International Conference on Machine Learning and Computational Intelligence-2017
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 7 | ISSN : 2456-3307

Sequence Labeling for Three Word Disambiguation in Telugu Language Sentences

Jinka Sreedhar*[1], Suresh Dara[1], Baijnath Kaushik[2], SK Althaf Hussain Basha[3]

[1]Department of Computer Science and Engineering B.V.Raju Institute of Technology, Narsapur, Telangana, India

[2]Department of Computer Science and Engineering Shri Mata Vaishno Devi University, Katra, Jammu and Kashmir, India

[3]Department of Computer Science and Engineering Gokaraju Rangaraju Institute of Engineering and Technoloy, Hyderabad, India

ABSTRACT

This paper is intended to apply sequence labelings which are introduced to find out the ambiguity in three-words. These words appear to give rise to ambiguity. They seem to be sequence words and this method can be applied only to these types of words. There is another theory of automata which is a mathematical model. By implementing this model to disambiguate the words of sequence it is found there is a kind of mathematical accuracy equal to that of sequence labeling method. The main aim of finding out these methods, is to find out solution to the problem of ambiguity in three-words sequences. Here designing of automata theory for three-words is dealt with the Three-Words disambiguation rules are explained with examples.

Keywords : Natural Language Processing(NLP),Information Retrieval Systems(IRS), Machine Translation(MT), Finite Automata(FA).

I. INTRODUCTION

To explain this theory clearly, five states have been identified. In this process state one may have more than one tag, state two may have more than one tag and state three may have more than one tag. Now one tag has been retained in the word one deleting the remaining tags. In the similar manner, the same procedure is continued in the second word order and third word also[1,2,3,4]. By doing so, the problem of ambiguity has been resolved. When this process comes to state four, it is treated as completed since it gives a complete sense. In the fifth state regarded as a dead state, all the unwanted tags will be appear[5,6,7,8].

With the help of transitional diagrams and transitional tables, the rules are explained. Transitional diagrams contain states, POS tags, start state and final state[5]. These diagrams can be represented with the symbols like Q, ∑, S, F. Here Q stands for states one, two, three and dead states, ∑ contains POS tags, S contains the starting state, that is, state one, and F contains the final state, that is, state four[13,14,15,16,17,18].

Transitional Table is also framed to show how these tags appear in different states and give a picture representation.

W1 :: W2 :: w3 => W1 :: W2 :: W3

Where

W1 ,W2 and W3 are sequence of words in that order.

II. DESIGNING AUTOMATA THEORY FOR THREE WORDS RULES

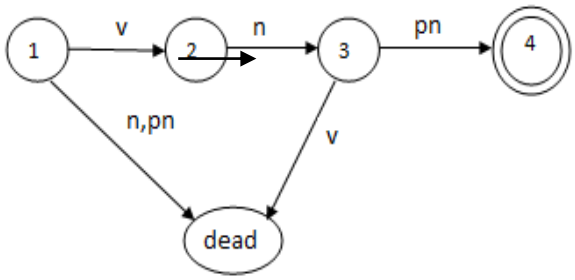

Figure 2.1. Three-word disambiguation for v, n, pn :: n :: pn,v

Where

1, 2, 3 and dead state belong to Q and n, v, pn belongs to ∑.

Here v denotes verb, pn denotes pronoun and n denotes noun.

Q: {1,2,3,4,dead}

∑:{v,n,pn}

S:{1}

F:{4}

Table 2.1. Three-word disambiguation for v, n, pn :: n :: pn,v

∂	v	n	Pn
1	2	dead	Dead
2	-	3	-
3	dead	-	4
4	-	-	-
dead	-	-	-

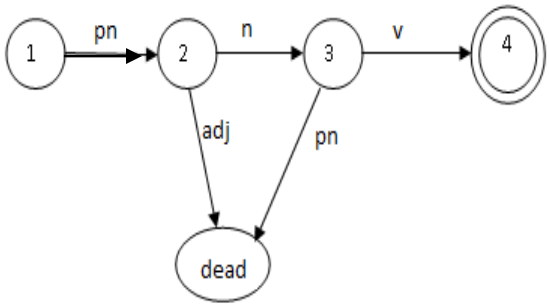

Figure 2.2. Three-word disambiguation for pn :: n, adj :: v, pn

Where

1, 2, 3 and dead state belong to Q and n, v, pn, adj belongs to ∑.

Here v denotes verb, pn denotes pronoun,n denotes noun and adj denotes adjective.

Q:{1,2,3,4,dead}

∑: {pn,n,v,adj}

S: {1}

F: {4}

Table 2.2. Three-word disambiguation for pn :: n, adj :: v, pn

∂	pn	n	V	adj
1	2	-	-	-
2	-	3	-	dead
3	dead	-	4	-
4	-	-	-	-
dead	-	-	-	-

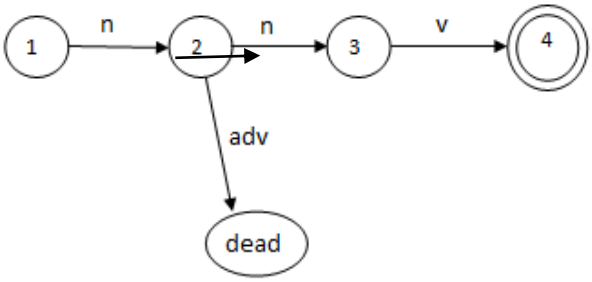

Figure 2.3. Three-word disambiguation for n :: n, adv :: v

Where 1, 2, 3 and dead state belong to Q and n, v, adv belongs to ∑.

