# <u>DESIKA</u> - AN APPRAISAL OF THE NATURAL LANGUAGE UNDERSTANDING PROJECT FOR SANSKRIT

Natural Language Understanding involves representation in Computer of knowledge pertaining to words of a natural language, their denoted objects, their environment, and their related objects.

Various linguistic theories describe the mechanisms of word sense determination and word-meaning relationships. This topic is quite complex for representation in a Computer owing to the fuzzy or qualitative nature of its logic/underlying principle.

To a large extent, a study based on ontological classification of the domain represented by linguistic elements and an empirical or heuristic-based connection and inference mechanism may help the issues involved to be understood better.

Ancient Indian Sciences have discussed this aspect in great detail and cover the subject of linguistic information exchange process exhaustively. All these works are in Sanskrit, a natural language with a well-structured grammar. It will be beneficial to pursue the study of the pertinent literature in this regard, model the concepts therein and build a prototype system to weigh the relative merits of the Indian system of linguistic analysis.

Following is a short appraisal of a possible effort with a fixed timeframe, clear output specification and a defined method.

OBJECTIVE: A. LINGUISTIC: -

'UNDERSTANDING' a natural language input (in the form of an isolated sentence, for example).

[This understanding should be demonstrated in one of the following ways:

1. BY PARAPHRASING THE INPUT {by changing the voice of the construction (active to passive and vice-versa), by synonymous explanation/description, or in a "classically structured format", described in the Indian Sciences and termed as "যাৰ্ব্ৰोध:"}.

2. BY ANSWERING QUERIES ON THE INPUT (which involves Inference, Connection, Coherence, Reference etc.).

3. BY SUMMARISING THE INPUT (in the form of a precis or gist, by picking the key thoughts in the input).

B. COMPUTATIONAL: -

1. Developing suitable custom-made tools for NLP, creation of function/ procedure libraries, data libraries in defined formats for networked mode operation, tools for lexicon building, creation of Dictionary, Thesaurus, Concordance etc.

2. Devising appropriate technologies for linguistic data storage, processing and 392 Parankushachar Institute of Vedic Studies®, Bengaluru retrieval. These could include Parallelisation, proper choice of the data structures, hierarchies, inheritance, chaining procedures, indexing, quick accessing, programming language modifications to accommodate linguistic processing needs, means of networking of databases created etc. (In fact, an attempt to simulate the features of Panini's Ashtadhyayi by a Programming Language would be amply rewarding, but quite challenging).

3. DEVELOPING A LANGUAGE-INDEPENDENT KNOWLEDGE REPRESENT-ATION SCHEME BASED ON ANCIENT INDIAN WORKS.

4. Design of a 'Linguistic Workstation' aimed at helping Language research through Computers.

5. Critical comparison of available AI languages for suitability to linguistic studies.

NOTE: THE PROTOTYPE WILL BE BUILT FOR SANSKRIT.

Reason for choosing Sanskrit for the prototype:

1. Richness of available literature on the subject and the depth and exhaustive nature of the analysis.

2. Clearly laid out frames of reference and grammatical and logical structure of the language being precise, thoroughly consistent and transparent.

3. The analytical methods are easily extendable to other Indian languages (and Pali & Prakrit), including some of the results. This would be helpful in Machine Translation.

4. The phonetic based nature of this language will be useful in Speech input case as well. The grammar also extensively covers the spoken aspects of the language including accent.

5. The resources are ENTIRELY INDIGENOUS.

# A. CONCEPTUAL ISSUES

# Introduction

Natural Language as a mode of giving expression to thought processes in written (or spoken) form is a powerful medium for information exchange. Since it is a conventional system based on symbols (for writing) and a set of myriad factors shared by a large number of native speakers, who also possess large amount of background or world knowledge, the effectiveness of communication in the human case is quite high, there being various means of self-correction/ orientation even with incomplete or incorrect data.

While attempting computational study of linguistic processes, a good knowledge about the theoretical basis for the denotation of meanings by words is essential. Also, multiple senses of words need to be analysed for the underlying principles (including the etymological and exegetical aspects). Here, the ancient Indian systems have devised analysis of what is termed as "शब्दशक्ति" or potency of words to denote one or more objects, physical as well as abstract, genus as

well as specific, entity or action, real or non-existent etc. This analysis describes the process of human cognition (understanding) of verbal inputs and is what is precisely meant by शाब्दबोध or शक्ति-ग्रह or व्युत्पत्ति.

There is also a need for typical classification of objects based on essential and additional characteristics for each type of a minimum set of well-defined categories. This also has been described very precisely and exhaustively in the ancient Indian treatises under the topic of "पदार्थविभागः".

Even about the method of analysis, there are copious details as to the frame of reference, sources and objects of knowledge, their domain of applicability, validity, means of testing, possible defects, theory of error, perception process per se, its fallacies, logic, its applicability and limitations, principles of deduction and inference, process of comprehension of meanings of words etc.

Thus, it is felt that a system based on the study of the three major branches of ancient Indian intellectual enquiry, viz. Logic (न्यायशास्त्र), Grammar (व्याकरणशास्त्र) and Study of Vedic text interpretation (मीमांसाशास्त्र) will throw light on many of the basic issues mentioned. We will also get a good insight into the computational aspects of language processing by this effort.

As शाब्दबोध is a key contribution to study of language and its understanding process (शक्तिग्रह) by the ancient Indian thinkers, the system proposed will be based on the technical material given in the following authoritative sources for the various issues and viewpoints:

1. Paninian Workbench: - Besides the five works of Panini, वार्तिक of कात्यायन; महाभाष्य of पतञ्जलि; काशिका of जयादित्य-वामन; सिद्धान्तकौमुदी of भट्टोजीदीक्षित; and its commentaries. This will be the main reference initially, especially for the choice of topics and modules.

2. शाब्दबोधः: - नैयायिकमतम्: शब्द-शक्ति-प्रकाशिका of जगदीशतर्कालङ्कार; न्यायमञ्जरी of जयन्तभट्ट; मुक्तावली of विश्वनाथ; तर्कसंग्रह of अन्नंभट्ट;

वैयाकरणमतम्ः वाक्यपदीय of भर्तृहरि; वैयाकरणभूषण of कौण्डभट्ट; वैयाकरण-सिद्धान्त-लघुमञ्जूषा of नागेश; लघुशब्देन्दुशेखर of नागेश

मीमांसकमतम्: भाट्टरहस्यम् of खण्डदेव; श्लोक/तत्र वार्तिक of कुमारिलभट्ट

वेदान्तमतम्: न्यायपरिशुद्धि of वेदान्तदेशिक; तत्त्वमुक्ताकलाप of वेदान्तदेशिक

With the help of domain experts, the finer aspects contained in works like शक्तिवाद, प्रामाण्यवाद, ज्ञप्तिवाद and व्युत्पत्तिवाद of गदाधर etc. would also be consulted and incorporated.

Technical Issues:

Any word in Sanskrit is made up of two components, namely, the base and the affix. The base denotes the object intended by the word while the affix part provides additional information by suitably modifying (or declining) the base. Affix is a common word to cover prefix (as in verbs and derivatives), in-fix (as in case of compound words) and suffix (as in nominal and verbal forms). Various grammatical case or governing relationships, number and gender are

denoted by these declensions.

There are six cases as Agent (कर्ता), Object (कर्म), Instrument (हेत्), Recipient/ Beneficiary (संप्रदान), Reference for separation (अपादान) and location (अधिकरण). These designate the functional relationship of the word with the activity marker (verb) present in a sentence. Thus, any word inherently and integrally contains the semantic information for sense coherence in a sentence.

This aspect of Sanskrit is responsible for simplifying the process of semantic extraction and makes it deterministic. The rules of grammar provide the framework for the formation (and so analysis, to an extent) of a word for a given information to be conveyed.

Words in Sanskrit are classified into two major types based on the affixes that get attached to the bases, as nominal class or सुबन्त (literally those which end with सुप suffixes) and verbal class or तिङन्त (literally those ending with तिङ् suffixes). The former ones decline (or get inflected) as aforesaid into cases to denote various functional relationships. There are about 325 types of सुबन्त based on the declensions and cover all nominal word forms including actival class or primary derivatives (कृदन्त), secondary derivatives (तिद्धित) and compound words (समास).

The latter type decline to indicate various tenses/moods, voices and other aspects like causal, desiderative, intensive, reflexive, abstract forms etc. along with person and number. There are about 2000 verbal roots in धातुपाठ and these decline into a few lakhs of forms on the above basis. This is the richest category of word-forms in Sanskrit.

There are certain Indeclinable word forms which are mostly particles, infinitives, participles etc. and these technically come under the nominal class. Semantically, these modify the word meanings in an adverbial, adjectival or substantive manner. Also, certain adverbial prefixes come under the indeclinable category. These are of about 100 types covering a list of about 700 words and procedures for deriving many more.

Since the word types are few, their classification and coding on parametric and declensional basis is relatively easier. This is useful in analysis of word senses to a great extent as will be shown.

The codes are devised for the ending, gender, sub-type (from the declension point of view and string pattern matching), suffix type or finished form, case and number for सुबन्तs, गण, पदी, इत्, कर्म, इट्, अन्त, स्वर, लुङ् and लिट् types etc. for तिङन्तs and प्रतिपदोक्त, लाक्षणिक, कृदन्त, तद्धितान्त, and their sub-types etc. for अव्ययs.

Having known the word types and codifying them, a lexicon is built containing the commonly used words with their codes. Here, the gender determination rules have enlisted certain words with given genders, while for the rest generic conditions are spelt out.

To begin with, a vocabulary of about 3000 words have been chosen, which represent all the possible word types, so that while growing the dictionary also it is done in a categorised manner rather than in a natural order. The धातुपाठ

## PIVS Silver Jubilee Souvenir

has been codified as described above and is the verbal root lexicon for the system. A list of the various types of  $3\overline{3}\overline{3}\overline{3}\overline{3}\overline{3}\overline{3}$  with codes aforesaid is the lexicon for indeclinables.

The database is then prepared which, in the case of  $\overline{4}$ ,  $\overline{4}$ ,  $\overline{4}$ ,  $\overline{4}$ ,  $\overline{5}$  is in the form of list of suffixes for each distinct type of base declension (modified  $\overline{4}$ ,  $\overline{4}$ ,  $\overline{4}$ ,  $\overline{4}$ ,  $\overline{5}$ ,  $\overline{$ 

The word level analysis then proceeds as follows:

The word is checked in the lexicon for अव्यय (प्रतिपदोक्त) first and in case of a match, the identification is put out in terms of the type and meaning category. (e.g., यत्, तदा, चिरस्य, आरात् etc.)

It is then checked for derived (लाक्षणिक) अव्ययs with the proper module (rulebased) and in case of a match, the proper identification is put out. (e.g., कृत्वा, अधिहरि, गन्तुम्, पायम्)

Then, it is checked for सुबन्त against finished forms. The results of both these checks notwithstanding, the given word is 'exploded' into each of its component vowel and consonant parts, starting from the right end, by a process of matching the RHS part with the list of suffixes and the LHS part against the lexicon for match in the string form as well as the type specification (code).

When a match is found, the identification of the word is put out and the process continued till the left end of the word.

For compound words, further level of explode is used which is described in a short while.

