



International Journal of Sanskrit Research

अनन्ता

ISSN: 2394-7519
IJSR 2016; 2(6): 78-81
© 2016 IJSR
www.anantaajournal.com
Received: 09-09-2016
Accepted: 10-10-2016

Chandana Bathulapalli
K. J. Somaiya College of
Engineering Vidyavihar,
Mumbai, Maharashtra India

Drumil Desai
K. J. Somaiya College of
Engineering Vidyavihar,
Mumbai, Maharashtra India

Manasi Kanhere
K. J. Somaiya College of
Engineering Vidyavihar,
Mumbai, Maharashtra India

Correspondence
Chandana Bathulapalli
K. J. Somaiya College of
Engineering Vidyavihar,
Mumbai, Maharashtra India

Use of Sanskrit for natural language processing

Chandana Bathulapalli, Drumil Desai and Manasi Kanhere

Abstract

Natural language processing with natural language inputs and outputs provides a better human-machine interface. But the natural language ambiguity poses a major hurdle in the processing. This paper proposes Sanskrit language as a possible natural language input to computers. The relatively unambiguous nature of this language and its well laid-out grammatical structure promote this language as the language for processing.

Keywords: Natural language, semantic net, ambiguity, Sanskrit, cases

1. Introduction

Artificial intelligence penetrates into a myriad number of fields, one of them being natural language processing. With man's continuous efforts to stay away from the jargon of machines, arose the need to feed natural languages as inputs to machines. All natural languages express a large amount of ambiguity. Though the languages are correctly interpreted by humans out of usage, for a computer that lacks the capability to distinguish between the various interpretations on a contextual basis, this ambiguity proves to be a vice. Hence, the present day is in need of a language that could to a great extent eliminate this ambiguity and at the same time be suitable for knowledge representation in artificially intelligent systems. Sanskrit which has been in prevalence since thousands of years and has not worn out and moulded in the course of human usage can be used for such knowledge representation.

2. Natural Language Processing

Natural language processing is the sub field of artificial intelligence devoted to make computers understand statements or word written or spoken in human language. The field of Natural Language Processing (NLP) involves making computers to perform useful tasks using languages used by humans. The two major parts of natural language processing include natural language understanding at the input side and the natural language generation at the output side. In natural language processing input and output could be text or speech.

Natural language understanding involves

- Mapping the given input in natural language into useful representation.
- Analysing different aspects of the language.

2.1 Natural Language Generation Involves

- Text planning-It includes retrieving relevant content from knowledge base.
- Sentence planning- It includes required word and forming meaningful phrases.

2.2 Application of Natural Language Processing

Natural language processing provides a better human-computer interface that could aid artificial intelligence systems to pervade more efficiently into the present day applications like:

- Natural language processing system for blind people to interact with computers with speech input.
- The chair of Stephan Hawking which converts text into speech.
- The translation program that could translate from one human language to another.
- A program which checks for grammatical errors in a given text.

3. Ambiguities in Natural Language Processing

Ambiguity refers to the property of words or sentences having more than one meaning or being interpreted in more than one way. Most of the natural languages are ambiguous, rendering computers unable to extract the appropriate interpretation of the input language in some situations. The ambiguities that can arise while processing natural languages are as follow

3.1 Scope Ambiguity

Example: I was eating ripped mangoes and banana.
Here the word ripe may be associated with only mangoes or both mangoes and banana.

3.2 Attachment Ambiguity

Example: I touched the girl with the rose.
Here the person may touch a girl who is holding a rose or the person touches the girl using a rose.

3.3 Semantic Ambiguity

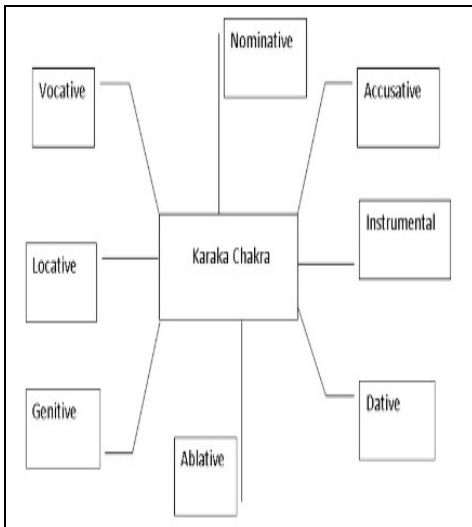
Example: The car hit the pole while it was moving.
Here the sentence can be interpreted as the car in motion hit a stationary pole or the car hit a moving pole. This ambiguity arises because of the computers inability to distinguish between what is logical and what is not.

