Global Journal of Comuter Science and Technology: G Interdisciplinary
Volume 14 Issue 1 Version1.0 Year2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

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

By J. Sreedhar, Dr. S. Viswanadha Raju & Dr. A. Vinaya Babu
JNTUH, India

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.

Keywords: *natural language processing, word sense disambiguation, rules, parts-of-speech.*

GJCST-G Classification: *F.2.0*

Strictly as per the compliance and regulations of:

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

J. Sreedhar [α], Dr. S. Viswanadha Raju [σ] & Dr. A. Vinaya Babu [ρ]

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.

Keywords: natural language processing, word sense disambiguation, rules, parts-of-speech.

I. Introduction

Natural 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 Translation, Text Processing, Anti plagiarism, Speech Processing and Search Engines etc.

Organization of this research article is as follows: Here Section 2 describes Word Sense Disambiguation approach for Two-Word Disambiguation, Rules, Theoretical Explanation, Before Disambiguation, After Disambiguation and Empirical Approach for Two-Word Disambiguation. Section 3 explains Word Sense Disambiguation approach for Three-Word Disambiguation, Rules, Theoretical Explanation, Before Disambiguation, After Disambiguation and Empirical approach for Three-Word Disambiguation. Section 4 deals with Conclusion and Future Research Direction followed by the References.

II. WSD 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:

$$W1 :: W2 => W1 :: W2 \quad (1)$$

Where W1 and W2 a sequence of words in that order.

Author α: Research Scholar-CSE, JNTUK,Kakinada.
e-mail: sreedharyd@gmail.com
Author σ: Dr. S. ViswanathaRaju,Professor& HOD-CSE, JNTUHCEJ, Jagityala, Hyderabad.
Author ρ: Professor-CSE, JNTUH, Hyderabad.

Table 1 : WSD Rules with Sentence id's in the Telugu Carpus

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)

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

S.NO	SENTENCE ID	BEFORE DISAMBIGUATION RULE	AFTER DISAMBIGUATION RULE (RESULT)
1	14784	n,adj :: n => n :: n	n :: n
2	274	n,pn :: n => pn :: n	pn :: n
3	153	n :: n,pn,v => n :: v	n :: v
4	2291	n :: v,pn => n :: pn	n :: pn
5	10349	avy :: v,pn => avy :: v	avy :: v
6	21560	v ,pn :: avy => v :: avy	v :: avy
7	16646	v,n :: n => n :: n	n :: n
8	24355	n :: n,v => n :: v	n :: v
9	13677	v,pn :: avy =>pn :: avy	pn :: avy
10	442	n :: v,n,pn => n :: pn	n :: pn
11	531	n :: v,pn => n :: v	n :: v
12	4552	n :: v,pn => n :: pn	n :: pn
13	25974	n :: v,n => n :: n	n :: n
14	12455	pn :: v,pn => pn :: pn	pn :: pn
15	656	avy :: v,pn =>avy :: v	avy :: v
16	1893	pn,v :: v => pn :: v	pn :: v
17	590	pn :: adj,n =>pn :: n	pn :: n
18	560	n :: v,pn => n :: v	n :: v
19	18714	n,adj :: n => adj :: n	adj :: n

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.

III. Theoritical Explanation with Example for Two Word Ambiguity

Let us consider a telugu sentence which has ambiguous words from telugu corpus like
Sentence: Adaxi aNacivewaku alavAtu padipoyiMxi.

a) Morph Output

Adaxi	Ada /adj,n
aNacivewaku	aNacivewa/n
alavAtu	alavAtu /n
padipoyiMxi	padu/v,adv,pn,n

b) Before Applying Disambiguation Rule

W1 = Ada
W2 = aNacivewa
w1 : w2 => w1 :: w2
n,adj :: n => n :: n

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.

c) After Applying Disambiguation Rule

Adaxi	a Nacivewaku	alavAtu	padipoyiMxi .
n	n	v	punc

Where punc is punctuation.

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%.