The process is repeated for तिङन्तs similarly with RHS part being checked against suffixes' list and LHS part against धातुपाठ, with proper checks for गण, विकरण, पदी, इट, and the tense/moods.

Additionally, check for presence of verbal prefixes (उपसर्ग) is done starting with the left end and in consultation with the list of उपसर्गs. The सन्धि rules are invoked at this stage.

By increasing the level of complexity, i.e., by invoking सन्धि at each "chop" of the 'explode' cycle and checking as above, even highly literary constructions (including compound words) can also be analysed.

A few illustrations will now make these points clear.

For रामः, the identifications are as a सुबन्त of अकारान्त पुल्लिङ्ग प्रथमैक वचन (with राम as प्रातिपदिक) and as a तिङन्त of अदादि गण, परस्मै पदी, लट्, उत्तम पुरुष बहुवचन (रा as धातु).

The word विश्वस्य will be identified as a सुबन्त of अकारान्त सर्वनाम पुम्/नपुंसक लिङ्ग षष्ठोक वचन (with सर्व as प्रातिपदिक) and a ल्यबन्त अव्यय of अदादि गण, परस्मै पदी (with श्वस् as धातु). A word like मात्रा is identified as a सुबन्त of ऋकारान्त स्त्रीलिङ्ग तृतीया एकवचन (with मातृ as प्रातिपदिक) and as a सुबन्त of आकारान्त स्त्रीलिङ्ग प्रथमैक वचन (with मात्रा as प्रातिपदिक).

A word like नरः is identified as a सुबन्त of अकारान्त पुल्लिङ्ग प्रथमा एकवचन (with नर as प्रातिपदिक) and a सुबन्त of ऋकारान्त पुल्लिङ्ग प्रथमा बहुवचन (with नृ as प्रातिपदिक).

The word श्वेतः is identified in the first level as a single सुबन्त word of अकारान्त पुल्लिङ्ग प्रथम एक वचन (with श्वेत as प्रातिपदिक) while exploding further, we get श्वा + इतः also as another valid result.

Thus, the word मातः is identified as a सुबन्त of ऋकारान्त स्त्रीलिङ्ग संबोधन एकवचन (with मातृ as प्रातिपदिक) and as a तिङन्त of अदादि गण, परस्मै पदी, लट्, प्रथम पुरुष द्विवचन (मा as धातु) initially and later as मा + अतः etc.

For compound words, the chopping process at the second level would yield options wherein more than one प्रातिपदिक will be there in the LHS part or an अव्यय and प्रातिपदिक or an उपसर्ग and धातु or प्रातिपदिक etc. Thus, applying the characteristic conditions to determine the समास information can be attempted.

For example, a word like शुक्लांबरधरम् during the second level of chopping (which includes सन्धि module), would give शुक्ल-अंबर-धर as possible प्रातिपदिकs, indicating a compound word. So also is the case with words like अधिहरि, अभिभवति, अधीश्वर etc.

All possible identifications are thus brought together and form the grammatically valid set for further sense determination and disambiguation. It can be stressed here that without such an exhaustive list of grammatically valid possibilities, a real help to linguistic research cannot be achieved and also that it is not very difficult to realise this till this stage on a computer. In fact, this stage is already nearing completion now.

This exercise is repeated with each word of the input sentence and all possible identifications of all the words are collected in this (word-analysis) stage.

Now is the question of using these word level identification to arrive at sentential import using empirical/heuristic rule base method. There are six case-relations (कारकसंबन्ध) and a host of generic relations (संबन्धसामान्य) for seven (or eight) cases (विभक्ति) as identified at the word analysis level. There are many word-specific case mappings (उपपदविभक्ति) as well. The mapping between case (विभक्ति) and its relation (कारक) is used to decide the possible functional relationships.

All the प्रत्ययs described in grammar rules have associated contextual or semantic based situations governing their choice. This aspect is used in determining the semantic/contextual fitness of syntactically valid alternatives. At this stage, certain options may get closed due to incompatibility in concept classes of words etc. If still several (more than one) valid options remain, then logical criteria like expectancy, proximity, compatibility etc. are applied to select the unique import(s).

The other aspect of Generation of all (or particular) forms of सुबन्तs and तिङन्तs can also be handled by using the database created as above. Here, input word base in case of सुबन्तs should be in the प्रातिपदिक form, e.g., राम, पितृ, राजन्, विद्वस्,

# <u>PIVS Silver Jubilee Souvenir</u>

रमा, गौरी, अप, छदिस्, ज्ञान, मधु, सम्यश्च, स्वनडुहू, तद्, किम्, त्रि etc. Then, the word is checked in the lexicon for type information and then, the सुप् forms are generated with the proper title etc.

If the word is not found in the lexicon, then the लिङ्गानुशासन rules are consulted to arrive at the gender and type (using rules for स/षत्व, न/णत्व etc.) and the proper forms are generated. Where लिङ्गानुशासन rules specify typical genders based on the suffixes, proper functions among कृदन्तs, उणादि etc. are invoked and the type information generated before the declensions are output.

A package for forming or analysing euphonic combinations in text matter, words or between words is a standalone function used for word-parsing analyses as mentioned above (in second level of chopping). The Vedic accent-related euphonic combinations are also incorporated (वर्ण/स्वर सन्धि) to complete the module and aid in Vedic analysis as well.

The objectives set out in the beginning are to address these types of basic issues which hold a key to success in NLP and it is here that the ancient Indian treatises promise an advantage and a possible breakthrough.

# B. IMPLEMENTATION ISSUES

# FEATURES OF THE SYSTEM:

1. A Computational Platform based on the five Paninian works related to Descriptive Grammar rules (अष्टाध्यायी), Vocabulary of verbal roots (धातुपाठ), List of nominal word groups based on syntactic properties (गणपाठ), Rules for Determination of nominal genders (लिङ्गानुशासन) and Rules of Phonetics (पाणिनीय शिक्षा). This Platform will form the core of the system.

2. Topic-wise modules for various aspects of the language will be developed, both for Generation and analysis, like nominal forms (सुबन्त), verbal forms (तिङन्त), Indeclinables (अव्यय), Actival Particles (कृदन्त), word combinations (सन्धि) and Compounding (समास), Technical Definitions (संज्ञा), Metarules (परिभाषा), Interpretive Rules (न्याय) etc.

3. Appropriately categorised database for all these modules aimed at standardisation of research efforts in the area and aid sharing of data through networking. Exhaustive coverage of the subject under each module ensures good utility and authenticity.

4. Good user interface with windows and proper messages at every stage of analysis/generation/consultation modes.

5. Suitable Application programs development for in-depth study of any particular topic, like Vedic grammar, word derivation and etymological analysis etc. by providing for a development shell.

# METHODOLOGY TO BE FOLLOWED:

1. Analysis of the total reach of the subject. (Total no. of rules applicable, the extent and type of words etc. covered, the kind of processing involved, frequency...)

2. Devising suitable parameters for classification of the subject from the point of view of computation.

3. Formulating a scheme of coding the subject in terms of the chosen parameters, for computational purposes.

4. Codifying the subject knowledge in terms of the codes for the parameters arrived at, in a classified manner.

5. Creation of a vocabulary with code cross-references and indexing as the local lexicon.

6. Identifying the set of global functions and procedures applicable for the topic from among the definitions (संज्ञा), metarules (परिभाषा) etc.

7. Defining the inputs, intermediate results and outputs from the module for various modes of use like analysis, generation, consultation/reference, tutoring etc.

8. Packaging for various possible applications as above.

# TECHNICAL DETAILS: LINGUISTIC: -

The various types of modules planned have been indicated earlier. A few illustrative details are enumerated here for better clarity.

For सुबन्तs:

- 1. SUPCODES Codes for the classification parameters.
- 2. SUPTYPES Characteristic distinction of types.
- 3. LEXSUP Categorised and coded Vocabulary.
- 4. SUPRULE Applicable set of Rules and functions etc.
- 5. SUPSUF Database of suffixes' list.
- 6. SUPFIN Database of finished forms.

For तिङन्तs:

- 1. TINCODES Codes for the classification parameters.
- 2. TINTYPES Characteristic distinction of types.
- 3. LEXTIN Categorised and coded Vocabulary.
- 4. TINRULE Applicable set of Rules and functions etc.
- 5. DHATUPAT Database of Verbal roots.
- 6. UPASARG Database of verbal prefixes and their effect

For अव्ययs:

1. AVYCODES - Codes for the classification parameters.

2. AVYTYPES - Characteristic distinction of types.

3. LEXAVY - Categorised and coded Vocabulary.

AVYRULE - Applicable set of Rules and functions etc.
For प्रत्ययs (कृदन्त & तद्धित):

1. PRTCODES - Codes for the classification parameters.

2. PRTTYPES - Characteristic distinction of types.

3. LEXPRT - Categorised and coded Vocabulary.

4. PRTRULE - Applicable set of Rules and functions etc. For सन्धिs:

1. SANCODES - Codes for the classification parameters.

2. SANTYPES - Characteristic distinction of types.

3. LEXSAN - Categorised and coded Vocabulary.

4. SANRULE - Applicable set of Rules and functions etc. For समासs:

1. SAMASCOD - Codes for the classification parameters.

2. SAMASTYP - Characteristic distinction of types.

3. LEXSAMAS - Categorised and coded Vocabulary.

4. SAMASRUL - Applicable set of Rules and functions etc. For वाक्यs:

1. SENTCODE - Codes for the classification parameters.

2. SENTTYPE - Characteristic distinction of types.

3. LEXSENT - Categorised and coded Vocabulary.

4. SENTRULE - Applicable set of Rules and functions etc.

For शाब्द-बोध:

1. BODHCODE - Codes for the classification parameters.

2. BODHTYPE - Characteristic distinction of types.

3. LEXBODH - Categorised and coded Vocabulary.

4. BODHRULE - Applicable set of Rules and functions etc.

A library of functions and procedures incorporating the metarule contents are developed typically for the following types:

1. For Technical Definitions -

वृद्धि	घ	टि	करण
गुण	संख्या	उपधा	अधिकरण
प्रत्याहार	षट्	वृद्ध	कर्ता
भातु	निष्ठा	सार्वधातुक	उपसर्ग
लोप	सर्वनामन्	आत्मनेपद	गति
आर्धधातुक	विभाषा	परस्मैपद	कर्मप्रवचनीय
इत्	समास	उदात्त	प्रथम
इट्	बहुव्रीहि	अनुदात्त	मध्यम
संयोग	द्वन्द्व	स्वरित	उत्तम
अनुनासिक	अव्यय	एकश्रुति	संहिता
सवर्ण	तद्धित	अपृक्त	अवसान
एकवचन	विभक्ति	तत्पुरुष	द्विगु
द्विवचन	कृत्	कर्मधारय	कृत्याः
बहुवचन	अव्ययीभाव	उपसर्जन	उपपद
प्रगृह्य	सर्वनामस्थान	प्रातिपदिक	आमत्रित
निपात	संप्रसारण	नदी	सत्
संबुद्धि	ह्रस्व	घि	ਲ
प्रथमा	दीर्घ	लघु	(अकर्मक)
द्वितीया	प्रुत	गुरु	तद्राजाः
तृतीया	विधि	अङ्ग	अभ्यास
चतुर्थी	पद	ਮ	अभ्यस्त
पञ्चमी	लुक्	(कारक)	आम्रेडित
षष्ठी	ર્સ	अपादान	
सप्तमी	लुप्	संप्रदान	
घु	प्रत्यय	कर्म	

