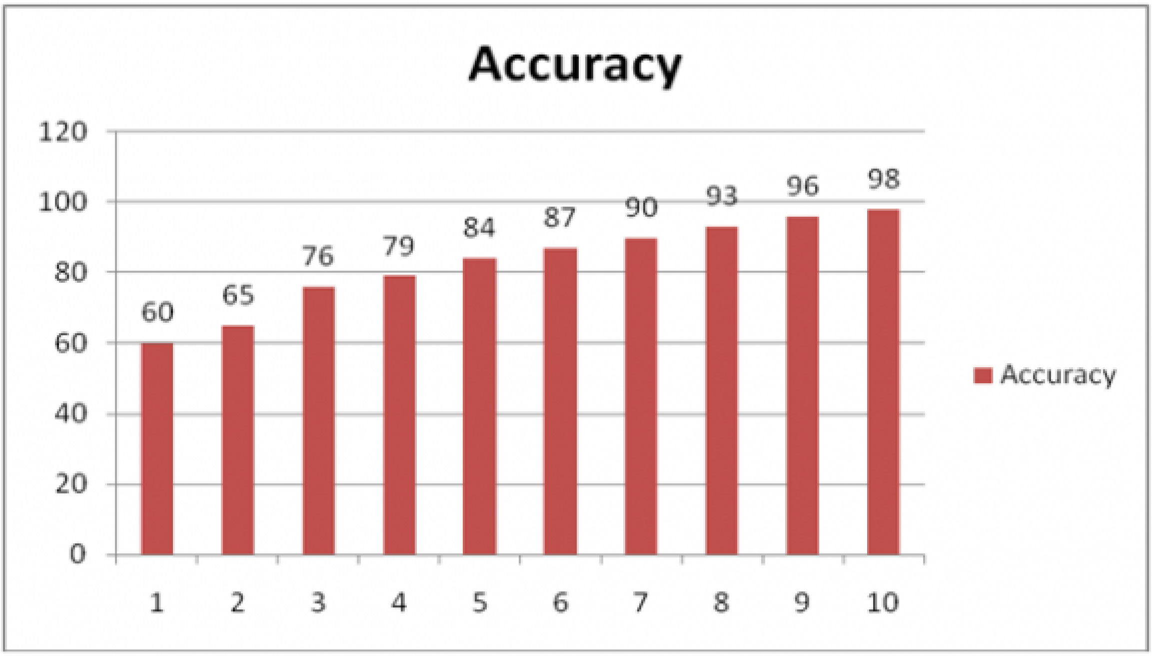
pn,v :: v => pn :: v---------------(17)

pn :: adj,n => pn :: n----------------(18)

n :: v,pn => n :: v -----------------(19)