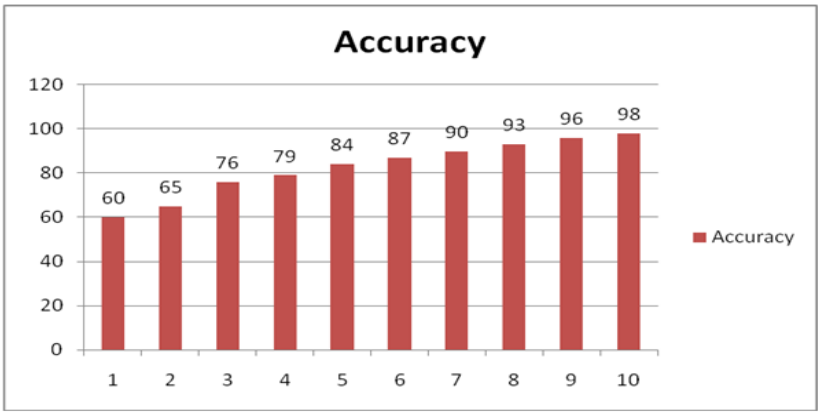

Figure 1 : Two word disambiguation rules accuracy

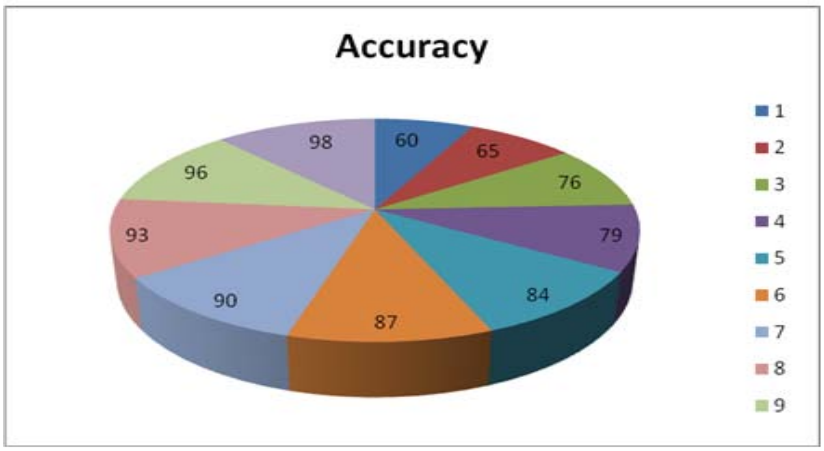

Figure 2 : Two word disambiguation rules accuracy

IV. WSD Approach for Three Word Disambiguation

a) Three Word Disambiguation Rules

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)

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

Table 2: Three word disambiguation rules

S.NO	SENTENCE ID	BEFORE DISAMBIGUATION RULE	AFTER DISAMBIGUATION RULE (RESULT)
1	876	n,v,pn :: n :: pn,v => v :: n :: pn	v :: n :: pn
2	25476	Pn :: n,adj :: pn,v => pn :: n :: v	pn :: n :: v
3	8357	n :: n,adv :: v => n :: n :: v	n :: n :: v
4	18476	unk :: n,pn :: v,pn =>unk :: n :: v	unk :: n :: v
5	5286	n :: n,v :: v,pn => n :: n :: v	n :: n :: v
6	20189	n :: v,pn :: n,adv => n :: v :: n	n :: v :: n
7	7514	v,pn :: n : pn,v => v :: n :: v	v :: n :: v
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

10	14007	unk :: n,adj :: v,pn => unk :: n :: n	unk :: n :: n
11	321	pn :: v,pn :: v,pn => pn :: v :: v	pn :: v :: v
12	3899	v,pn,n :: v,pn,n,adj :: v,pn =>pn :: v :: v	pn :: v :: v
13	16295	avy :: n,adv ::v,pn =>avy :: n :: pn	avy :: n :: pn
14	23539	n,adj :: n :: v,pn,n => n :: n :: v	n :: n :: v
15	2735	n :: n,adv :: v,pn => n :: n :: v	n :: n :: v
16	1094	n,adv :: adv :: v,pn =>n :: adv :: pn	n :: adv :: pn
17	28440	v,pn,n :: v,pn ::avy =>n :: pn :: avy	n :: pn :: avy
18	489	adv,n :: n,adj :: v,pn=>adv:: adj :: v	adv :: adj :: v
19	16963	punc :: v,pn,n,adj :: v,pn =>punc :: adj :: v	punc :: adj :: v
20	6804	n :: n,adj :: v,pn => n :: n :: v	n :: n :: v

b) Theoritical Explanation With Example For Three Word Ambiguity

Let us consider a telugu sentence which has ambiguous words from telugu corpus like

Sentence:

waMdri ceVppina viRayAlu AlociMcevAdu.

i. *Morph Output*

waMdri	waMdri/n
ceVppina	ceVppu/n,v,pn
viRayAlu	viRayaM/n
AlociMcevAdu	AlociMcu/pn,v

ii. *Before Applying Disambiguation Rule*

W1 = ceVppu
W2 = viRayaM
W3 = AlociMcu

w1 :: 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).

iii. *After Applying Disambiguation Rule*

waMdri	ceVppina	viRayAlu	AlociMcevAdu.
n	v	n	pn punc

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%.