2. Interpretation rules (परिभाषा) (about priority etc.)

3. Rule hierarchy and inheritance (partial/full)

4. Domain of applicability/jurisdiction of sectional rules

\_\_\_\_\_

5. Chaining/nesting of rule applicability

6. General and special (or exception) rules etc.

Ending	Masc.	Fem.	Neu.	M&F	M&N	F&N	M,F,N	Total
अ	19	-	15					34
आ	2	12	-	-	-	1		15
ন্থ	10	5	11	-	-	1	1	28
क्र	10	7	-	2	-	-	-	19
उ	4	2	4	-	-	-	-	10
জ	4	6	-	-	-	-	-	10
ऋ	3	2	1	-	-	-	-	6
ॠ	1	-	-	-	-	-	-	1
ल	1	-	-	-	-	-	-	1
ए	1	-	-	-	-	-	-	1
ऐ	-	-	-	1	-	-		1
ओ	-	-	-	1	-	-		1
औ	-	-	-	1	-	-		1
Total	55	34	31	5	-	2	1	128
Ending	Masc.	Fem.	Neu	M&F	M&N	F&N	M,F,N	Total
क्	-	-	-	-	-	-	-	-
শ্ব্	-	-	-	-	-	-	-	-
ख् ग्	-	-	-	-	-	-	-	-
ख् ग् घ्	- -	- -	- -	- -	- - -	-	-  -	- - -
ख् ग् घ् ङः	- - -	- - -	- - -	- - -	- - -	- - -	-  -	- - -
ख ग घ छ च	- - - 12	- - - 1	- - - 10	- - - -	- - - -	- - - -	-  - -	- - - 23
र्ख ग् घ् छः च्	- - - 12 -	- - - 1 -	- - - 10 -	- - - -	- - - - -	- - - - -	-  - - -	- - - 23 -
र्ख ग घ छ च छ ज	- - - 12 - 7	- - - 1 - 1	- - - 10 - 4	- - - - - 1	- - - - - -	- - - - -	-  - - - -	- - - 23 - 13
र्ख ग घ छ च छ ज झ	- - - 12 - 7 -	- - - 1 - 1 -	- - - 10 - 4 -	- - - - - 1 -	- - - - - -	- - - - - -	-  - - - - -	- - - 23 - 13 -
জ্ ग घ ङ च छ ज झ ञ	- - - 12 - 7 - -	- - - 1 - 1 - 1 -	- - - 10 - 4 -	- - - - - 1 -	- - - - - - -	- - - - - - -	-  - - - - - -	- - - 23 - 13 - -
জ্ ग घ ङ च छ ज झ ज ट	- - - 12 - 7 - 7 - -	- - - 1 - 1 - - -	- - - 10 - 4 - - -	- - - - - 1 - - - -	- - - - - - - -	- - - - - - - - -	-  - - - - - - - -	- - - 23 - 13 - - - -
জ্	- - - 12 - 7 - - - -	- - - 1 - 1 - - - -	- - - 10 - 4 - - - -	- - - - - 1 - - - - -	- - - - - - - - -	- - - - - - - - - - -	-  - - - - - - - - - -	- - - 23 - 13 - - - - -
ৰ্	- - - 12 - 7 - - - - - -	- - - 1 - 1 - 1 - - - - - -	- - - 10 - 4 - - - - - -	- - - - - 1 - - - - - - -	- - - - - - - - - - - -	- - - - - - - - - - - - - -	-  - - - - - - - - - - - - -	- - - 23 - 13 - - - - - -
জ্ ग घ छ च छ ज द ठ छ द छ छ छ छ छ छ छ छ छ छ छ छ छ	- - - 12 - 7 - - - - - - - - -	- - - 1 - 1 - - - - - - - - - -	- - - 10 - 4 - - - - - - - -	- - - - - 1 - - - - - - - - -	- - - - - - - - - - - -	- - - - - - - - - - - - - - - - -	-  - - - - - - - - - - - - - - - -	- - - 23 - 13 - - - - - - - -
জ্ ग घ छ च छ ज इ च छ छ च छ छ च छ छ च छ छ च छ छ च छ छ छ छ छ छ छ छ छ छ छ छ छ	- - - 12 - 7 - - - - - - - 1	- - - 1 - 1 - 1 - - - - - - - - - - - -	- - - 10 - 4 - - - - - - - - - -	- - - - - 1 - - - - - - - - - - -	- - - - - - - - - - - - - - - - - - -	- - - - - - - - - - - - - - - - - - -	-  - - - - - - - - - - - - - - - -	- - 23 - 13 - - - - - - 1

# ANALYSIS OF SUBANTA TYPES

## **<u>PIVS Silver Jubilee Souvenir</u>**

थ्	1	-	-	-	-	-	-	1
द्	6	5	6	-	-	-	8	25
ধ্	1	1	-	-	-	-	-	2
न्	19	1	11	-	-	-	2	33
प्	1	1	1	-	-	-	-	3
फ्	-	-	-	-	-	-	-	-
ब्	-	-	-	-	-	-	-	-
મ્	-	1	-	-		-	-	1
म्	4	3	4	-	-	-	-	11
य्	-	-	-	-	-	-	-	-
र्	2	4	2	-	-	-	-	8
ल्	1	-	-	-	-	-	-	1
व्	-	-	1	1	-	-	-	2
য্	4	1	1	-	-	-	-	6
ष्	8	2	3	-	-	-	1	14
स्	16	5	8	-	-	-	1	30
ह्	8	3	2	-	-	-	-	13
Total	150	64	90	7	-	2	13	326

# ~@&&&&&

**<u>PIVS Silver Jubilee Souvenir</u>** 





रामानूजमहोदय : 'सि-डाक्' संस्थाया : गणकयन्त्र-सम्बन्धिनः संस्कृतकार्यस्य प्रमुखः । एलेक्ट्रानिक्स् इञ्जनियरिङ्ग् पदवीं कृष्णयजूर्वेदस्य प्राप्तवान् एषः तैत्तिरीयशाखायाः अष्ययनं कृतवान् अस्ति अपि । विशिष्टाद्वैतवेदान्ते अपि एतस्य विशेषपरिश्रमः अस्ति । बेङ्गलूरुनगरस्थया वेदधर्मपरिपालन-सभया 'वेदवारिधि:' इति बिरुदया सम्मानितः अपि अस्ति एषः । विविधासु संस्कृत-गणकयन्त्र-सम्बद्धासू गोष्ठीषू एतेन बहवः प्रबन्धाः उपस्थापिताः । बहूनां संशोधकच्छात्राणां मार्गदर्शनम् अपि एतेन कृतम् ।

समग्रा अष्टाध्यायी गणकयन्त्रे आरोपणीया इति रामानूजवर्यस्य प्रयतः । 'देशिक' नामकस्य प्रक्रमस्य सज्जीकरणे एतस्य महत्त्वभूतं पात्रम् अस्ति । 'अष्टाध्यायी गणकयन्त्रे केन क्रमेण आरोपयितूं शक्या' इति कश्चन मार्ग: एतेन निर्मित: इति वदाम: चेत् न अतिशयोक्तिः तत्र ।

तिरुमल-तिरुपति - देवस्थानस्य सहयोगेन सि-डाकसंस्थायाः कार्यं तिरुपतिनगरे विस्तार्यते । गणकयन्त्रे चतुर्णाम् अपि वेदानाम् आरोपणार्थं कृता महती योजना तत्र कार्यपथम् आनीयते ।

'यत् क्रियते तत् समीचीनं इत्यत्र रामानूजवर्यस्य स्यात्' विशेषनैर्भर्यम् । अतः एव सारभूतं किश्चित् कार्यं कर्तुं शक्तवान् सः , अग्रे शक्ष्यति अपि ।



मः कृष्णः च दशमकक्ष्यायाः छात्रौ । उभयोः अपि इच्छा - वार्षिकपरीक्षायाम् अधिकाः 🛛 अद्भाः प्राप्तव्याः इति । अतः रामः प्रतिदिनं श्रद्धया पठति । गुरून् पृष्ट्वा सन्देहानां निवारणं प्राप्नोति । ग्रन्थालयं गत्वा विषयसम्बद्धान् ग्रन्थान् परिशीलयति अपि ।

किन्तु कृष्णः न तथा । यद्यपि सः बुद्धिमान्, किन्तु सः पठने अनासक्तः । क्रीडादिभिः एव तस्य दिनानि यापितानि भवन्ति । रामेण सज्जीकृतां टिप्पणीं गृहम् आनीय स्वपूस्तके कदाचित् लिखति सः ।

परीक्षा आगता । राम: ज्ञानस्य आधारेण उत्तराणि लिखितवान् । किन्तू कृष्ण: अन्यानि 'तन्त्राणि' अश्रित्य उत्तराणि लिखितवान् । परीक्षापरिणामः अपि आगतः । रामेण ६९% अड्झाः प्राप्ताः । कृष्णेन तावत् ७०% अङ्गाः ।

परीक्षापरिणाममात्रं पश्यामः चेत् कृष्णः एव उत्तमः इति निर्णेतव्यं भवति । किन्तु अध्ययनमार्गः, भाविपरिणामः इत्यादीन् परिशीलयामः चेत् भासते यत् रामः एव उत्तमः इति ।

संस्कृत-गणकयन्त्रक्षेत्रे इदानीं बहुविधाः प्रयत्नाः दृश्यन्ते । किन्तु तेषु केचन रामप्रयत्नसदशा:, पुन: केचन कृष्णप्रयत्नसदशा: । केचन प्रामाणिकं प्रयतं कूर्वन्त: दश्यन्ते । पून: केचन अल्पेन प्रयासेन प्रसिद्धि प्राप्तुम् उद्युक्ताः दृश्यन्ते |

कथम् अल्पप्रयत्नेन प्रसिद्धिः प्राप्यते ? अत्र एकं उदाहरणं पश्याम तावत् ।

गणकयन्त्रद्वारा धातुरूपाणि प्राप्तव्यानि सन्ति इति भावयाम । तदा द्विधा मार्गः -१. धातुरूपप्रकाशिकाम्, अन्यं धातुकोषं वा दृष्ट्रा गणकयन्त्रे सर्वाणि रूपाणि आरोप्य तद्द्वारा धातुरूपाणां प्राप्तिः । २. गणकयन्त्रे धातुरूपसिद्धिसम्बद्धानां पाणिनिसूत्राणाम् (संज्ञा-परिभाषादिसूत्राणाम् अपि) आरोपणं कृत्वा यन्त्रम् एव रूपाणि यथा साधयेत् तथा करणम् ।

# 'सि-डाक्' संस्थया संस्कृत-गणकयन्त्रक्षेत्रे क्रियमाणस्य प्रयत्रस्य परिचयः ।

प्रथममार्गस्य निमित्तं विशेषपरिश्रमः करणीयः नास्ति । बौद्धिककार्यम् अपि न एतत् । दत्तांशानाम् आधारेण निर्मातुं शक्य: प्रक्रम: एष: । (Data Based Programme)

द्वितीयमार्गस्य निमित्तं तावत् महान् परिश्रमः करणीयः भवति । पाणिनीयसूत्राणाम् अर्थः यन्त्रे सम्यक् आरोपणीय: भवति । परिभाषासूत्रादीनाम् अर्थम् अवगत्य यन्त्रं कार्यं यथा कुर्यात् तथा प्रयतः करणीयः भवति । अत्र बुद्धिकार्यम् अधिकम् ।

