Pre-Conference Papers

National Conference
On
"Samskritam and Computers"

27-28 April 1989

SANKARA
c/o Sankara Vidya Kendra,
Vasant Vihar
New Delhi - 110057

Contents

Part I

COMPUTER REPRESENTATION OF SABDA - BODHA

*Shri P. Ramanujan**

A. ABSTRACT

(It is a translation with some amplification where felt necessary of the companion article in Sanskrit "masthishka yanthre shabda-bodha prathiroopanam" by the Author submitted to the `First National Conference on Samskritam and Computers' held at New Delhi during 27-28 April 1989)

It is the considered opinion of Computer Scientists that Artificial Intelligence generation would greatly enhance the utility of Computers in all modern scientific investigations. Also, it is accepted that generation of Artificial Intelligence could be either in Inference or Verbal Import/Semantic Extraction form. The latter is described in this article in some detail. If the machine interactions are possible in Sanskrit, then generation of Verbal Import would be easier. Though there are different approaches in the definition and format of Verbal Import among Logicians (**Naiyayikas**), Grammarians (**Vaiyakaranas**) and Vedic Scholars (**Mimamsakas**), the computer representation aspect of Verbal Import, in general, only is discussed here. As the grammar of Sanskrit is well-structured by Panini's Suthras, it could be useful to write programs, formulate commands and instructions and perform other interactions in Sanskrit itself for representing Verbal Import in particular and knowledge in general.

Firstly, in order to extract the Semantics of a sentence; split it for getting the constituent words; and get individual meanings, connect these suitably to arrive at the sentential import are discussed. Secondly, the human processes involved in intuition development and its representation are considered. The words denoting their own meanings together with the syntactic connections required for getting the sentential import are also dealt with. Lastly, the outputting of the sentential import in a particular format is described.

Utilising such a facility of representing sentential import, various application areas for in-depth study in all the 14 branches of traditional Learning are enlisted. Thus, an approach for representation of Verbal Import and its various applications are covered in this paper.

* 188/2, 2nd Phase, 4th Main,
Manjunatha Nagar,
Bangalore-560010.

B. Introduction

When a learned person addresses another through a sentence, the knowledge generated in the listener, communicated in the form of another explanatory sentence conveying the message in an unambiguous and explicit form is defined as Sabda-Bodha (Verbal Import or Semantics). Here, a sentence is a collection of words and is potent. Words could be classified into different types like noun-forms (subantha - ending with 'sup' affixes), verb- forms (thingantha - ending with 'thing' affixes) etc. The connection (correlation) between a word and its meaning (object) is natural, i.e., eternal. This is why, by the mere utterance of a word, the knowledge about the object denoted by it is generated. {By saying word and meaning are naturally (and so, eternally) connected, what is meant is that according to the axioms "ya eva vaidikah sabdah tha eva laukika api" and "loke prayukthah sarvopi vede kvapi prayogavan", all words used in colloquial language are a sub-set of vedic words and, since vedas are eternal, all their words denote their respective objects always. Thus, in Sanskrit, word meanings are not conventional. This very concept is a subject of lengthy discussion elsewhere, but is presently out of context}.

Now, if we examine the process of verbal cognition (vyutpatti) in humans, the first development of verbal cognition, i.e., knowing word-meaning relations, is caused by various sources like the parents and relatives pointing to certain objects like Moon, Father, Mother etc. simultaneously using the words denoting them, e.g., "see, that is the Moon", "This is Mother" etc. whereby the child, on repeated usage gets a relationship between them (as the denoter and denoted), where the word is the denoter and the meaning, the denoted (object), well registered in mind. Similarly, following conversations between elders, carried out using gestures and actions as well, the observant child gets the "implied" meaning. Feelings like happiness, sorrow, anger etc. expressed facially, are also comprehended likewise. The *"intuition"* thus developed, is further enhanced by formal modes of 'Learning' through Dictionary, Grammar, Simile, Verbal Testimony, Conversations, Inference, Explanations/Commentaries, Proximity etc. Factors like concentrated study, practice, chance, discussion, experience etc. greatly reform/improve intuitive 'understanding'.