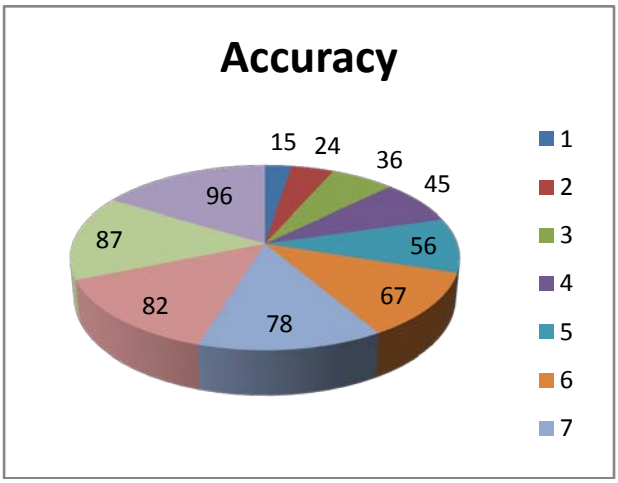

Figure 3 : Three word disambiguation rules accuracy

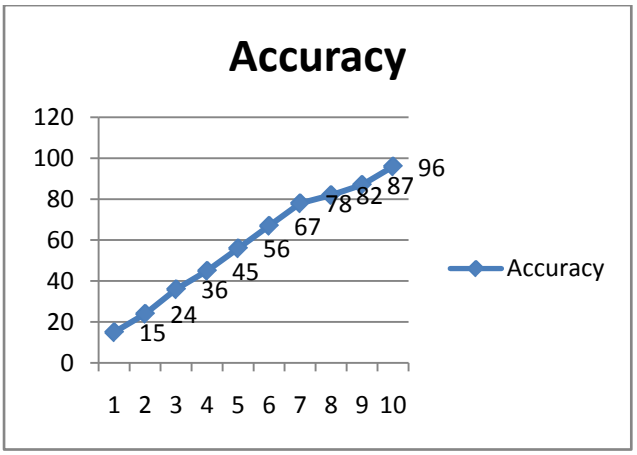

Figure 4 : Three word disambiguation rules accuracy

V. Acknowledgements

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

VI. Conclusion and Future Research Direction

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

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.

References Références Referencias

1. Richard Wicentowski, Emily Thomforde, and Adrian Packel. 2004. The swarthmore college senseval-3 system. In Proceedings of Senseval-3, Third International Workshop on Evaluating Word Sense Disambiguation Systems.
2. Ide,Nancy; Jean Véronis. 1998. Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art. Computational Linguistics 24(1), 1-40.
3. Stevenson, Mark; Yorick Wilks. 2001. The Interaction of Knowledge Sources in Word Sense Disambiguation. Computational Linguistics 27(3), 321-49.
4. L Y. Halevi, "Part of Speech Tagging ", Seminar in Natural Language Processing and Computational Linguistics, School of Computer Science, TeL Aviv University, Israel, April, 2006.
5. D.Jurafsky, J.H.Martin, "SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistic and Speech Recognition", Prentice-Hall, 2000.
6. D. Jurafsky and J. H. Martin "chapter 8: Word classes and Part of Speech Tagging", Speech and Language Processing, Prentice Hall, 2000.
7. K.K. Zin, N.L. Thein," Hidden Markov Model with Rule Based Approach for Part of Speech Tagging of Myanmar Language", Yangon, 2009.
8. Q. Ma, M. Murata, K. Uchimoto, H. Isahar, "Hybride Neuro and Rule-Based Part of Speech Taggers", International Conference on Computation Linguistics", 2000, pp. 509-515.
9. W.P. Pa, N.L. Thein "Disambiguation in Myanmar Word Segmentation", "Proceedings Of the Seventh International Conference On Computer Applications", Yangon, Myanmar, 2009, PP. 1-4.
10. Dhanalakshmi V., Anand Kumar M., Rekha R.U., Arun Kumar C., Soman K.P., Rajendran S., Morphological Analyzer for Agglutinative Languages Using Machine Learning Approaches, in International Conference on Advances in Recent Technologies in Communication and Computing (ATRCOM), IEEE Press,doi: 10.1109/ARTCom.184. Pp. 433-435 2009.
11. Dieter Wunderlich, Why is there Morphology, in 23th Annual Meeting of the DGIS, AG 12, 2-4, 2004.
12. A. Gelbukh, M. Alexandrov, and S. Y. Han, "Detecting Inflection Patterns in Natural Language by Minimization of Morphological Model", Progress in Pattern Recognition, Image Analysis and Applications: Lecture Notes in Computer Science, Volume 3287/2004, pp. 110-14, 2004.
13. G. Minnen, J. Carroll, and D. Pearce, "Applied morphological processing of English", Natural Language Engineering, pp. 207-223, Cambridge University Press, 2001.