Here v denotes verb, n denotes noun and adv denotes adverb.

Q: {1,2,3,4,dead}

∑: {n,v,adv}

S: {1}

F: {4}

Table 2.3. Three-word disambiguation for n :: n, adv :: v

∂	n	v	adv
1	2	-	-
2	3	-	dead
3	-	4	-
4	-	-	-
dead	-	-	-

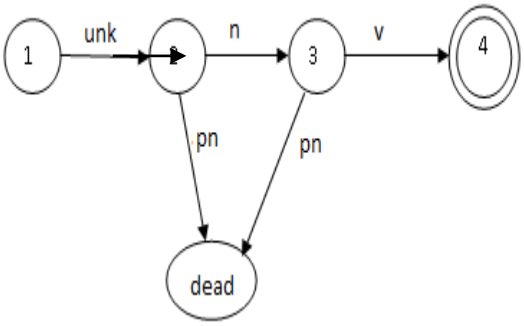

Figure 2.4. Three-word disambiguation for unk :: n, pn :: v, pn

Where

1, 2, 3 and dead state belong to Q and unk, v, pn, nbelongs to ∑.

Here v denotes verb, pn denotes pronoun, n denotes noun and unk denotes unknown.

Q: {1,2,3,4,dead}

∑:{unk,n,v,pn}

S:{1}

F: {4}

Table 2.4. Three-word disambiguation for unk :: n, pn :: v, pn

∂	unk	n	v	pn
1	2	-	-	-
2	-	3	-	dead
3	-	-	4	dead
4	-	-	-	-
dead	-	-	-	-

III. THREE WORD DISAMBIGUATION RULES

W1::W2::W3 ⇒ w1::w2::w3…..(1)

N,v,pn :: n::pn,v ⇒ v::n::pn…..…...(2)

Pn::n,adj::pn,v ⇒ pn::n::v…..(3)

N::n,adv::v ⇒ n:n:v …. ….(4)

Unk::n,pn::v,pn ⇒ unk::n::v….(5)

n::n,v::v,pn ⇒ n::n::v ….(6)

n::v,pn::n,adv ⇒ n::v::n..…(7)

v,pn::n:pn,v ⇒ v::n::v.(8)

n::v,n::v,pn ⇒ n::n::v……..(9)

n,v:avy::v,pn,adj ⇒ n::avy:v…….....(10)

unk::n,adj::v,pn ⇒ unk::n::n.....(11)

pn::v,np::v,pn ⇒ pn::v::v…...(12)

v,p,n,n::v,pn,n,adj::v,pn ⇒ pn::v::v…...(13)

avy::n,adv::v,pn ⇒ avy::n::pn...(14)

n,adj::n::v,pn,n ⇒ n::n::v.....…(15)

n::n,adv::v,pn ⇒ n::n::v..……..(16)

n,adv::adv::v,pn ⇒ n::adv::p……(17)

v,pn,n::v,pn::avy ⇒ n::pn::avy..…...…(18)

adv,n::n,adj::v,pn ⇒ adv::adj::v…….....(19)

punc::v,pn,n,adj::v,p ⇒ punc::adj::v…...…...(20)

3.1 Case Study for three word ambiguity

Here is a Telugu sentence which has ambiguous words from Telugu corpus, like

Sentence:

waMdri ceVppina viRayAlu AlociMcevAdu.

Morph Output:

waMdri waMdri/n

ceVppina ceVppu/n,v,pn

viRayAlu viRayaM/n

AlociMcevAdu AlociMcu/pn,v

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) followed by the second word carrying the tag n and followed by a third word carrying the tags (pn,v). Then the tag v is retained from the first word and pn is retained from the third word eliminating the (n,pn) from (n,v,pn), and v from (pn, v).

After Applying Disambiguation Rule:

waMdri ceVppina viRayAlu AlociMcevAdu.

n v n pn punc

3.2 Analysis of Three Word Disambiguation

Here the following figure 3.1 gives an analysis of the Accuracy. While X-axis indicates the number of test sessions, Y-axis indicates the Accuracy. As a result, the proposed method can disambiguate nearly 96% of the ambiguity [9,10,11,12].

