

Sanskrit as a Programming Language: Possibilities & Difficulties

Vipin Mishra

Abstract

In the past twenty years, much time, effort, and money has been expended on designing an unambiguous representation of natural languages to make them accessible to computer processing. These efforts have centered on creating schemata designed to parallel logical relations with relations expressed by the syntax and semantics of natural languages, which are clearly cumbersome and ambiguous in their function as vehicles for the transmission of logical data. Understandably, there is a widespread belief that natural languages are unsuitable for the transmission of many ideas that artificial languages can render with great precision and mathematical rigor. But this dichotomy, which has served as a premise underlying much work in the areas of linguistics and artificial intelligence, is a false one. There is at least one language, Sanskrit, which for the duration of almost 1,000 years was a living spoken language with a considerable literature of its own. Besides works of literary value, there was a long philosophical and grammatical tradition that has continued to exist with undiminished vigor until the present century. Among the accomplishments of the grammarians can be reckoned a method for paraphrasing Sanskrit in a manner that is identical not only in essence but in form with current work in Artificial Intelligence. This article demonstrates that a natural language can serve as an artificial language also, and that much work in AI has been reinventing a wheel millennia old.

Keywords: Computer, Sanskrit, Grammar, Panani Ashtadhyayi, Vibhakti, Karaka, NLP, Sandhi

1. Introduction

पदवाक्यप्रमाणशास्त्रेभ्यः (व्याकरणमीमांसान्यायादिभ्यः) अर्थनिर्धारणार्थानां विधिकल्पानां प्रकाराणाञ्च

(Algorithms and methods for determining meanings) सङ्कलनम् अत्र ।

Language, in other words the storehouse of all human Knowledge is represented by words and meanings. Language by itself has an Ontological structure, Epistemological underpinnings and Grammar. Across languages, even though words /usages differ, the concept of meanings remain the same in respective communications. Yet the "Meanings" are understood by human beings

based on Contextual, Relative, Tonal and Gestural basis. The dictionary meanings or 'as it is' meanings are taken rarely into consideration, thus human language is ambiguous in one sense and flexible in other.

Computers on the other hand are hard-coded to go by the dictionary meanings. Thus teaching (programming) Computers to understand natural language (human language) has been the biggest challenge haunting Scientists ever since the idea of Artificial Intelligence (AI) came into existence. In addition this has led to the obvious question of "What is intelligence" from a Computation perspective. Defining intelligence precisely being impossible, this field of study has taken many shapes such as Computational Linguistics, Natural Language Processing and "Machine Learning" etc. Artificial Intelligence instead of being used as a blanket term, is now being used increasingly as "Analytics" in many critical applications.

Sanskrit being the oldest is also the most Scientific and Structured language. Sanskrit has many hidden Algorithms built into it as part of its vast scientific treatises, for analyzing "Meanings" or "Word sense" from many perspectives since time immemorial. "It is perhaps our job to discover and convert the scientific methods inherent in Sanskrit into usable Computational models and Tools for Natural Language Processing rather than reinventing the wheel" - as some Scientists put it. This blog's purpose is to expose some of the hidden intricate tools and methodologies used in Sanskrit for centuries to derive precise meanings of human language, to a larger audience particularly Computational Linguists for further study, analysis and deployment in Natural Language Processing.

In addition, Sanskrit even though being flexible as a human language, is the least ambiguous as the structure of the language is precisely defined from a semantically and syntactical point of view. From a Psycholinguistic perspective this blog could also give us a glimpse of the advanced linguistic capabilities of our forefathers as well their highly disciplined approach towards the structure and usage.

2. Sanskrit in Computational Linguistics

First of all the confusion that needs to be cleared is whether Sanskrit is best suited for Computing or Computer Programming - my view is both. Yet this paper is not about Sanskrit as a computer programming tool - even though there are scientists and academicians who are developing programming languages based on Paninian principles, however this paper deals with Sanskrit as a Computing tool. Computing here refers to concepts, algorithms and methodologies.

Computer Programming is an entirely different thing as it deals with a human being generating code in a high level computer language, which in-turn translated to a low level code through

compilers /linkers, which in-turn translated to operating system instructions, which in-turn translated to microprocessor instructions (based on CPU instruction set) which internally converted into binary instructions which further converted to digital electronic (electrical) signals for flip-flops /counters etc.

The entire chain of programming is based on mathematics /symbol language and not any human language spoken or written - even though the symbols consists of few human understandable characters such as numeric 1-9, alphabetical a-z, and some signs of mathematics such as +, -, /, %, etc. all these constitute the ASCII - which has 255 characters or symbols of computer codes - in other words each symbol can fit into a single byte. These symbols are assigned to certain operative values in digital electronics - thus the programming languages are not human languages. That's the precise reason why human beings want Natural Language Processing or human language processing capabilities in computers - which literally means our languages being understood by computers. So far computer understands only computer language and human beings only human language.