In such a situation, realising the potency of words in Computers must somehow imitate this process of human acquisition of knowledge. In Shasthras, three factors are said to cause the 'comprehension' of the import on the hearing of a sentence, in a coherent and unambiguous way, viz. Expectancy, Proximity and Compatibility. Of these, expectancy means the process whereby each word 'fills-up' the sentential import by "requiring" proper connectives (setting up questions the answers for which cohere the sentential import). Proximity is just the presence in the same sentence (collocation). Compatibility is consistency of meaning (in relation to other 'instruments of knowledge').

The potency of words (ability to denote a particular object) is also termed as vrithi (sense). It is of two types, denoted and implied. The meaning occurring by the intuitive cognition is the denoted, e.g., the word 'simha' when used to mean a lion. Denotation of meaning is of many types due to category, quality, topic, deduction, connotation, a combination of the last two etc. When the denoted meaning is inadmissible, implication is taken in an object 'closer' to the denoted one. This 'closeness' may be connection with the denoted object or its attributes. The former sense is termed primary implication (lakshana). The same could be further extended to other levels. The latter sense is secondary implication, e.g., by saying 'Devadatta is a Lion' is meant the valour and such other qualities present in Devadatta. Thus, to determine the import of a sentence, the denoted and implied meanings are to be established for the words and the applicable one in the context has to be chosen. These are to be borne in mind while representing verbal import.

There are two schools of thought describing the process of Semantic Extraction. The first holds that words describe the syntactic connection required for cohering the sentential import besides their own meanings, i.e., separate words denote (remind) their objects since denotion is unique. Further, the words occurring in a single sentence put out a coherent meaning, and hence, by Logic, Congruence etc. connect the unrelated word meanings to form a sentential import. This view is called 'anvithabhidhanam'. The second school contends that words denote their respective meanings, word meanings collectively denote sentential import where the word meanings also meet separately. Thus, the issues pertaining to Semantic Extraction are enumerated.

C. Sentential Analysis

Unlike the cognition of sentient beings which is due to Consciousness, machine representation of Cognition needs careful study. Hence, word-by-word analysis of sentence is desirable. Annexure-I details the classification of words to choose the best approach for sentential analysis and representation. As an illustration, as per the definition of sentence "thing subantha cayo vakyam kriya va karakanvitha" (it is a collection of words ending with *thing* and *sup* affixes or action connected by karakas) words are first classified as *subantha* (noun-forms), *thingantha* (verb-forms) and then into indeclinables, compound and combined words etc. with further sub-divisions as shown. These being codified by *Panini*, are *to be formalised and represented*, both tough but necessary tasks indeed. In accordance with the adage "gacchath pipilika pankthih...." a *mobile* battalion of Ants would eventually reach even the Ocean, while even *Garuda* (reputed to be super fast), if *static*, would reach nowhere), a humble start is being made in this direction. Disregarding the question of authenticity etc. of a given sentence, understanding the denoted import and extracting the Semantics alone is the topic of the present discussion. Hence, only coherent sentences are considered initially. Simple words of common usage with unique meanings and used in conventional sense are analysed first, the process gradually including other more

difficult forms like combined words and compound words, in that order. This graded approach would help the problem-understanding better, as a part of analysis of the previous stage is utilised in the subsequent stages. So, to begin with, if *data* regarding noun- forms, verb-forms and indeclinables could be stored in the memory, the processing could start. As indeclinables are quite limited in number and variety, their classification and storage is straight-forward. Regarding noun-forms and verb-forms, two approaches could be considered. Either all the possible *types* are established and the modifications the concerned affixes undergo while applied to any of these types are stored beforehand *or* the problem of formalisation if the grammar rules could be launched straightaway. In the first case, noun forms are categorised ending and gender-wise and the modified sup affixes for the various cases and number with typical bases for each category are to be stored. For verb-forms, the roots (denoting action), the various affixes, group, usage (padhi), mood etc. are to be stored alongwith the modified *thing* affixes for the various persons and numbers. A simple illustrative analysis is shown at Annexure-2. These facilitate identification of a given noun-form in terms of ending, gender, case and number or give all the declensions (in case and number) for a word with given ending and gender, and in case of verb-forms, identification of a given verb-form in terms of root, group, usage, mood (tense), person and number or forming all the declensions for a given root etc.