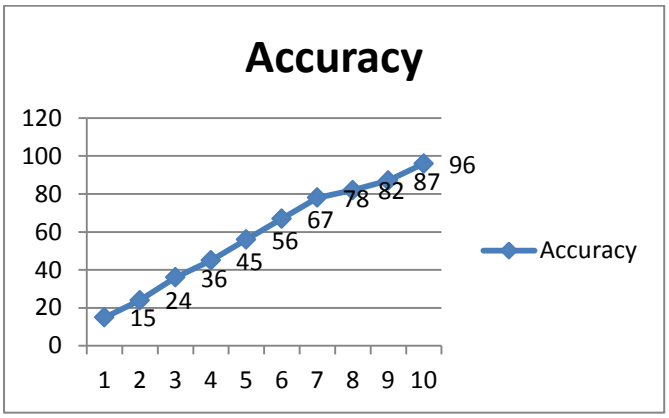

Figure 3.1. Three word disambiguation rules accuracy

IV. CONCLUSION

Here dealing with the designing of three-word rules for Telugu Language Sentence word order. To make things easy to understand some rules have been made which can be applied for the word order of Telugu Language Sentences and it clarifies the ambiguity. All these things are so vividly explained with the help of case studies and theoretical explanations. When these rules are applied whenever needed, they help the user easily to eliminate the ambiguity. These theories help understand the study of disambiguation. By applying disambiguation rules it is found that the proposed method can disambiguate nearly 96% of the ambiguity. The theoretical explanation and disambiguation rules have resulted in the accuracy of evidences.

V. REFERENCES

[1]. Noam Chomsky, "Logical syntax and semantics: their linguistic relevance", vol.31, No.1, pp: 36-45, 1955.

[2]. Noam Chomsky, "On Certain Formal Properties of Grammars", Information and Control, Vol. 9, pp: 137-167, 1959.

[3]. Nou, Chenda and WataruKameyama,Khmer, "POS Tagger: A Transformation-based Approach with Hybrid Unknown Word Handling", Proceedings of the First IEEE International Conference on Semantic Computing (ISCS), Irvine, CA. pp: 482-492, 2007.

[4]. PawanGoyal,LaxmidharBehera,Thomas Martin McGinnity, "Query Representation through Lexical Association for Information Retrieval", IEEE Transactions on Knowledge and Data Engineering, pp: 2260-2273, 2012.

[5]. PengJin,XingyuanChen,"A Word Sense Probabilistic Topic Model", 9th International Conference on Computational Intelligence and Security (CIS), pp: 401-404, 2013.

[6]. PengYuan Liu, "Another View of the Features in Supervised Chinese Word Sense Disambiguation", 9thInternational Conference on Computational Intelligence and Security, ISBN: 978-0-7695-4584-4, pp: 1290-1293, 2011.

[7]. Pengyuan Liu, YongzengXue, Shiqi Li, Shui Liu, "Minimum Normalized Google Distance for Unsupervised Multilingual Chinese-English Word Sense Disambiguation", International Conference on Genetic and Evolutionary Computing, ISBN: 978-0-7695-4281-2, 2010.

[8]. Ping Chen,BowesC,Wei Ding, et..al, "Word Sense Disambiguation with Automatically Acquired Knowledge",IEEE INTELLIGENT SYSTEMS, 2012.

[9]. PrashanthMannem, "Bidirectional Dependency Parser for Hindi, Telugu and Bangla", Proceedings of ICON09 NLP Tools Contest: Indian Language Dependency Parsing, India, 2009.

[10]. Quinlan, J. R, "Induction of decision trees", Mach. Learn. 1, 1, pp: 81–106, 1986.

[11]. Quinlan, J. R, "Programs for Machine Learning", Morgan Kaufmann, San Francisco, CA, 1993.
[12]. R.M.K.Sinha and K. Sivaraman, "Ambiguity Resolution in Anglabharati",TRCS-93-174, Department of Computer Science and Engineering, IIT,Kanpur, India, 1993
[13]. R. Mahesh K. Sinha,"Learning Disambiguation of Hindi Morpheme 'vaalaa' with a Sparse Corpus" 4th International Conference on Machine Learning and Applications, ISBN: 978-0-7695-3926-3, pp: 653-657, 2009.
[14]. J.Sreedhar, S. Viswanadha Raju, A. Vinaya Babu, Amzan Shaik, P.Pavan Kumar "Word Sense Disambiguation : An Empirical Survey" International Journal of Soft Computing and Engineering(IJSCE), Volume-2,Issue-2,May-2012,ISSN:2231-2307.
[15]. J.Sreedhar, S. Viswanadha Raju, A. Vinaya Babu, Amzan Shaik, P.Pavan Kumar"A critical Approaches to Identification of Disambiguation Words in NLP : A Current State of the Art" International Journal of Engineering Trends and Technologies (IJETT), Volume-3,Issue-3,May-2012,ISSN:2231-5381.
[16]. P.Pavan Kumar, J.Sreedhar "Innovative Techniques and Technologies in Translation in a Multilingual Context" 3rd International Conference on Translation Technology and Globalization in Multilingual Context" Delhi, June 23-26, 2012.
[17]. P.Pavan Kumar, J.Sreedhar "Language teaching and MLE in the context of the third revolution" DLA Conference on Dravidian Languages and Translation Technology" HCU, Hyderabad, June 18-20, 2012.
[18]. Hyuk-Chul Kwon, Minho Kim, Youngim Jung, "Hybrid word sense disambiguation using language resources for transliteration of Arabic numerals in Korean", International Conference on Hybrid Information Technology (ICHIT '09), ISBN: 978-1-60558-662-5, pp: 314-321,2009.