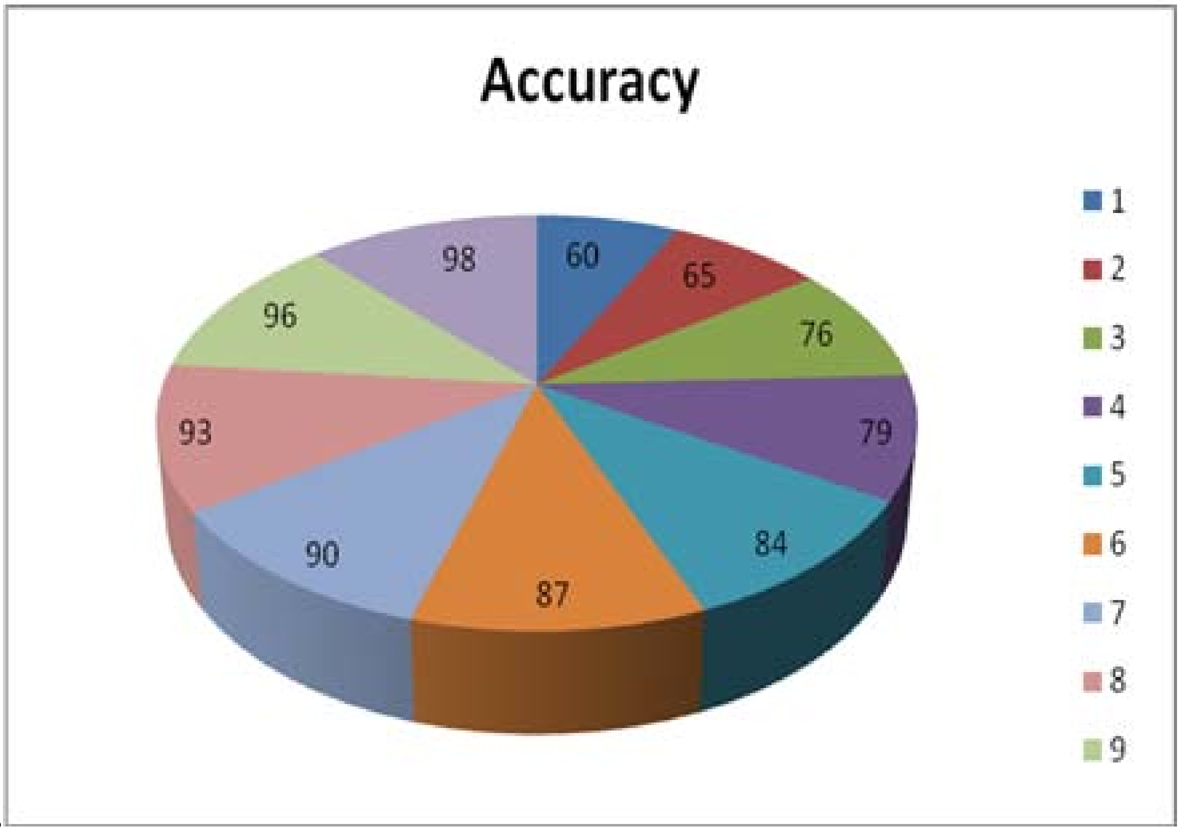
1 n,adj :: n => adj :: n ---------------(20)

Figure 2: Figure 1 :



2

Figure 3: Figure 2 :



3

Figure 4: Figure 3 :

W1 :: w2 :: w3 => w1 :: w2 :: w3--------------------(21)

n,v,pn :: n :: pn,v => v :: n :: pn ---------------------(22)

Pn :: n,adj :: pn,v => pn :: n :: v ---------------------(23)

n :: n,adv :: v => n :: n :: v ---------------------(24)

unk :: n,pn :: v,pn => unk :: n :: v---------------------(25)

n :: n,v :: v,pn => n :: n :: v ---------------------(26)

n :: v,pn :: n,adv => n :: v :: n ---------------------(27)

v,pn :: n : pn,v => v :: n :: v ---------------------(28)

n :: v,n :: v,pn => n :: n :: v ---------------------(29)

n,v : avy :: v,pn,adj => n :: avy :v ---------------------(30)

unk :: n,adj :: v,pn => unk :: n :: n---------------------(31)

pn :: v,pn :: v,pn => pn :: v :: v ---------------------(32)

v,pn,n :: v,pn,n,adj :: v,pn =>pn :: v :: v ---------------------(33)

avy :: n,adv :: v,pn => avy :: n :: pn---------------------(34)

n,adj :: n :: v,pn,n => n :: n :: v ---------------------(35)

n :: n,adv :: v,pn => n :: n :: v ---------------------(36)

n,adv :: adv :: v,pn => n :: adv :: p ---------------------(37)

v,pn,n :: v,pn ::avy => n :: pn :: avy ---------------------(38)

adv,n :: n,adj :: v,pn => adv :: adj :: v---------------------(39)

4 punc :: v,pn,n,adj :: v,pn => punc :: adj :: v--------------------(40)

Figure 5: Figure 4 :