द्वितीयविधम् एकम् उत्कृष्टं कार्यं 'सि-डाक्' नामिकायाः संस्थायाः आश्रये बेङ्गलूरुनगरे प्रचलति । प्रगत- सङ्गणक-विकास-केन्द्रम् (Centre for Development of Advanced Computing) एव सङ्ग्रहेण सि-डाक् (C-DAC) इति निर्दिध्यते । एषा संस्था तावत् गणकयन्त्रसम्बद्धानि विविधानि कार्याणि करोति । तेषु संस्कृतसम्बद्धं कार्यम् अपि अन्यतमम् ।

'सि-डाक्' संस्थया यन्त्रद्वारा भाषा-विश्लेषणस्य साधनार्थं 'जिस्ट्' (Gist) नामकं दृढपत्ररूपं यन्त्राङ्गं निर्मितम् अस्ति । एतस्य उपयोगत: भारतीयलिपीनां विदेशीयलिपीनां च प्रयोग: गणकयन्त्रे कर्तुं शक्य: । वैदिकवाब्यूयस्य सस्वरं लेखनम् अपि एतदद्वारा शक्यते ।

'समग्रा 'अष्टाध्यायी' गणकयन्त्रे आरोपणीया' इति सि-डाक्संस्थायाः सङ्कल्पः । तदनुगुणं प्रकृतं कार्यं प्रचलत् अस्ति । तस्य पूर्वाङ्गरूपेण 'देशिक'नामकः एकः परिष्कृतः प्रक्रमः (Programme) रचितः अस्ति ।

देशिके चत्वार: विभागा: सन्ति | ते च -

१. निष्पत्तिः २. विश्लेषणम् ३. सन्धिः ४. वेदः चेति ।

१. निष्पत्तिः - निष्पत्तिविभागे तावत् पुनः चतुर्धा विभागः । i) सुप् ii) तिङ् iii) कृत् iv) अव्ययं चेति ।

i) सुप्-विभागः - अत्र यस्यकस्यापि शब्दस्य सर्वाणि विभक्तिरूपाणि प्राप्तुं शक्यानि । तदर्थम् अत्र एषः क्रमः आश्रितः - प्रथमं तावत् प्रातिपदिककोषः रचितः । तत्र ५०,००० अधिकानि प्रातिपदिकानि सन्ति । कतिविधानि प्रातिपदिकानां तत• अन्ताक्षराणि भवितुम् अर्हन्ति इति परिगणनं कृतम् । परिशीलनानन्तरं निर्णीतं यत् ३२ अन्ताक्षराणि भवन्ति इति । लिङ्गादिभेदेन, सर्वनामादिभेदेन, तृतीयादिषू यत् णत्वादिकं भवति तद्भेदेन च ३२७ शब्दस्वरूपभेदाः निर्णीताः । (तेषु १५० विशेषशब्दाः, अन्ये सामान्याः ।) ततः लिङ्गानुशासनम् अपि अत्र योजितम् ।

एतस्य सर्वस्य बलात् अत्र यस्यकस्यापि शब्दस्य रूपाणि प्राप्तुं सौलभ्यं सिद्धम् । यदि प्रातिपदिककोषे अविद्यमानः शब्दः पृष्टः तर्हि यन्त्रं प्रथमं तस्य लिङ्गं पृच्छति । लिङ्गे निर्दिष्टे,



#### (चित्रम् - १)

तदन्ताक्षरयुक्तनां शब्दानाम् आवलीं प्रदर्श्य -'एतेषु केन सदश: अस्ति स: शब्द: ?' इति पुन: पृच्छति । तत: रूपाणि निर्दुष्टं निर्दिशति।

एतस्य परिष्कृतत्वं निरूपयितुं श्री रामानुजवर्यः (सि-डाक्संस्थायाः संस्कृतकार्यस्य प्रमुखः) एतां घटनाम् उदाहरति - ''कदाचित् एवं जातम् -मुम्बयीपत्तने गणपतिवाक्यार्थसदसि श्टङ्गेरीश्रीचरणाः एतां योजनां परीक्षमाणाः पृष्टवन्तः - ''सदस्-शब्दस्य स्त्रीलिङ्गे कानि रूपाणि ?'' इति । मम मनसि भीतिः आसीत् - यन्त्रं तानि रूपाणि प्रदर्शयितुं शकुयात् वा न वा इति । तथापि यन्त्रे आदेशाः पूरिताः । आश्चर्यं नाम यन्त्रेण - 'सदाः, सदसौ, सदसः ...' इत्येवं रूपाणि प्रदर्शितानि एव !'' इति ।

ii) तिङ्विभागः - अत्र पाणिनीयधातुपाठे स्थितानां सर्वेषाम् अपि धातूनां कर्तारे दशसु लकारेषु ('लेट्'लकारे चापि, यश्च वेदेषु प्रयुज्यते ।) अपि ददाति । न केवलं तावत्, णिच् - सन् - यङादिरूपाणि अपि ददाति यन्त्रम् । यङादीनां रूपाणां योजनप्रक्रियायाम् इतोऽपि कश्चन परिष्कारः करणीयः अस्ति । स च क्रियमाणः अस्ति एव ।

तिङ्रूपाणि

एतानि सर्वाणि '

(धातुरूपाणि) पाणिनिसूत्राणाम् अर्थम् अवगत्य, सूत्राणां बलाबलं निर्णीय निर्दिशति यन्त्रम् । अतः एतत् दत्तांशाधारेण (Date based) निर्मितं न, अपि तु पाणिनिसूत्रावगतेः आधारेण निर्मितम् । सामान्यतः ७०० - ८०० सूत्राणि उपयुक्तानि अत्र।

सूत्राणाम् आधारेण एषः विभागः सजीकृतः इत्यतः रूपसिद्धिसमये या प्रक्रिया आश्रिता, यानि सूत्राणि उपयुक्तानि, यः विकारः तत्तत्पदे जायते, तत्सर्वं द्रष्टुं शक्यम् अत्र।

iii) कृत्-विभागः - पाणिनिना निर्दिष्टानां कृत्प्रत्ययानां योगेन यानि रूपाणि भवन्ति तानि अत्र द्रष्टुं शक्यानि | प्रकृतं प्रसिद्धाः दश कृत्प्रत्ययाः (सर्वेषां धातूनाम्) योजिताः | अन्येषां योजने प्रयत्नः प्रचलति |

'गृ' (निगरणे) धातोः (गिलति - इति वर्तमाने प्रथमपुरुषैकवचने) क्तवतुप्रत्यये कृते 'गिलितवान्' इति रूपम् इति वयं प्रमादवशात् वदेम । किन्तु गणकयन्त्रं तावत् सूत्रादिकं परिशील्य 'गीर्णवान्' इत्येव रूपं प्रदर्शयति । एवम् एव इच्छितव्यम्, खनितवान्, चयितवान् इत्यादीनि असाधुरूपाणि वयं वदेम प्रमादवशात्, किन्तु यन्त्रं तावत् एष्टव्यम्, खातवान्, चितवान्

### विज्ञानम्

इत्यादीनि शुद्धानि रूपाणि एव प्रदर्शयति ।

iv) अव्ययविभागः - अत्र प्रसिद्धानि सामान्यानि अव्ययानि, कृदन्ताव्ययानि, तद्धितान्ताव्ययानि इत्यादीनि सङ्ग्रहीतानि ।

२. विश्लेषणम् - विश्लेषणं तावत् पदस्तरे वाक्यस्तरे चेति द्विविधम्।

i) पदस्तरीयम् - अत्र यः कोऽपि शब्दः दत्तः चेत् तस्य विभक्तिः, वचनम्, अन्तम् इत्यादिकं निर्दिशति यन्त्रम् । अन्तविभक्त्यादिभेदतः बह्वचः सम्भावनाः सन्ति चेत् ताः अपि प्रदर्शयति यन्त्रम् ।

ii) वाक्यस्तरीयम् - वाक्यस्तरे विश्लेषणं करणीयं चेत् कारकसम्बन्धः नितराम्

आवश्यकः । ततः योग्यतादि -निर्णयार्थम् उपायाः अपि आवश्यकाः।

इदानीम् एतद्विभागे तावत् कारकप्रकरणं समग्रतया योजितम् । (सूत्रपुरस्सरम्) कारक-विभक्तिसम्बन्धः अपि स्फुटीकृतः अस्ति । यद्यपि ६ कारकाणि शास्त्रे प्रसिद्धानि, तथापि अत्र यन्त्रसौलभ्यं मनसि निधाय षण्णाम् एव कारकाणां ६२ अवान्तरभेदाः कल्पिताः । (चित्रम् - १ पश्यन्तु ।)

अतः एव 'लक्ष्मीः कमले अघितिष्ठति' इति वाक्यं दत्तं चेत् एतत् वाक्यम् असङ्गतम् इति सूचयति यन्त्रम् । 'लक्ष्मीः कमलम् अघितिष्ठति' इति उक्तं चेत् पदविश्लेषणं करोति !!

विश्लेषणानन्तरं शाब्दबोधः अपि अत्र द्रष्टुं शक्यः । वैयाकरणानां, नैयायिकानां, मीमांसकानां च मतम् अवलम्ब्य त्रिविधः अपि शाब्दबोधः यन्त्रेण निरूप्यते ।

वाक्यविश्लेषणार्थं 'शब्दशक्तिप्रकाशिका', 'मञ्जूषा', 'भाट्टरहस्यम्' इत्येतान् त्रीन् ग्रन्थान् परिशील्य तत्र उदाहृतानि वाक्यानि सङ्गृहीतानि । एवं सङ्गृहीतानि ११६७ वाक्यानि प्रातिनिधिकानि इति परिगणय्य वाक्यविश्लेषणप्रयतनः अनुवर्तितः । प्रकृतं केषाञ्चित् एव वाक्यानां विश्लेषणम् अत्र शक्यते ।

वस्तुतः वाक्य-विश्लेषणे प्रधानः क्लेशः तावत् योग्यतादिनिर्णये । अतः तदर्थम् अपि अत्र प्रयासः कृतः । 'शिलाखण्डः विद्यालयं गच्छति' इति वाक्यं दत्तं चेत् वाक्यस्थानां शब्दानां विश्लेषणं तावत् यन्त्रं करोति, किन्तु शाब्दबोधं तु न प्रदर्शयति । कुत्र सम्बन्धः विच्छिन्नः इति निर्दिशति तत्।

३. सन्धिः - सन्धिविभागे विच्छेदः संयोगश्चेति द्वैविध्यम् । संयोगः तावत् अल्पश्रमसाध्यः । अतः संयोगः अत्र पूर्णतः निरूपितः । पदद्वयं दत्तं चेत् सन्धिं कृत्वा प्रदर्शयति यन्त्रम् । कृत्प्रकरणे तिङ्प्रकरणे च सन्धिः अपरिहार्यः । तत्र सर्वत्र सूत्रपुरस्सरं



गणकयन्त्रकार्यरतः श्री रामानुजमहोदयः ।

सन्धि करोति एव यन्त्रम्।

उदाहरणार्थं - 'प्रक्षालयति' इत्यस्य ('प्र' इति उपसर्गयुक्तस्य 'क्षल'धातोः) भूतकालरूपं प्राक्षालयत् इत्येव प्रदर्शयति यन्त्रं, न तु अप्रक्षालयत् इति । अडागमविषये सन्धौ वा न प्रमाद्यति तत् ।

