

XI. TECHNOLOGY

(Under each head, the various approaches/models used, applications, current state of technology, software tools available, future directions, etc. will be given along with relevant references to organizations/ institutions/ companies, products, experts in the field, books, Journals, articles, websites, CDs, Cassettes etc.)

A. Corpus and Corpus Management Tools:

- i. General, Tagged corpus, Parallel Corpus, Aligned Corpora.
- ii. Corpus, Indexing tools (Concordance, K.WIC index, etc.
- iii. Corpus compression and encryption tools
- iv. Text processing tools
- v. Statistical analysis tools.

Academy of Sanskrit Research Melkote Dist. Mandya, Karnataka:

Academy of Sanskrit Research Institute, Melkote, Dist. Mandya, Karnataka is actively engaged in the above noted items such as tagged corpus, parallel corpus, corpus indexing and text processing tools.

Siddhaganga Math, Tumkur, Karnataka:

Siddhaganga Math, Tumkur, Karnataka is also actively involved in the these studies.

B. Text Editors and Word processors:

1. Text editing tools
2. Word processing tools
3. D.T.P. tools
4. Fonts

I.I.T. (Indian Institute of Technology) :

Indian Institute of Technology, Kanpur, U.P. is doing work in this area of Text editing word processing, D.T.P. Tools and Fonts.

Prof. B.N.Patnaik of I.I.T Kanpur has succeeded in developing a system for the analysis of Indian Languages on the model of Pāṇiyan grammar with selected vocabulary. The passer developed in names after Pāṇini.

Academy of Sanskrit Research, Melkote is also doing some work in the above mentioned areas.

C. Dictionary Tools:

1. Word lists/Vocabulary
2. Electronic/Online dictionaries
3. Electronic/On line thesaurus
4. Morphological analyzers/generators.

Deccan College, Pune, Maharashtra:

Deccan College pune has done work on dictionaries both electronic and online.

I.I.T. Kanpur also has done work in this area.

Yāska who was earlier to Pāṇini (4th century B.C.) has written a dictionary called 'Nighantu'. It gives information on derivation of certain vedic words and other relevant information.

Ganga Ram Garg (P. XXVi) says "The great dictionary is the Amara - kośa by Amarasimha, who probably lived early in the 6th century A.D". The other important works in this field are (Abhidhāna - ratnamāla by Halayudha (10th Century A.D.), Vijayani, Abhidhāna - kośa, Anekartha - śabda kośa, the lexical works of purushottama deva, vācaspatya and śabda-kalpadruma. (Apte, V.S. Practical Sanskrit-English Dictionary; Monier-Williams: Sanskrit-English Dictionary.

Cappeller: A Sanskrit English Dictionary; Ghatage: An Encyclopedic Dictionary of Sanskrit on Historical principles; Mayshofer: A concise Etymological Sanskrit Dictionary; Sūryakanta: A practical vedic Ditionary (1981)].

D. Spell checkers/Grammar checkers/style checkers:

IIT kanpur and Academy of Sanskrit Research have done work in this area.

E. Parsing systems:

1. Phonological
2. Morphological
3. Syntactic

IIT Kanpur has succeeded in developing a system for the analysis of Indian Languages on the model of pāṇiniyan grammar with selected vocabulary. The parser developed in names after pāṇini "Vanita Ramaswamy".

Academy of Sanskrit Research, Melkote and Taralabālu Jagadgura peetha of Tumkur are working in this area.

F. Machine Translation and Translation Tools:

1. Translation memories
2. Terminology data Books
3. Post-editing tools
4. Word sense Disambiguation (WSD) tools

Prof. Vanita Ramaswami in her paper entitled "Computer compatibility of Pāṇini's Grammar and its utility" pp. 188 (First Edition 1998) says that "One of the first linguistic applications of Computers to be envisaged was Machine translation (MT) means the translation of one natural language called the output or the 'target' language". But natural language processing is a major problem of Artificial Intelligence.

Artificial Intelligence (AI):

AI means to provide human intelligence artificially to the computer core. But intelligence requires knowledge. The problem is how to represent this knowledge in the machine system. For this purpose many expert systems have been developed.