With respect to computer language the instructions (let's say commands /actions /verbs) are very limited - widely used are about 15 - such as go to, break, compare, copy, reverse, assign, operators (+,-,*,/), receive, display, etc. Also few other actions (verbs) can be written as functions such as sort, list, etc. Thus the computer language's capability in comparison with human language is very limited.

In comparison in Ashtadyayi - Panini's 1000s of years old Sanskrit grammar treatise - the meta language used inside Ashtadyayi not Sanskrit but uses certain words and rules of Sanskrit - which is used to teach Sanskrit grammar to the readers of Ashtadyayi. That meta language has more instruction sets - yet without using any explicit verb. Thus if one can make a high-level programming language exactly mimicking the meta language of Ashtadyayi - we will have a

powerful tool - with which computer can generate words, form sentences, etc. - Yet associating meanings will be the biggest challenge.

The hypothesis is that - if there is a highly structured human language, can then that language be used for Natural Language Processing? - The answer is yes and to wonderful degree containing complex human sentences - how? = In Linguistics and most importantly computational linguistics the following are essential for scientific analysis (for computers to do the analysis) of the language - which consists of sentences - and sentences have inherent meanings.

How then analysis can take place? –

Without ambiguity the sentence meaning being conveyed is first and most important thing; because computers don't have intelligence - computer's understanding of language is based on a particular structure (let's say a word (or) phrase and its meaning) and it tries to combine or mix and match to a particular meaning. Again here the computer doesn't care about the meaning but it responds for a question which has a particular meaning and based on that from a set of answers the most suitable answer is chosen and given - which again based on a particular individual, popularity, number of occurrences, place, time, etc.

In natural language for example sentences like –

a- The committee chair chairs the meetings where the chair is elected as the chair for one more chair-term.

b- All committees' chairs chair their meetings to elect the chair and the past chair is elected as the new chair for the next chair-term.

Now, we can easily understand the meanings of these sentences, but computer can't understand - here is where the language's ambiguities with respect to word meanings and words' declensions, usage, phrase meanings and along with other phrases and within a sentences - many such things matter.

English is the most complex language - it takes even for a native speaker 8-10 years to achieve proficiency. It takes just 2 years to achieve proficiency in Sanskrit another 2 years in literary Sanskrit. In addition, the written form of English is again non-phonetic which adds its own problems in converting text to voice. In addition due to many borrowed words - spellings and pronunciations are again differ and add complexity. More over the regional flavours.

The interpretation of a text in computing goes through - first Parsing, part of speech tagging, lexical analysis, morphological analysis, syntactical analysis, and then semantical analysis. Thus the more structured and scientific the language is the less problems in computer based natural

language processing and its applications such as Machine Translation and Machine Assisted Translation.

In general Language means - collection of sentences; A sentence means - collection of meaningfully associated words such as Subject, Object, Verb etc.; Each word in a sentence should have clear and easily understandable verbal and nominal declensions if not confusion starts. In the above example the word "Chair" is both verb and noun - this gives enormous confusion to computer.

Let's explore further... Human language is highly ambiguous. Primarily - because of the ambiguities of word sense (meanings) word meanings (on their own) and in association with another word (in a phrase) and in association with verbs and other words in a sentence - the complexities multiply. With quotations and idioms complexities only increase further in a sentence or a part of speech. Now add acronyms and what we get? - Most complex thing known to human being next only to human mind (or) both language and mind are one and the same?

Thus in linguistic terms the complexity is exponentially increased in each corresponding step as per the 6 most important things in the order in Linguistics and how they are in Sanskrit are given below:

(1) Phonetics and Phonology —knowledge about linguistic sounds - In Sanskrit it is known as Shiksha shastra - Sanskrit has over 40 Shikshaas for each shaaka of Veda but for language in general Paniniya shiksha is most suitable as it correspondingly connects to the Grammar and the rules of the grammar also abide by the rules of the Phonetics.

(2) Morphology —knowledge of the meaningful components of words from stems and their generation and usage - In Sanskrit this is called as 'pada vyutpatti' in Sanskrit - 4 types of vrittis (word generators) are there for this purpose namely - Krit, Taddhita, Samasa and Sannaadyanta - In addition the method for generating words are also explained step-by-step in Panini's Ashtadyayi like a mathematical equation - thus programming to generate words are easiest.

(3) Lexical —knowledge of meanings and equivalent words. Every Sanskrit lexical item has a one-one correspondence. So a particular word used in some place means the same when used elsewhere too from a semantics point of view. Amara Kosha, Nirukta, Nighantu all have the complete lexical database of Sanskrit words and associated word connections.

(4) Syntax —knowledge of the structural relationships between words - declensions of nominal

forms /stems - In Sanskrit Vibhakti play this role - we have very tight rule thus there is no ambiguity. Also Sanskrit is a language without prepositions thus a major complexity is removed - this is also explained step-by-step in Ashtadyayi like a mathematical equation.

(5) Semantics —knowledge of meaning of words in a sentence - In Sanskrit this is one discussed in detail in many works and in Sanskrit vyakarana called as "Kaarakam" - Many ways of sentence meanings and their analysis on a scientific basis are available in Sanskrit with respect to different schools of linguistic sciences such as Vyakarana, Nyaya and Mimamsa.