सन्धिविच्छेदे तावत् समस्याः बह्वथः । 'रामो वनं गच्छति' इत्यत्र 'रामः अवनं गच्छति' इत्यपि सन्धिविच्छेदः भवितुम् अर्हति । 'नमेयम् उच्चैः इतरा गतिर्न मे' (विश्वगुणादर्शचम्पूः) इत्यत्र 'नमेयम्' इत्येकस्य पदस्य १७ धा सन्धिविभागं कृत्वा प्रदर्शयति यन्त्रम् । (चित्रम् - २ पश्यन्तु ।) प्रतिविच्छेदं साधुत्वं कथम् इत्यपि निर्दिशति तत् । तत्र कतमं रूपं स्वीकरणीयम् इत्यत्र विनिगमकं किं भवितुम् अर्हति इति विचिन्त्य तच ज्ञानं यन्त्रे आरोपणीयम् । अतः सन्धिविच्छेदक्षेत्रे न बहु कृतम् अत्र ।

४. वेदः - अत्र त्रिधा विभागः अस्ति १. सन्धिः, २. विकृतयः, ३. वर्णक्रमः
चेति ।

वेदे स्वरस्य प्राधान्यं भवति । अर्थनिश्चये सन्धिनिश्चये वा स्वर: एव विनिगमक: भवति । अत: वेदक्षेत्रे यन्त्रं शब्दसंयोगं यथा करोति तथैव सन्धिविच्छेदम् अपि लीलया करोति । सन्धि कृत्वा स्वरविपरिणामादिकम् अपि करोति । 'तैत्तिरीय-प्रातिशाख्य-नियमा:', 'व्यासशिक्षा' च अत्र आरोपिता: इत्यत: एतत् सौलभ्यं सिद्धम् अस्ति ।

कमः, जटा इत्यादयः पाठाः वेदविकृतयः इत्युच्चन्ते । प्रकृतं तावत् तैत्तिरीयशाखायाः यस्यकस्यचित् मन्त्रस्य पदपाठः दत्तः चेत् तस्य कम-जटा-घनादिकं प्रदर्शयति यन्त्रम् ।

तृतीयः विभागः वर्णक्रम-सम्बद्धः । शास्त्रेषु वर्णानां धर्माः निर्दिश्यन्ते । एकस्य 'हल्' (व्यञ्जन) वर्णस्य २६ धर्माः सम्भवन्ति । (अष्टौ स्वरसम्बद्धाः, अष्टौ व्यञ्जनसम्बद्धाः, १० उदात्तादि- स्वरसम्बद्धाः चेति ।) एतत्सर्वं विस्तरेण निरूपयति यन्त्रम् ।

'देशिक:' इत्यस्य स्थूलं विवरणम् एतत् । 'देशिक:' इदानीं विक्रयणार्थम् अपि सिद्ध: अस्ति । 'सि-डाक्' संस्थात: य: कोऽपि एतं केतुम् अर्हति । देशिकस्य 'जिस्ट्काई' इत्यस्य च सम्बन्ध: अस्ति । अत: जोस्ट् काई, देशिक -इत्येतदुभयं क्रीतं चेत् गणकयन्त्रकार्य स्वयं कर्तुं विशेषसौलभ्यं सिद्धयति । (सङ्केत: तावत् - C-DAC, Ramanashree Plaza, 2/1, Brunton Road, Bangalore - 85)

'देशिक'निमित्तं पाणिनीयायाः अष्टाध्याय्याः कानिचन सूत्राणि उपयुक्तानि | समग्रा अष्टाध्यायी अपि यन्त्रे आरोपणीया इत्यत्र प्रयत्नः प्रचलति इदानीम् | अष्टाध्याय्याः यन्त्रे आरोपणं नाम 'तत्र स्थितानां ४,००० सूत्राणां सूत्रवृत्तीनां च आरोपणम्' इति अर्थः न । एतत् कार्यं तावत् सामान्य-संस्कृतज्ञानेन अपि कर्तुं शक्यम् । अर्थानुगुणं यन्त्रं कार्यं यथा कूर्यात् तथा करणीयं चेत् महान् परिश्रमः करणीयः भवति ।

सूत्रमात्राणि तु अत्रापि आरोपितानि सन्ति एव । सूत्रसङ्ख्या दत्ता चेत् सूत्रस्य पूर्णं विवरणं

(त्रिपाद्याम्) अस्ति । असिद्धकाण्डविषये अस्माभिः एतावता अपि किमपि न कृतम् । अतः यन्त्रम् इतोऽपि 'प्रभवन्ति' इत्यत्र यत् णत्वनिषेधसूत्रं प्रवर्तते तत् सूत्रं न जानाति । समग्रा अष्टाघ्यायी यदा योज्यते तदा एषः दोषः परिमार्ज्यते । 'प्रभवन्ति' इति रूपम् एव द्रष्टारः यथा पश्येयुः तथा कर्तुम् इदानीम् अपि

गणकयन्त्रविज्ञानिनाम् अभिप्रायः । यदि संस्कृतगणकयन्त्रक्षेत्रे सारभूतानि कार्याणि बहूनि प्रचलेयुः तर्हि ताद्यां दिनं प्रायः दूरे न स्यात् यस्मिन् दिने गणकयन्त्रविशेषज्ञाः अपि संस्कृतं सश्रद्धं नमेयु: ।

सादरं स्वीकृतानि

विज्ञानम्



अस्ति अत्र ।

सि-डाक्संस्थायाः एवं कार्यस्य परिणामतः भाषायाः केषुचित् अंशेषु परिष्कृतः कश्चन मार्गः स्पष्टीकृतः अस्ति । तथापि फलप्राप्तिसमये अल्पाः केचन त्रुटयः दश्यन्ते । यथा - 'प्र' इति उपसर्गेण युक्तस्य भूघातोः प्रथमपुरुषबहुवचनरूपं 'प्रभवण्ति' इति निर्दिशति यन्त्रम् । णत्वविषये अत्र प्रमाद: I

एतद्विषये श्रीरामानूजमहोदयः वदति -'णत्वप्रकरणं तावत असिद्धकाण्डे

अष्टाध्यायीकार्यादिकं तु तत्रापि अनुवर्तते एव ।

'कम्प्यूटर्' इति शब्दः कम्प्युटेषन्-(Computation) शब्दात् निष्पन्न: | अत: 'कम्पूटर्' इति शब्द: एव सूचयति यत् एतत् यन्त्रं गणनाकार्यादिनिमित्तम् एव सृष्टम् इति । गणकयन्त्रद्वारा यद्यपि इदानीं बहुविधकार्याणि कार्यन्ते, किन्तु तेषां सर्वेषां मूलतत्त्वं तावत् गणनासम्बद्धम् एव । भाषासम्बद्धा प्रणाली काऽपि नास्ति तत्र । 'तादशी प्रणाली करणीया चेत् संस्कृताश्रयणं विना नान्या गति:' इति

* नारायण	गुरोः संस्कृतकृतयः			
म -	नारायणगुरुकुलम्			
	श्रीनिवासपुरम्,			
	वर्कला-६९ँ५ १४५ केरलराज्य			
पृष्ठानि -	५२, प्र.मु १९९३			
मूल्यं -	न निर्दिष्टम्			
* संवित्				
(Samskrit	& Computer based			
linguistic	s)			
सम्पादकौ -	एम्.ए.एस्. राजन् &			
	एस्.एच्. श्रीनिवासन्			
म -	संस्कृतसंशोधनसंसत्			
	मेलुकोटे - ५७१ ४३१,			
	कर्णाटकम्			
पृष्ठानि -	१३८, प्र. मु १९९३			
मूल्यं -	रू. ३०.००			
* संस्कृतशि	ाक्षणे प्रायोगिकाध्यापन-			
	विधानम्			
ले -	डा. कानाल नलचक्रवर्ती			
Я −	श्री वेङ्कटाम्बा पब्लिकेषन्स्,			
	तिरुपति (आ.प्र.)			
नृ -	१४९, प्र. मु १९९४,			
मूल्यं -	रू. ४५			
* संस्कृतसङ्गमः (स्मरणसञ्चिका)				
म -	भारतसंस्कृतपरिषत्, ५२८,			
	नारायण पेंट्, विश्वशोभा को.			
	आप्. हौ. सोसैटी,			
	पुणे - ४११ ०३०			
पृष्ठानि -	२८, प्र. मु १९९४,			
मूल्यं -	रू. ५.००			

सादरस्वीकारार्थं ये स्वकृति प्रेषयितुम् इच्छन्ति तैः प्रतिकृतिद्वयं प्रेषणीयम् । एवं प्राप्तानां कृतीनां नाम-प्रकाशकादिविवरणम् एतस्मिन् विभागे यथावकाशे प्रकाशयिष्यामः ।

- -



# Software for Sanskrit analysis developed

By our staff reporter At first glance, one may tend to mistake him for the friendly local Brahmin who performs vedic rituals. But few may realise that Prof. P. Ramanujan, a research scientist with the Centre for Development of Advanced Com-puting (C-DAC) has developed a software package for the systemat-ic analysis of Sanskrit grammar and its fundamentals. and its fundamentals.

Called 'Desika', the software package has been developed under the Knowledge-Based Computer Systems (KCBS) and funded by the Department of Electron-ics(D oE) and United Nations Development Programme (UND-P). It claims to offer a computational rendering of Panini's gram-mar, an advantage for linguistic sludies and research. Prof. Ramanujan, a participant

Prof. Ramanujan, a participant of the three day All-India Oriental Conference organised by the Bhandarkar Oriental Research Institute (BORI), who has been on the project since September 1990 pointed out that the package, which is a substrate for develop-ment of similar packages in Indian

languages, offers immense possi-bilities of standardisation of the script on the normal English keyboard.

Prof. Ramanujan ,essentially an aeronautical engineer by profes-sion, is also a scholar of the vedic sciences. He realised the potential of utilising Sanskrit grammar for the analysis and preservation of our ancient texts. Since the Sastras are comprehen-Since the Sastras are comprehen-sive and exhaustive in nature, they have been transmitted over generations through the oral tradi-tion. By representing their multi-dimensional character, we can help define a knowledge represen-tation system based on the ancient Indian tradition and, comparitive-by study the madern development Indian tradition and, comparitive-ly, study the modern developments in this light, he pointed out. The West also developed an interest in utilising these systems for the preservation of literary traditions, thus making it imperative for us to take the lead and make use of our rich cultural heritage to our advan-tage herevolution.

tage he explained. The intention was to apply com-puters for the various branches of study, yet work in effective coor-

dination. In a similar vein, recently the C-DAC had developed the GIST (Graphics and Intelligence based Script Technology) card which provides facility for prolinguistic data in any cessing



Research scientist S.Ramanujan at work many techniques are available, he added.