D. Combined/Compound Word Analysis

Representation of combined words is dealt next. There are five types of combinations of words, viz. vowels, consonants, form retention, with 'visarga' and 'svadi', certain special forms. Further, these are of 48 categories as per Annexure-1. The representation of these is quite difficult. To begin with, if the input sentence is suitably annotated to indicate the splitting point for combined words, the analysis could be easier. Next, a 'calibration' phase wherein combined words of all possible types are input with their identification for future use. 'Learning' by the program is thus possible and progressively correct "guessing" of the particular *sandhi* improves. The result after analysis of a combined word is identification of the constituent (split) words and that of the sandhi. In all cases, various possibilities would have to be listed out initially and choice of the relevant answer is to be effected later. This process helps program improvement as well 'calibration', as finer issues involved in representation which are not transparent in human processing would come to the fore. The reasoning behind the choice of a particular answer also needs to be stored simultaneously to make the program 'learning'. Though there are only limited number of combinations (sandhis), there is no general criterion to formalise all the possible combinations since it is subjective and intentions matter. Thus, the interactive approach is inescapable.

Compound words are dealt next in the order. The analysis and representation of this class of words is by far the toughest since practically anything could be intended to mean anything (in a loose manner of speaking, of course!). That is to emphasise

the subjectivity and flexibility associated with the formation of compound words. Here also the type analysis and representing the various cases with examples suitably are required as the base. The general characteristics of the compounding like 'the first word is prominent in avyayibhava, the latter in tatpurusha, both in dvandva, the other (denoted by the constituent words) in bahuvrihi etc. may be used for guidance. Certain technical terms used to denote the parts of speech are to be defined. The initial approach is similar to combined word analysis here also. The output of analysis here is identification of the constituent words with their parts of speech and the samasa type. This has to be in the standard vigraha vakya style. Listing of the possible answers and choice of the appropriate one are to be done as mentioned before. There are certain other word-formations having more than one word. The explanation of these formations termed as vrithi is of five types including compounding. The other four are also to be treated appropriately. Annexure-1 lists these as well. Verb-forms with prefixes also are similar to compound words though they are special forms of verbs.

Karakas (government clauses) fit into both noun and verb form categories due their nominality and action features. These are of six types and their case and auxiliary meanings are needed for processing of word meanings. Meanings of typical noun bases from lexicons etc., verb root meanings from dhatu-pata, and meanings of indeclinables etc. are to be represented in a suitable data structure. Thus word type and meaning analysis leads to sentence signification. Here, various criteria to be applied for Semantic extraction could be expectancy, proximity, compatibility, congruence aspects like commencement, conclusion, emphasis, novelty, effect, figurativeness, propriety, primary and secondary significations, denotation, context, portents, suitability, place, time, correlation with other sources of knowledge etc.

E. The Three Formats of Sentential Import

Logicians, grammarians and Vedic Scholars propound three different forms of verbal import. Logicians treat the agent or object as the main substratum in sentential import. However, this is not too rigidly applied, and exceptions/special cases do exist. Here the verb affix ('thing') meaning qualifies the agent/object, which is in nominative (singular) case. Thus, the doer, qualified by the actions meant by the various governing clauses, is denoted in the sentence. Grammarians treat verb root meaning as primary in a sentence. The agent, instrument etc. qualify the activity denoted by the verb. The governing clauses connect their meanings with the action meant by the verb root suitably. Mimamsakas hold the intention of the agent, i.e., verbal activity with the agent's will, to be the signification of a sentence. Here, verb affix meaning is prominent and the agent is the supporter of the activity while the object is subjected to the activity etc. The Semantics to be extracted from a given sentence is to be in any one of these three formats. The next section, in original, illustrates the formats applied to various sentences.