3.4 Pragmatic Ambiguity

Example: I love you too.
Here the sentence can be interpreted in many ways like I love you (just like you love me) I love you (just like someone else does) I love you (and I love someone else) I love you (as well as I like you).

4. Sanskrit for Natural Language Processing

Sanskrit is one of the very few languages which has formal defined grammar. The learning of this language starts with understanding of the fundamental rules and norms which are to be abided by unlike other natural languages which are learnt in due course by continual communication. The grammatical "Treatise" 'Ashtadhyayi' with 3959 sutras (formulae or rules) composed by the ancient grammarian Maharshri Panini is referred to by many scholars as the crude form of coded efficient language. This language has a rich and often rigid declension of noun indicating their relationship to each other in sentences and this makes the language more conducive to be approximated to a semantic net model used in artificial intelligence system.



4.1 Vibhakti in Sanskrit

Sanskrit language has 8 predefined vibhakti's (cases). Words in a given sentence decline in one of these eight cases together convey the correct, precise, specific and unambiguous meaning of the sentence to the reader. The usage of eight cases of vibhakti's is as follows:-

- 1. Nominative:** To address nouns (proper/common) Example-Ram eats a fruit.
- 2. Accusative:** The accusative part of the sentence is generally the object on which the action is being performed. Example-Ram gives the book.
- 3. Instrumental:** It denotes the instrument with which the action is being performed. Example-Ram writes with a pen.
- 4. Dative:** This case denotes the object for which the object is being performed. Example- Ram works for his children.
- 5. Ablative:** It expresses the point of separation of the object. Example- Ram fell from the chariot.
- 6. Genitive:** The word declined in the genitive case is the possessor of the object. Example- Ram's house is made of bricks.
- 7. Locative-** It signifies the location of the object. Example- Ram is sitting on the floor.
- 8. Vocative:** This case is used to address a person or an object. Example- Oh Ram! Please help us.

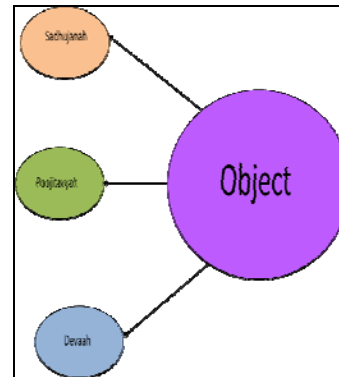
5. Sanskrit's Association with Programming Language

Sanskrit, with all its vibhakti's and grammatical rules can be more or less approximated to a programming language with classes and objects. All the words that are used to make a sentence in Sanskrit are fundamentally properties when appended with a proper case these words can be treated as object. Consider an example,
"Sadhujanahpoojitavyahdevahkhaluprithivyaam"
Meaning: The good one worshipped who is indeed a God on earth.

Here, 'saadhujana' is the property of being good, 'poojitavya' is the property of worshipping, and 'deva' is the divine property. When declined in the nominative case, all the three words act as pointers to one common object i.e. words declined in the same case behave as pointers to a single object. In this case, the object may be a person whose attributes (properties) are being good, worthy of worship and of godly stature.

6. Consider another example

"Chhatrah pustakam pathati"- Here, Chhatraah is a Chhatral instance in the nominative case and can also be declined in seven other cases which will be pointers pointing to a certain object.



7. Semantic Nets

Semantic nets are techniques for knowledge representation used for propositional information. They are two dimensional representations of knowledge. Mathematically, they are directed labelled graphs consisting of

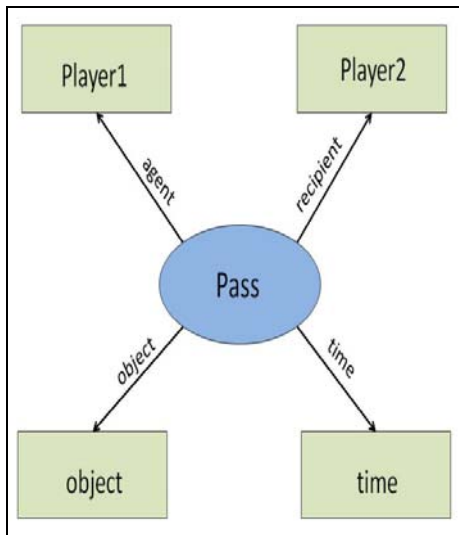
- Nodes
- Links(edges)
- Link labels

Nodes appear as circles, ellipses or rectangles. They represent physical objects, concepts or situations. The links express the appropriate relationship between the objects.