Indian scripts and also a few foreign In or scripts In order to utilise knowledge contained in natural language in free form within a machine, it is ential to devise gertain formats,

and the knowledge should be represented suitably in a machine understandable form, Prof. Rama-nujan explained. To achieve the conversion in free form in natural. language to formal machine form,

The C-DAC was chosen by the DoE and UNDP for the proper

understanding of the ancient Indi-an scripts and artificial intelli-

Seven nodal centres gence issues. were set up all over the country to specialise in the areas of frontier

specialise in the areas of frontier research, he pointed out. So far, the response to the pack-age has been promising, he said. Henceforth the same yardstick could be applied to different lan-guages for easier standardisation and a uniform level of syllabi. The package is amenable to

The package is amenable to foreigners and non-resident Indito ans wanting to establish links with our culture, he observed. Prof. Ramanujan, who has been recently been transferred to the C-DAC branch at, Bangalore, stated that the ardous process of grammar modelling would be completed within three years.

Life for Prof. Ramanujan has been hectic for the past six months. Dashing across the coun-try providing demonstrations of the package, he even has an ap-pointment with Prime minister Narasimha Rao for a demonstra-tion during the first week of June. A beginning has been made in this important area. A lot more remains to be done, he said.

# संस्कृत और कम्प्यूटर से राष्ट्रीय एकता की कोशिश

नगर संवादताता सागर तेर शाखों में दर्ज अमुल्य ज्ञान को आधुनिक तकनीक की सहापता से जनसुलभ का दिया जाए तो समुचा भारत अपनी संस्कृति के माध्यम् से डी एकस्नीकविद्य और संस्कृत के ज्ञानी इस रवननकी संस्कृत के ज्ञानी इस रवन की संकार करने का प्रयास कर रहे हैं। सूचना एवं प्रसारण मंत्रालय के बंगलुर स्थित सेंटर ऑफ डेवलपमेंट फॉर





तिलक, खियडी वाडी, सिर पर लंबी बोटी, देरे मे बयाल, प्रन्टदिवर संस्कृत, कुएं का जनी पीना, खुद का बनाया भौजन ही मिरायों का पालन करते हैं तो दूसरी निर्ययां का पालन करते हैं तो दूसरी तरफ मौबाहल, लेपटॉप से लेस प्रतेष्ठ वें होकर और सॉफ्टेक्यर इंजीनियर है। पंचा पर रह ह 1969 में तत्कालीम राष्ट्रपति ठों. शकरदयाल शर्मा की मौजुदगी में पहली बार कम्प्युटर पर संस्कृत का प्रयोग करके दिखाया और तब से इसी पर कार्य कर रहे हैं। अब तक वे संस्कृत के 300 गुंथों का हा अब तक वे संस्कृत के 300 ग्रंथों का कम्प्यूटर पर डेटा बेस आत युक्ते हैं। इंटाबंस पूरी तरह तेरार होने के वाद तिया बदतेगी, भाषा वही रहेगी। इससे संस्कृत के प्रंथ और भारत का ज्ञान भी सुरक्षित रहेगा।

#### वाणी पर आधारित होगा कम्प्यटर

पाणा भर जापारारा लागा जन्मपुरः विदेशी कहते है कि जो लोग (खासकर अंग्रेजी) लिख-पढ़ नहीं सकते ये अनयह है जबकि पी. रामाजुनन का मानना है कि जो बोल सकता है, वह भी साक्षर है । वे कहते है, हम न्यीच बेरह ! बोली पर आधारित) कम्प्यूटर का ईजाद कर लेंगे तथ यह संभव हो जाएगा। अद्भुत समन्वय

अदभुत राभानवा जातुर के प्रसिद्ध सन्मुक विवान परगलुषावार्थ के यहा जन्मे थी. रामानुजन देवी के जाता है। स्वरूत तरापिय राशिज से नादीय मेरिंट मे रहे और स्वारुद न ही इरविष्ठालक इजीवियाँग की 1530 मिलिस की देश की जानी-मानी हिंदुआन करियान सिंग्रेड के यारेवालन जिजावीन पर काम किया। 1983 से 1990 तक क्रिकेस सिंहत एंड डेरातमांट उजनीवालोगन (डोजारीओ) के प्रसार देश स्वार में उनके कि जिता इसी तराम द्वार्यात्रा कि प्रारंधन देश्वारीक डो, एडी.अ., आवुल राशमा के सकनीकी स्वीरक के रस्य में हरूरे तडाकू जिना पर काम कि 20 का दाय किया 3 जाने के सीडेक में कम्प्रदार और सन्द्रक प्रध पर कार्य कर रहे हैं।

#### सीडेक : एक नजर में

(1164) : (भूष गधार म अमेरिको ने उस पर को कप्प्युटर टेक्नोतोंजी देने से मना बर दिया था, तो तत्कातीन प्रधानमंत्री तजीव गांधी ने सेम विश्वेय को धारत में सुपर कप्प्युटर के निर्माण का जिम्मा तौया। 1969 में मैंसैके अस्तित्व में आया इस संस्थान ने 1991 में भरत के लिए धारता सुपर कप्प्युटर 'परम' गोव सिर्वा था। अब ने करके कई एडवास्ड वर्जन आ गए हैं। अब यह संस्थान भाषा सक्तीक के क्षेत्र में कार्य कर रहा है।



# कम्प्यूटर की गति और मति का उपयोग करें

2.1 दिनी राष्ट्रीय कार्यवाला का तीस्य दिन त्रार बेंडरवला कार रेव भाव संस्कृत भारत में जन्मे लेकिन वह विश्व दी भे तकार के उनकी रही 105 भाग हो बोकि वह विश्व दी भन्म रेव भाव संस्कृत भारत में जन्में लेकिन वह विश्व दी भन्म तेन प्रदेर वन तो है 1 यदि भारत सरका छा का परेट का तो हो मान सिरंग जो प्रकार को उनकी स्वीन वह विश्व दी भन्म सार्व कि उनकी रही 105 भाग हो बोकि वह किया दी भन्म मान सरका के उनकी दिवा भा का स्वार के बात स्वार के से तो है 1 विश्व भारत सरका छा का परेट का मान सरका के उनका दिवा भा का स्वार के प्रकार के प्रकार की का स्वार का मान सार की उनकी रही 105 भाग हो बोकि स्वार कि प्रकार की है 1 विश्व भारत सरका का स्वार का स्वार का स्वार मान सरका के उनका दिवान के स्वार विश्व में का स्वार का का का स्वार के स्वार का स्वार सार का सरका के उनका दिवान के सार किया थी भा सार का स्वार का स्वार का स्वार स्वार का सरका के उनका दिवान के सार की स्वार कि प्रकार के सार 1 जे तका की स्वार का स्वार का सार सरकार होना 1 जित का प्रकार का स्वार के से ता ती ही कि प्रकार का प्रकार का ना 1 जा सार्ववारी, जावका वास स्वार का दोर्सिक रोग होने के मुत्यों विष्ठ मां के दो ती

सी-डेक के निदेशक डॉ. रामानुजन् ने कहा-बैंगलोर में वैज्ञानिकों का दल तकनीक विकसित करने में जुटा



#### समीर पाटिल, उज्जैन

बेदों की ऋचाओं के उच्चारण से निकलने वाली ध्वनि अब कम्प्यूटर पर भी आसानी से उतारी जा सकेगी। इस तकनीक को विकसित करने के लिए सी-डेक (सेंटर फॉर डेवलपमेंट ऑफ एडवांस्ड कम्प्यटिंग) बैंगलोर में वैज्ञानिकों का दल कार्यरत है। ऐसा होने पर समस्त वैदिक साहित्य की स्वर आधारित सारणी (इंडेक्स) बन सकेगी। 'माउस' या 'की' के माध्यम से किसी एक स्वर को पाठ में ढंढना व उस स्वर को सनना भी

संभव हो सकेगा। यह प्राच्य व वैदिक विषयों के शोधकर्ताओं के लिए 🖩 डॉ. रामानुजन् विशेष मददगार साबित होगा। यह बात सी-डेक बैंगलोर के निदेशक डॉ. पी. रामानुजनु ने भास्कर से चर्चा में कही। वे विश्व वेद सम्मेलन में भाग लेने उज्जैन आए हैं। उन्होंने कहा यह कार्य होने पर सारा वैदिक साहित्य कम्प्यूटर पर उपलब्ध होगा। अंग्रेजी भाषा को कम्प्यूटर पर टेक्स्ट-ट-स्पीच यानी लिखित को ध्वनि में डालने के लिए 40 वर्ष से अधिक का समय लगा। संस्कृत सहित भारतीय भाषाओं में यह कार्य तीन वर्षों में ही हो सकता है क्योंकि ये भाषाएं ध्वन्यात्मक हैं। ध्वनि के आधार पर ही लिखे जाने की वजह से वे तार्किक रूप से अधिक शुद्ध हैं और उन पर कार्य करना सुगम है। इंडियन स्टेटेस्टिकल इंस्टिट्यूट एवं सी-डेक कोलकाता में बांग्ला भाषा के कम्प्यूटर स्वरांकन पर गहन शोध हुआ है। वेदों के स्वरांकन में इस शोध कार्य की मदद भी ली जाएगी। सी-डेक बैंगलोर में इंडिया हेरिटेज ग्रुप के अंतर्गत काम कर रहे डॉ. रामानजन एवं अन्य विद्वान इंडिया हेरिटेज पोर्टल बना चके हैं। वैदिक अध्ययन के लिए एक वैदिक रीडर पर भी काम हो रहा है जहां पर वेदों के संबंध में समग्र जानकारी मिल सकेगी।

#### पांडलिपियों का संरक्षण

कोई भी ग्रंथ कई पांडुलिपियों में होता हैं। ग्रंथ विशेष के क्रिटीकल एडीशन या आलोचनात्मक अध्ययन के लिए सभी प्राप्त प्रतियों का अध्ययन करना होता है। इसके लिए आईलिप सॉफ्टवेअर बनाया है। इसका निर्माण करने वाली टीम के सदस्य रहे डॉ. रामानुजन् ने बताया इस सॉफ्टवेयर की मदद से विभिन्न पाठों का एक साथ शोध सरल हो गया है। इससे एक पांडुलिपि का देश की प्रमुख भाषाओं की लिपियों में परिवर्तन भी संभव है।

### वेद में ध्वनि का महत्व

वेद में ध्वनि का खास महत्व है। यही एकमात्रं ऐसा साहित्य है, जिसमें स्वर को यथावत रखने के नियम है।

इन नियमों से ही उच्चारण एवं भाषा में सदिया गुजर जाने के बाद भी कोई विकृति या परिवर्तन नहीं आया है |

यही कारण है वेद पाठ करने के लिए कडा अभ्यास करना पडता है।

#### किसी ऋषि की जीवनशैली

डॉ. रामानुजन् कम्प्यूटर विज्ञान के देश के शीर्षस्थ विद्वानों में से हैं साथ ही वे वेदों सहित व्याकरण, मीमांसा, न्याय व वेदांत आदि शास्त्रों के निष्णात पंडित हैं। शिखा (चोटी), धोती-उत्तरीय (गमछा) धारण करने वाले डॉ. रामानुजन् का व्यक्तित्व कई लोगों को अचरज में डाल देता है। वे संस्कृत में वार्तालाप को प्राथमिकता देते हैं। वे स्वयंपाकी (खुद का भोजन तैयार करने वाले) हैं और स्वयं के हाथ से निकाले कुएं के पानी से ही स्नान व पूजा आदि करते हैं।



TOI Bombay 24-5-97

# **DIGITISING THE SNASTAS** Hiren K Bose met P Ramanujan who has computerised ancient Vedic texts for posterity

Should our ancient heritage, he knowledge contained in our *shastras*, be computerised? Traditional scholars, men for whom digital knowledge is an alien concept, would frown upon such an idea. Having derived their knowledge through oral tradition, reservations about using information technology to preserve our heritage is natural.