(6) Pragmatics — knowledge of the relationship of meaning with respect to the context - this is the most complex as meanings change based on context and many other factors - In Sanskrit there is a wonderful Vyakarana treatise available for pragmatics called as "Vakyapadiyam" by Maharishi Bhartrhari - it is pity that many Sanskritists are not aware of this. But this treatise is very popular among European linguists in particular German, Belgian, and French.

All most all kinds of meaning analysis based on relationship between 2 words - sameness, opposites, connection, association, context, etc. are dealt in detail in "Vakyapadiyam", in the West Scholars are been inspired by this and made their theories of Semantics and Pragmatics.

3. Analysis

"Pāṇini's work became known in 19th-century Europe, where it influenced modern linguistics initially through Franz Bopp, who mainly looked at Pāṇini. Subsequently, a wider body of work influenced Sanskrit scholars such as Ferdinand de Saussure, Leonard Bloomfield, and Roman Jakobson. Frits Staal (1930-2012) discussed the impact of Indian ideas on language in Europe. After outlining the various aspects of the contact, Staal notes that the idea of formal rules in language – proposed by Ferdinand de Saussure in 1894 and developed by Noam Chomsky in 1957 – has origins in the European exposure to the formal rules of Pāṇinian grammar. In particular, de Saussure, who lectured on Sanskrit for three decades, may have been influenced by Pāṇini and Bhartrihari; his idea of the unity of signifier-signified in the sign somewhat resembles the notion of Sphoṭa. More importantly, the very idea that formal rules can be applied to areas outside of logic or mathematics may itself have been catalyzed by Europe's contact with the work of Sanskrit grammarians.

Pāṇini, and the later Indian linguist Bhartrihari, had a significant influence on many of the foundational ideas proposed by Ferdinand de Saussure, professor of Sanskrit, who is widely considered the father of modern structural linguistics. Saussure himself cited Indian grammar as an influence on some of his ideas. In his *Memoire sur le systeme primitif des voyelles dans les langues indo-europennes* (Memoir on the Original System of Vowels in the Indo-European Languages) published in 1879, he mentions Indian grammar as an influence on his idea that "reduplicated aorist represent imperfects of a verbal class." In his *De l'emploi du genitif absolu*

en sanscrit (On the Use of the Genitive Absolute in Sanskrit) published in 1881, he specifically mentions Pāṇini as an influence on the work."

4. Conclusion

Sanskrit referred as a Devabhasa (Gods language) is not because it is the oldest language - it is because it is very perfect in its structure, morphology, semantics, etc. which have not changed for 1000s of years - Only a good linguist understands that such a perfect language can't be created by Human beings, neither Cavemen nor evolved - What's the evidence: we have seen English evolving from a structured languages and having some formal structures and usage initially to now with no structure and highly ambiguous! - Not with respect to human understanding but with respect to Linguistics. With respect to human understanding it has become easy and flexible - as a result can we say that human mind has become unstructured and dull !, may be past century scientists without labs and tools have found many things !. Mathematician Ramanujan's tools were just a pencil and a paper!

5. References

- [1] The science of language, Chapter 16, in Gavin D. Flood, ed. The Blackwell Companion to Hinduism Blackwell Publishing, 2003, 599 pages ISBN 0-631-21535-2, ISBN 978-0-631-21535-6. p. 357-358.
- [2] George Cardona (2000), "Book review: Pāṇini's Grammatik", Journal of the American Oriental Society 120 (July– September, 2000): 464–5, JSTOR 606023.
- [3] Leonard Bloomfield (1927). "On some rules of Pāṇini". Journal of the American Oriental Society (American Oriental Society) 47: 61–70. doi:10.2307/593241. JSTOR 593241.
- [4] Ashtadyayi Reference: <http://avagraha.wordpress.com/>
- [5] Sanskrit Programming - Reference 2 sites both contains lot of information - <http://vagartham.blogspot.in/> and <http://uttishthabharata.wordpress.com/>
- [6] Functional programming - Reference: <http://vishk.wordpress.com/2007/02/11/backus-naur-form-and-ashtadyayisanskrit-grammar/>

[7] Parser /Tokenizer for Samasa - Vaakkriti: Sanskrit Tokenizer, Aasish Pappu and Ratna Sanyal, Indian Institute of Information Technology, Allahabad (U.P.), India, Proceedings from the paper submitted in Third International Joint Conference on Natural Language Processing, 2008, Hyderabad, India

[8] The methods used inside the ashtadyayi is similar to today's arrays, inheritance (including multiple inheritance), polymorphism, etc. used in OOPS - Reference: Recent Research in Science and Technology, 2011, 3(7): 109-111, ISSN: 2076-5061, www.scholarjournals.org

[9] Computational Lingusitics - Reference: Hyman Malcolm D., "From Pāninian Sandhi to Finite State Calculus", Sanskrit Computational Linguistics: First and Second International Symposia, Revised Selected and Invited Papers, ISBN: 978-3-642-00154-3, Springer-Verlag, 2009.