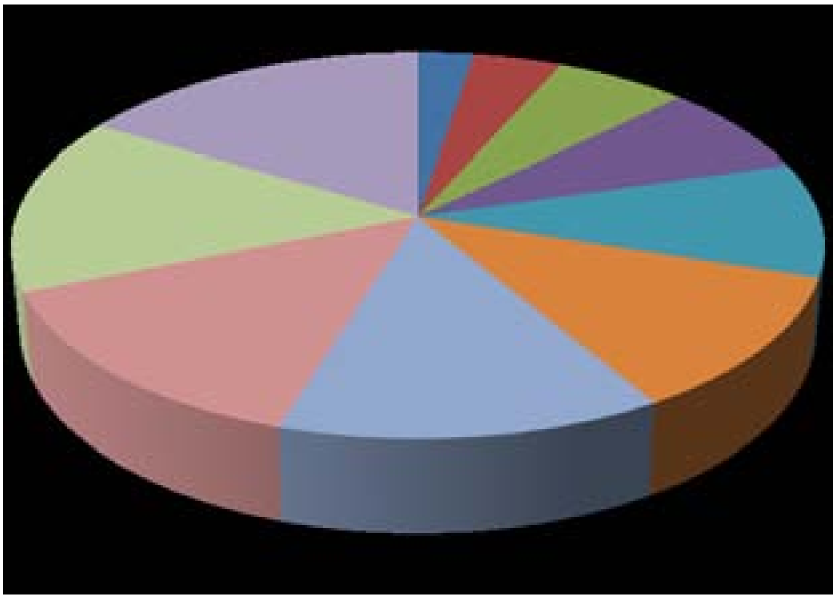


Figure 6:

2

8	926	n :: v,n :: v,pn	=> n :: n :: v	n :: n :: v
9	11634	n,v : avy :: v,pn,adj =>	n :: avy :v	n :: avy :v

Figure 7: Table 2 :

Here based on the context, linguist will decide which tag will be retained and which one has to be eliminated. We observed that if two-word and three-word sequences is unable to resolve unique tags, then four-word, five-word sequence rules may be useful for disambiguation.

[Minnen et al. ()] 'Applied morphological processing of English'. G Minnen , J Carroll , D Pearce . *Natural Language Engineering* 2001. Cambridge University Press. p. .

[Jurafsky and Martin ()] 'chapter 8: Word classes and Part of Speech Tagging'. D Jurafsky , J H Martin . 10.1109/ARTCom.184. *Speech and Language Processing*, 2000. 2009. IEEE Press. p. .

[Gelbukh et al.] *Detecting Inflection Patterns in Natural Language by Minimization of Morphological Model*, A Gelbukh , M Alexandrov , S Y Han . Progress.

[Dhanalakshmi et al.] V Dhanalakshmi , Anand Kumar , M Rekha , RU , Arun Kumar , C Soman , K P Rajendran , S . *Morphological Analyzer for Agglutinative Languages*,

[Pa and Thein ()] 'Disambiguation in Myanmar Word Segmentation'. W P Pa , N L Thein . *Proceedings Of the Seventh International Conference On Computer Applications*, (Of the Seventh International Conference On Computer ApplicationsYangon, Myanmar) 2009. p. .

[Disambiguation Rules for Telugu Language Sentences: A Practical Approach] *Disambiguation Rules for Telugu Language Sentences: A Practical Approach*,

[Zin and Thein ()] *Hidden Markov Model with Rule Based Approach for Part of Speech Tagging of Myanmar Language*, K K Zin , N L Thein . 2009. Yangon.

[Ma et al. ()] 'Hybride Neuro and Rule-Based Part of Speech Taggers'. Q Ma , M Murata , K Uchimoto , H Isahar . *International Conference on Computation Linguistics*, 2000. p. .

[Ide and Véronis ()] 'Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art'. Nancy; Jean Ide , Véronis . *Computational Linguistics* 1998. 24 (1) p. .

[Halevi (2006)] 'Part of Speech Tagging'. L Y Halevi . *Seminar in Natural Language Processing and Computational Linguistics*, (Israel) April, 2006. School of Computer Science, TeL Aviv University

[Jurafsky and Martin ()] 'SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistic and Speech Recognition'. D Jurafsky , J H Martin . *Pattern Recognition, Image Analysis and Applications: Lecture Notes in Computer Science*, 2000. 2004. 2004. Prentice-Hall. 3287 p. .

[Stevenson and Wilks ()] 'The Interaction of Knowledge Sources in Word Sense Disambiguation'. Mark; Yorick Stevenson , Wilks . *Computational Linguistics* 2001. 27 (3) p. .

[Wicentowski et al. ()] 'The swarthmore college senseval-3 system'. Richard Wicentowski , Emily Thomforde , Adrian Packel . *Proceedings of Senseval-3, Third International Workshop on Evaluating Word Sense Disambiguation Systems*, (Senseval-3, Third International Workshop on Evaluating Word Sense Disambiguation Systems) 2004.

[Wunderlich ()] 'Why is there Morphology'. Dieter Wunderlich . *23th Annual Meeting of the DGIS*, 2-4, 2004. 12.