"Such Vedic scholars are few in number," says Ved Varidhi P Ramanujam, an Iyengar Brahmin,

Ramanujam, 59, an engineer at present with C-DAC at Bangalore, is qualified to make such a statement. Son of an Atharva



Veda scholar, in his initial years he studied in a *gurukul*, graduated in electrical engineering and is reading for an M.Sc. degree at the Indian Institute of Science at Bangalore.

Following his graduation, Ramanujam joined the Bangalorebased Hindustan Aeronautics Limited. He was among the handful of Indians to be trained on IBM computers.

In 1989 he presented a technical paper titled 'Computer representation of *Shabdha-bodha*during the first national conference on 'Computers and Sanskritam' which was attended by President Shanker Dayal Sharma. "It was the first time that flow charts were used to explain our ancient texts and I could see that the President understood what I was trying to convey," says Ramanujam about the experience.

This was followed by a paper titled 'A case for Sanskrit as computer programming language' at the annual convention of the Computer Society of India in Bangalore.

Impressed by his arguments in favour of computerisation of Vedic texts at a conference on Vedas and shastras, the Tirumala Tirupati Devasthanam awarded a grant of Rs 1 crore to set up a library with computers.

He has edited Mahakavi Venkatadhwari's Lakshmi Sahasram, written a series of articles on *Vedic sanskaras* in a Kannada quarterly and was awarded the Veda Varidhi title at 31 by the Veda Dharma Paripalana Sabha of Bangalore.

To date, he has completed 15 software projects guiding students from institutions like the University of Kerala, PICT, Pune, IIT Mumbai, MS University Baroda and others. At present, Ramanujam is working on Panini's gramniar.

Since 1990, he has been associated with 'Desika', a natural language understanding system for Sanskrit, based on the Vyakarana, Nyaya and Mimansa Shastras,' including Vedic processing. Besides, he has been rendering Vedic databases into a form that can be analysed by the computer.



412

# THE TIMES OF INDIA, BANGALORE . 19 APRIL 1997 . 8 PAGES SATURDAYTIMES ·

phould our ancient heritage, the knowledge contained in our *shastras*, be computerised? Traditional scholars, people who have gained an insight into the literature of our past, men for whom digital knowledge is an alien concept, would frown upon such an idea. Having derived their knowledge through oral tradition, it's but natural for these men to have reservations about using information technology to preserve our heritage.

"Such Vedic scholars are few in number," says Ved Varidhi P. Ramanujam, an Iyengar Brahmin, "We need to acquire from them the deep insight they have of the shastras. If we don't, we shall be losing a lot, for they will take their knowledge to the grave."

39 Ramanujam, 59, an engineer at present with C-DAC at Bangalore, is qualified to give such a statement. Son of an Atharva Veda scholar, in his initial years he studied in a gurukul, graduated in electrical engineering and is reading for an M.Sc. degree at the Indian Institute of Science at Bangalore.

From the age of three till he was 16, Ramanujam studied at his father's gurukul. His father was originally from Tamil Nadu but settled in Bangalore, learning nearly 4,000 songs of Alwars composed in Tamil in the Sangam era. Like his father, he continues the *guru-shishya* parampara and still has 15 shishyas under his care who take lessons in Taittireeya Krishna Yajur Veda, songs of Alwars and Vishista Dvaita Vedanta which he delivers in the traditional pravachana style.

Following his graduation, Ramanujam joined the Bangalore-based Hindustan Aeronautics Limited and worked for a couple of years on the Light Combat Aircraft project.

He was among the handful of Indians to be trained on IBM computers.

In 1989 he presented a technical paper titled 'Computer representation of Shabdha-bodha' during the first national conference on 'Computers and

convey," says about the experience. This was followed by a paper Bangalore. titled 'A case for Sanskrit as

Ramanujam at a young age of 31 by the Veda nce. Dharma Paripalana Sabha of 1991 September In

Ramanujam made a presentacomputer programming lan-With the virtual death of the oral tradition, P. Ramanujam has decided to computerise the ancient Vedic texts for posterity. Hiren K. Bose met the Ved Varidhi with a unique mission

Digitising haistras



Ved Varidhi P. Ramanujam

guage' at the annual convention been an oral tradition which has of the Computer Society of India laid emphasis on memorising in Bangalore.

favour of computerisation of Vedic texts at a conference on Vedas and shastras, the Tirumala Tirupati Devasthanam awarded a grant of Rs 1 crore to set up a library with computers. "One may not derive immediate results by computerising the ancient texts - the accumulated wisdom of seers and sages contained in the Vedas and shastras. But it has to be made presentable and

mar. "Ours has and some scholars believe that if Impressed by his arguments in the ancient texts are computerised, it will only encourage laziness," says Ramanujam. "Well, they do have a point. But when I say computerise the texts, I an emphasising the analysis part." In other words, open-ended R&D.

Take, for instance, the word dharma or satya. It is difficult to translate into languages such as English which do not have the relevant words. If Sanskrit becomes a computer language, ready for analysis for future gen-erations using the tools of infor-proficient in the language

about the Vedas - and at th touch of a few keys.

Ramanujam dreams of future when Sanskrit could 1: used as a platform for compu ers. The machine is able t recognise Sanskrit. "Rather tha digitising the shastras, it would be great if the learned schola recite the shlokas with th machine recognising the Ved phonetics and understanding th relation between words an meaning," he says.

Ramanujam's zeal to compu erise the shastras led to a ic offer from C-DAC in Pur where, since 1990, Ramanuja has been associated wit 'Desika', a natural languas understanding system fe Sanskrit, based on †1 Vyakarana, Nyaya ar Mimansa Shastras, includir Vedic processing. Besides, 1 has been rendering Ved databases into compute analysable form.

to

IIT

of

Panini's gram-

But what excites Ramanuja is his work on 5th century E grammarian Panini, who w the first to compose form grammar with unbelievable th oretical insight. Panini Astadhyavi deals with comple and complicated issues abo language, thought and the rel tionship between form ar meaning.

The Upanishads, Aranyak and the Vedas speculate on la guage and Panini codified a this in crisp statements on which long commentaries are writte "For artificial intelligence or needs to know about what knowledge and that's availab in our ancient shastras, incluing Panini's works," explain Ramanujam.

A man who has tasted tl fruits of information technolog and is in touch with schola and others worldwide, peop who have put the shastras of WWW.IndoNet tl Indologists on the Net Ramanujam strongly feels th computerisation in colleges ar universities should not 1 restricted to the sciences by extend also to those who wor on languages. In fact, Sag University has a course for M.Phil students who work c the multilingual word process

# Browse or search, in Sanskrit!

Thanks to Ramanujan's effort, one can now access Sastras through the computer.

#### SUGANTHY KRISHNAMACHARI

(This is the first of a two-part article on Dr. P. Ramanujan's work on Sanskrit and computers)

I n the late 1920s, Ghanapathi Parankusachar Swami won a prize in Sanskrit. When asked whether he wanted the prize of Rs 3,000 in cash or kind, he asked for books! Thus he acquired a wonderful library. This enabled his son Ramanujan to pore over the books every day.

Ramanujan spent seven years putting the contents of the Sastras into a database. He culled 30,000 sutras from all the Sastras, classified the different aspects of the Sastras, and gave his compendium the name, Sakala Sastra Sutra Kosa.

When a retired professor of Physics from IIT Madras, who became a sanyasi after being initiated by Sringeri Pontiff, Paramananda Bharati, organised a conference in Delhi on Sanskrit and Computers, Ramanujan told him about the kosa and was asked to present a paper at the conference.

The paper was on using computers for Sanskrit. Many IIT professors were present and what caught their attention was that Ramanujan had come up with a flow chart in Sanskrit, and a programme for the generation of nouns. The then President of India, Dr. Shankar Dayal Sharma, was so impressed that he suggested that Dr. Bhatkar- founder director of Centre for Development of Advanced Computing (C-DAC) - make use of Ramanujan's services. In 1990, Ramanujan joined C-DAC, Pune. While in Pune, Ramanujan developed DESIKA, a comprehensive package for generating and analysing Sanskrit words. What does DESIKA do? "Given a Sanskrit word, it gives you the hidden meanings, the meanings with which it is packed. Key in a word and DESIKA



A Dr. P. Ramanujan. PHOTO: G.P. SAMPATH KUMAR

gives you the noun attributes like paradigm, ending type, noun base, number and case, and similarly for verbs."

When Ramanujan joined C-DAC, their ISCII standard was in the testing stage. Ramanujan wrote the Vedic part of the standard.

Around this time, a question was raised in Parliament about what Indian scientists were doing in the field of Computers and Sanskrit. Ramanujan was asked to make a presentation in Parliament. He presented DESIKA, and later gave a demo in the Parliament annexe. The then Prime Minister P.V. Narasimha Rao, who held the Science and Technology portfolio, attended the demo and was amazed at the simplicity of DESIKA.

Ramanujan made a second presentation in Parliament in 1993. The question now was about how to handle differences between Vedic and classical Sanskrit. Ramanujan replied that this would pose no

In 1994, C-DAC began work on Vedic fonts and today, all the Vedas have been rendered machine readable. problems, and showed a 73 by 26 matrix, which he had prepared (73 individual characters in the Vedic part and 26 parameters). For every Vedic syllable, there are three

components- consonant, vowel and accent, and each syllable has 26 parameters, which define it fully.

which define it fully. In 1994, C-DAC began work on Vedic fonts and today, all the Vedas have been rendered machine readable. Searchable. analysable Sastraic contents, Itihasas, Puranas, Divya Prabandham are all now available too, with value added features such as retrieval as word, stem, compounds, including Boolean search. You can use the same keyboard layout for any script. Ramanujan entrusted to students of Veda Pathasalas, the task of typing out old texts. "One lakh pages have been typed, and 600 texts covered. But the task of annotation still remains. because there are not enough knowledgeable people to do the job." Aren't people who study for many years in pathasalas competent to

do this? "Not necessarily. Most of the pathasalas concentrate on rote learning. I feel we can dilute the memorising part and concentrate on analysis. We need to make this kind of study monetarily attractive as

well' Ramanujan was the Principal Investigator for the TARKSHYA (Technology for Analysis of Rare Knowledge Systems for Harmonious Youth Advancement) project, which envisages providing Sanskrit institutions across the country with high speed connectivity, for promoting heritage computing activities. Content has also been developed for online study. Three courses have been designed: Vedic processing, Sastras and manuscript processing. "We have video lectures by 40 scholars. Students can

# Helpful for scholars

Ramanujan has a website parankusa.org, in which he gives the Arsheya system for the Krishna Yajur Veda. This is a topical arrangement of contents. What is actually followed today is the Saarasvatha system, which does not have such an ordering. Giving the Arsheya system alongside the Saarasvatha ordering, has been of great help to many Sanskrit scholars.

access the lectures through their mobiles. If a student wants to search something later, he can do so, for a verbatim transcript is available.' For manuscript processing, a computer application program, called Pandu-lipi Samshodaka has been developed by C-DAC, which has browse, search, index, analyse and hyperlinking features. Ramanujan takes me round his library, which has many rare manuscripts, some of them more than 400 years old. They have all been digitised. He feels students must seek out old manuscripts, for who knows what treasures lie hidden in them? How can we tweak education for students of traditional learning? "A student of Indian logic should study Western logic too. A student of vyakarana must study modern theories of linguistics. Study should be interdisciplinarymathematics in ancient Sanskrit texts and in modern texts; transdisciplinary- that is different areas within Sanskrit such as vyakarana, mimamsa, nyaya; multi disciplinarya student of ayurveda could perhaps study the therapeutical aspects of music.'