Consider the example: "Player1 passed the ball to player2"
Here, the information can be represented in a set of triples as follows:

- Pass, agent, Player1
- Pass, agent, Player2
- Pass, agent, ball
- Pass, time, past

The action of passing forms the central node out of the five nodes. The action of passing was being performed on the ball (object). The ball was being passed from Player1 (agent) towards Player2 (recipient) sometime in the past (time).



8. Equivalence of Sanskrit with Semantic Nets

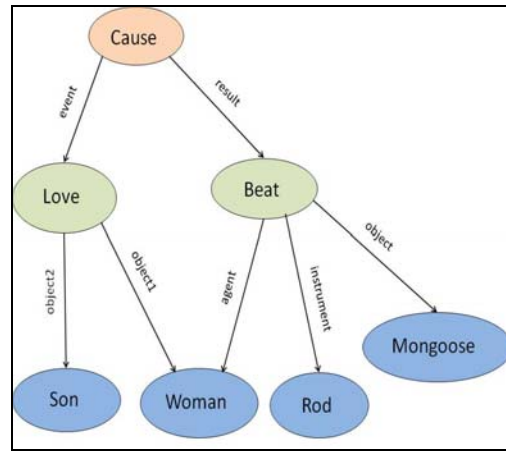
The Sanskrit language with all its Vibhaktis (cases) provides an efficient way of segmenting the sentences into logical constructs for natural language processing. The breakdown of the sentences in Sanskrit is very similar to the semantic net models used for artificial intelligence systems.

The following example is from Panchatantra written by Vishnu Sharma in the 3rd century B.C

“Swaputrasnehaat sa brahmami nakulam danden taadayat”
It means: “Out of love for her own son, the woman hit the mongoose with a rod”

8.1 This information can be broken down into triples as follows

- Beat, agent, woman
- Beat, instrument, rod
- Beat, object, mongoose
- Love, woman, son
- Beat, because of, love



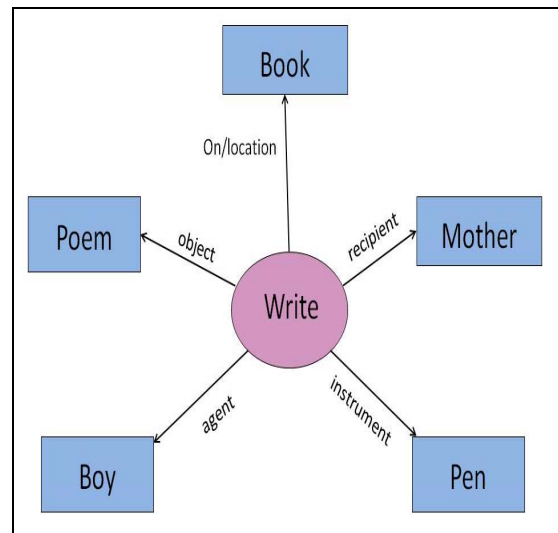
Another example which includes the locative case is:

“Balakah maatre pustake kalamena kavitam likhati”

Meaning “Boy writes a poem for his mother in a book with a pen”

This information can be broken down into triples as follows:

- Write, agent, Boy
- Write, recipient, mother
- Write, object, poem
- Write, instrument, pen
- Write, location, book



9. Conclusion

The specific and unambiguous nature and at the same time, the vast literature and vocabulary of the Sanskrit language provides a gateway for implementing this language in a way that a computer can understand. NASA scientists believe that Sanskrit language can provide a huge impetus to the development to the field of artificial intelligence.

10. References

1. Anjali MK, Babu Anto P. Ambiguities in Natural Language Processing, 2014. [online],rroj.com <http://www.rroj.com/open-access/ambiguities-in-natural-language-processing.pdf>
2. Abhimanyu Chopra, Abhinav Prashar. Natural Language Processing, 2013. [online],ijtee.org <http://www.ijtee.org/final-print/nov2013/Natural-Language-Processing.pdf>

3. AI – Natural Language Processing, tutorialspoint, 2015. [online]
http://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processing.htm
4. Robin. Natural Language Processing-Overview, 2009. [online] language.worldofcomputing.net,
<http://language.worldofcomputing.net/nlp-overview/natural-language-processing-overview.html>
5. Shah GS. Similarities between Sanskrit and Programming Languages, Uttishthabharatam. Wordpress, com, 2011. [online]
<https://uttishthabharata.wordpress.com/2011/05/30/sanskrit-programming/>
6. Rick Briggs. Knowledge Representation in Sanskrit and Artificial Intelligence, 1985. [online] aaii.org
<https://www.aaii.org/ojs/index.php/aimagazine/article/download/466/402>
7. Kale MR. Panchatantra in Vishnuserman Motilal Banarsidass Publishers Pvt Ltd. Delhi, India. Reprint Ed, 2005.