F. Examples of Sentential Import

The nine types of sentences based on voice are shown in the Annexure. The companion article is elaborate while only the types are mentioned here with examples:

1. Pure causal active voice - Ex: 'Rama shines".
2. Active voice with agent & object - Ex: "Rama ruled the Earth".
3. Passive voice - Ex: "Unique fame was earned on earth by Rama".
4. Active voice (2 agents) - Ex: "Rama (got) built a bridge by monkeys".
5. Active voice (2 objects) - Ex: "Rama took (sent) Ravana to the other world".
6. Passive voice (2 agents) - Ex: "Lanka was rendered to ashes by Hanuman and Rama".
7. Passive voice (2 objects) - Ex: "Ravana was taken (sent) to the other world by Rama".
8. Passive voice (object-agent) - Ex: "Rama is being worshipped".
9. Pure passive voice - Ex: "(It is) being existed by Rama".

G. Implementation Aspects

The various types of 'data' described hitherto are to be put into some suitable structure and a program is to be written for implementation. The various analysis involved are to be broken down into elements and certain *processes (primitives)* are to be defined to cover the general functions involved. With the availability of Sanskrit keyboards and displays, Users would be able to interact with the Computer in Sanskrit. A typical set of such 'processes' follows:

1. Standard ones like Read, Write, Print, Save, etc.
2. Look-up : Refer to stored date, perform certain comparisons (logical) etc.
3. Combine : Expect words, type of sandhi and form combined word.
4. Compound : Similar to the above, in respect of samasas.

5. Separate : Splits a given combined word into constituents.
6. Deform : Splits a compound word into constituents.
7. Identify : Gives relevant details about a particular word.
8. Connect : Uses various criteria and coheres word meanings.
9. Describe : Explains the sentential import in a given format.
10. Convert : Changes voices of sentences, formats of sabda-bodha
11. Complete : Gives all declensions of noun bases, verb roots...

Different users could define further processes as required. If some uniformity is achieved in the 'instruction set' and the basic functions used, it would better interaction and sharing of experiences of various researchers and the like. Even modularity in attempting solutions to various aspects may be possible. A flow-chart appended as Annexure-3 outlines the principle of Sabda-Bodha representation which could be discussed and collective efforts may be launched to realise its implementation.

H. Application Areas

The advantages of Computer being Speed, Large memory content and ability to repititively carry out tasks as per instructions. These factors have inspired exploring the use of Computers for tasks that are too demanding manually, but which require reasoning to precede action. This is how knowledge representation became important and once we succeed in it, the following tasks could be achieved using the machine.

Traditional Sanskrit study has been classified into 14 distinct branches as described further. The four vedas, i.e., rg, yajura, sama & atharvana, their six angas (auxiliaries), i.e., siksha (exegetics), vyakarana (grammar), chandhas (prosody), niruktha (etymology), jyauthisha (astronomy) and kalpa (vedic ritual), mimamsa (study of vedic text for interpretation of meaning), nyaya (syllogism for vedic applications), puranas (epics) and dharma shastharas (codes of moral rectitude) are the 14 vidhyasthanas (branches of Learning) referred above. In all these branches, the basic tenets are condensed in a few pithy, terse and precise suthras (Aphorisms) which are in sentential form but require critical study and analysis. These form the fittest subject for in-depth study to be able to grasp the fundamentals of definition and representation of knowledge as found in Oriental works. The Sanskrit article gives detailed topics for in-depth study and analysis in all these areas.

I. Conclusion

The Author wishes to express his gratitude for the opportunity provided to submit this paper to this Conference. The Author's suggestion is to divide the task into three segments aimed at entrants, graduates and researchers categories. The first could learn Sanskrit using Computer, while the graduate level would provide deeper understanding of the basic system for specialisation. The researchers would actually write 'Application Programs' in the areas of their interest.

As a small step, the Author has set out to represent simple nouns and verbs as aforesaid, though using english script. Usage and potential having a mutual dependence is hampering Sanskrit's use as a programming medium. With the awareness increasing gradually, the situation might improve is what is hoped. Towards this direction, Conferences like the present one, Symposia, Courses, Workshops, Training etc. may be necessary. The treasure contained in Oriental works could be unearthed with committed and sustained efforts and trends could be reversed wherein India could once again lead the world in Intellectual exercises.

With institutional support and timely positive action, the last genre of traditional scholars left in the country at present could still be motivated to revive the fast-fading virtue of intellectual honesty/enquiry.

COMPUTER REPRESENTATION OF SABDA-BODHA — ANNEXURE-2