## FRIDAY REVIEW | THE HINDU • Friday, April 3, 2015

# Where Sanskrit meets computer science

Proving the compatibility of Science and Sastras, Dr. P. Ramanujan headed a project on 'Computational Rendering of Paninian Grammar,' writes Suganthy Krishnamachari.

(This is the last of the two-part article on Dr. P. Samanujan's work on computer and Sanskrit. The first part appeared in Friday Review on March 1990 (Init's hittp:// www.thelinduy.com/ theorem. The sanstras-through-theorem. The sanstras-through-the said this is true even of plistorical Principles. There is howledge is that even the said this is true even the said this is true even of plistorical Principles. There is howledge is that even the said this is true even the said this is true even the said this is true even of plistorical Principles. There is howledge is that even the said this is true even of plistorical Principles. There is howledge is that even the said theore were possible for a speaker to "use in language. This unconscious knowledge is howledge is than even the said there were provided theorem. The said there were possible for a speaker to "use in language. This unconscious knowledge is howledge is than even the said there were possible for a speaker to "use in language. This unconscious knowledge is howledge is thome even the said the said the said the said the possible for a speaker to "use in language. This unconscious knowledge is howledge is the even the said the said

"most tainous and case." Walter Eugene Clark, who was Professor of Sanskrit at Harvard University, and who translated Aryabhatta's Aryabhatiya into English, wrote that "Panini's

grammar is the earliest scientific grammar in the world, and one of the greatest." He said the "Indian study of language was as objective as the dissection of the body by an antomist." dissection of the body by an anatomix<sup>17</sup> dy, there are scientists who study Paninian grammar, with a view to seeing what application they have in the area of Natural Longage Programme Co-ordinator, Indian Heritage Group-C-DAC, Bengalurn, is an authority on Paninian grammar, With a tuft, a namam on his forehead and a traditional dhoti, he doesn't look like a typical scientist. Ramanujan is

a traditional dhoti, he doesn't look like a typical circle traditional education need not stand in the way of a career in science, for it is his traditional learning which has brought him to where he to be the strady of the hose brows of the brows of the strady of the other strady of t

The richness of Sanskrit comes from the fact that everything is pre-determined and derivable. – Dr. Ramanujan



Rendering of Paninian Grammar.' The richness of Sanskrit comes from the fact that everything is pre-determined and derivable. "There is a derivational process, and so there is no ambiguity. You can explain everything structurally. There is a base meaning, a

suffix meaning and a combination meaning. The base is the constant part, and the suffix is the variable part. The variables are most potent. With suffixes one can highlight, modify or attenuate."

can highlight, moday or attenuate." Two different words may denote an object, but you can't use them interchangeably, for the functional aspect is what matters. For example you can't replace 'Agni' with Yvahni, 'for 'Agni' has its own componential meaning. An object may be denoted by the base. An object can have sets of relationships and interactions with other bitmes in the world. For have sets of relationships and interactions with other things in the world. For example, 'Rama', in relation to other objects, may be an agent of some activity or the recipient etc. "Even the

recipient etc. "Even the interactions have been codified nicely and briefly. Clarity and brevity are the hallmarks of Panini's work. His rule-based approach is

his biggest plus point." Isn't it true that in Sanskrit you don't have to coin words for a new invention or discovery, and you can derive a word to suit the functionality of the object? "Yes. You have all the components with you to derive a word. You can use multiple suffixes, if need be, to show the particular function of an object." Does meaning vary

the particular function of an object." Does meaning vary according to accent? 'It does. For the same suffix, different meanings are derivable because of accent differences. So you have the Divine Couple, Jaganmatha and Jagathpitha. How do yoo show the difference betweer our parents for all time and our parents for all time and our parents in this life alone? Accent helps here. This is how the Vedas are most apt, and this has been fully noted by Panini. 'He gave us a conceptual, functional system. You take

an example, apply the rules and get clarity about what it means. So the structures are clarity about what it means to the structures are clarity about what it means to the structures when the work of a Finnish scholar, who found fault with "The Finnish scholar structures" and the structures of the scholar, who found fault with "The Finnish scholar structures" is de only in the Vedas, and prime rules relating to Vedic are claring to the Vedic are claring

use the Sutra to generate the rest. When Ramanujan explained that his program was done mainly to show how the rules worked, Dr. Awasthi conceded that Ramanujan did have a point. "This just shows that people can be won over, if we are able to show the purpose of something."



नई दिल्ली, ७ मई (वार्ता) । कम्प्यूटर भाषा के रू में संस्कृत के इस्तेमाल के सिलसिले में इलेक्ट्रानिकी विभाग ने पाणिनि व्याकरण के हर सूत्र का पूरा साफ्टवेयर तैयार कर लिया है। यह नेबा कम्प्यूटर कार्यक्रम 'सी-डेक' ने विकसित किया है।

विभाग का दावा है कि इस नए साफ्टवेयर के जरिए संस्कृत का हर शब्द, रूप और धातु रूप तैयार किया जा सकता है। इसके लिए संपूर्ण पैकेज विकसित कर लिए गए हैं।

विभिन्न विश्वविद्यालयों और संम्थानों में संस्कृत व अन्य भारतीय भाषाओं के साफ्टवेयर विकसित करने के कि अभिकलनात्मक प्रतिपादन और मशीनी अनुवाद के प्रयास चल रहे हैं। संस्कृत कम्प्यूटर कार्यक्रम खासतौर से पुणे के इलेक्ट्रानिकी विभाग के संस्थान 'सी-डेक' ने शुरूआती नतीजे सकरात्मक रहे हैं। तैयार किया है।

पिछले तीन साल के दौरान संस्कृत से संबंधित परियोजनाओं के लिए सरकार ने ५८.१२ लाख रुपए उपलब्ध कराए है।

संस्कृत कम्प्यूटर परियोजनाओं के तहत अंब तक हुई उपलब्धियों में संस्कृत भाषा के मूल पाठ का पूरा विकास शामिल है। इसमें आठ लाख से भी अधिक अविकारी शब्दों का पाठ रखा गया है।

इलेक्ट्रोनिकी विभाग के लिए भेद के संबंध में संस्कृत वाक्यों का विश्लेषण करने के लिए पद परिचय

भी विकसित कर लिया गया है और लाल बहादुर शास्त्री के जरिए एक मंच दिया है। राष्ट्रीय संस्कृत विद्यापीठ नई दिल्ली में बीएड के विद्यार्थियों के लिए कम्प्यूटर शिक्षण को एक विषय के रूप में शामिल किया गया है ।

अमेरिको कम्प्यूटर वैज्ञानिकों ने १९८५ में अपने अनुसंधान के बाद यह दावा किया था कि संस्कत सर्वश्रेष्ठ कम्प्यूटर भाषा सिद्ध हो सकती है । क्योंकि यह नियमों पर आधारित भाषा है।

इलेक्ट्रोनिकी विभाग ने १९८९ में पाणिनि व्याकरण लिए संस्कृत भाषा पर काम शुरू किया था जिसके

पुणे के आधुनिक कम्प्यूटर विकास केंद्र ने अपने अध्ययन के दौरान धातुओं से आरंभ होने वाले संस्कृत शब्दों और वाक्यों को सुव्यवस्थित ढंग से तैयार करने वाक्यों व शब्दों का विश्लेषण करने और उनकी पहचान पर विशेष ध्यान दिया गया।

भाषा संसाधन के कार्यकलापों को बेहतर बनाने और संचालित करने की दुष्टि से इलेक्ट्रोनिकी विभाग ने भाषा संस्थानों और कम्प्यरर क्षेत्र में विशेषज्ञता रखने वाले संगठनों को भारतीय भाषा प्रोद्योगिकी विकास कार्यक्रम

संस्कृत में सूचना संसाधन से संबंधित चार बड़ी परियोजनाए शुरू की गई हैं जिन्हें अलग-अलग संस्थान चला रहे हैं।

वाराणसी के सम्पूर्णानंद संस्कृत विश्वविद्यालय के संस्कृत के मूल पाठ का विकास करने और संस्कृत शास्त्र में सूचना संसाधन संरचना का पता लगाने की जिम्मेदारी सौंपी गई है।

नई दिल्ली के लाल बहादर शास्त्री राष्ट्रीय संस्कृत विद्यापीठ को बीएड और एमएड के विद्यार्थियों के लिए कम्प्यूटर की संहायता से शिक्षा देने का साफ्टवेकर विकसित करने की जिम्मेदारी दी गई है।

मंत्रालय के अनुसार जवाहर लाल नेहरू विश्वविद्यालय इस परियोजना में विद्यापीठ को मदद दे रहा है।

इस विद्यापीठ राष्ट्रीय संस्कृत विद्यापीठ तिरूपति और कुल कांगड़ी विश्वविद्यालय हरिद्वार में भाषान्त्रिदों और कम्प्यूटर वैज्ञानिकों को संस्कृत व्याकरण मीमांसा और निरक्त से परिचित कराने के लिए पाठय साल्मग्री तैयार की जा रही है।

कर्नाटक में मेलकोट की संस्कृत अनुसंधान अकादमी में कम्प्यूटर पर आधारित अनुसंधान में संस्कृत उपयोग पर अध्ययन किया जा रहा है।

#### *ଝଔର୍ଚ୍ଚାର୍ ଅଧି*ନ

Parankushachar Institute of Vedic Studies<sup>®</sup>, Bengaluru

a job soon, and so suggested he do a diploma course. After obtaining his diploma, Ramanujan joined HAL, Later on, he graduated in engineering, and did his Masters in Engineering from the second structure of a General Purpose Sanskrit Parser. What would make a study of Sanskrit useful to a student of Compute for a sword, it is ambiguous, but when Sanskrit has many meanings for a word, it is ambiguous, but when Sanskrit has many meanings for a word, it is rich!" says for, a word, it is rich!" says for a word, it is rich!" says for, a word, it is rich!" says for a word, it is r

# COMPUTER AND INFORMATION SCIENCES Current Trends in Applications

Proceedings of the XXIV Annual Convention of Computer Society of India, 20th to 23rd September, 1989 Bangalore, India

Editors

## V. B. Kaujalgi

Professor & Coordinator Computer Aided Management Centre Indian Institute of Management Bangalore

## K. A. Ranganatha Setty

Professor & Head Department of Computer Science & Engineering R.V. College of Engineering, Bangalore

# H. Sivaramakrishnan

Deputy Chief Engineer Microelectronics and Computer Division Indian Telephone Industries Limited, Bangalore



# Tata McGraw-Hill Publishing Company Limited NEW DELHI

McGraw-Hill Offices

New Delhi New York St Louis San Francisco Auckland Bogotá Guatemala Hamburg Lisbon London Madrid Mexico Milan Montreal Panama Paris San Juan São Paulo Singapore Sydney Tokyo Toronto