An Expert system describes the grammatical details of the language fed to the computer and also gives the techniques for logically storing this knowledge in the computer. The more subtle the description the more accurate the translation could be. Morpheme is a meaningful Linguistic Unit. But sentence is the basic unit of translation (Vākyasphota) according to Sanskrit grammarians. This is also called the contextual unit of the language. The context sensitive rules of

pāṇini's grammar appear very significant when we look back into programming languages developed for NLP since 1971....

Briggs writes the "It is not surprising that the attention of the computer scientists was drawn to pāṇini's Aṣṭhādhyayī, an expert system". Pāṇini's grammar is scientific and logical. The technique of knowledge representation is similar to the schemes of AI in computers. Paying glorious tributes to pāṇini and the achievements of ancient Sanskrit grammarians, Briggs writes that 'It is tempting to think of them as computer scientists without hardware'.

The study of pāṇinian grammar has taken a new turn in the twentieth century. Because pāṇini's grammar represents the first attempt of the world to describe and analyze a spoken language on scientific lines. Pāṇini writes rules in the same order the human mind proceeds processing words and sentences. Hence, it is the grammar of man and his mind.

According to Linguists a man has certain innate and finite set of rules capable of generating infinite number of words and sentences. In the case of the computer, it is necessary to make algebraic patterning of the finite rules and then input it to the logical unit of the machine. For example the rule 'ikoyanaci' contains three code words which may be represented as X ay (z) where X & Y are the invariable parts of the rule, Z is optional. To prevent generation of undesirable forms patanjali puts forth 'lokapramanya vada' theory of the authority of usage.

Relevance of Pāṇini's techniques to computational Linguistics:

Pāṇini's grammar reveals the basic views of cognition. Cognitive rules are applied at the user level. Hence, the description of the language is passed on sound morphemes and analysis of words.

1. Pāṇini's classification of words into subanta, tiṅanta and avyaya implied by the sūtras 'Suptiṅatam padam' (1.4.14) and 'avyayad aṣṭupah' (11.4.82) are scientific. Pāṇini makes a sharp distinction between śabda (concept word) and pada (grammatical form) or the word functioning in a sentence. The śabda is used in a sentence only after inflection. Further the eight parts of speech are grouped under subantas (noun, pronoun, adjective), and avyayas (adverbs, preposition, conjunction, interjection) and verb. The function of the word as used in the sentence is very important. The main characteristics of subanta are gender, number and case. There are both grammatical and natural gender.

Sanskrit to architect tomorrow's systems: The aim is to highlight the fact that pāṇini has attempted to describe a language the way the linguists were looking for. He has only shown the direction for natural language processing. He has not exhausted all the possibilities. Further semantic disambiguation may be done with the help of the contributions of later grammarians and other Sastric texts. Then it is possible to build up a core grammar for all Indian languages which is the need of the hour.

The structure of sentences in Indian languages is not linear but hierarchical and the grammatical mechanism underlying them is almost the same. Once the syntactic and semantic inaccuracies are solved in the source language it is for the computer to look up for substitution. The Lexical, Syntactical and morphological tables integrate and interact to give the right meaning. Thus, the prospects of mechanization of translations are bright.

The existing method is based on preprogramming techniques. But this is not advocated as it deadlocks progress. Creating a knowledge base as shown in pāṇini's grammar with self revealing property and necessary basic theories helps automatic information retrieval.

Secondly, a knowledge base structured on pāṇiniyan model may act as a teacher for self learning purpose. The problem of communication seen more in knowledge bases open to public use gets solved, as the pāṇinian model will respond more intelligently and interactively to urgent

queries.

IIT kanpur has succeeded in developing a system for the analysis of Indian languages on the model of pāṇiniyan grammar with selected vocabulary. The parser developed in names of the pāṇini. (Vanita Ramaswamy)

G. Optical Character Recognition (OCR) :

1. Single font/Multifont/Omnifont OCR Systems.
2. Printed/taped/Handwritten/Shorthand
3. Online/Offline.

Information has to be collected on the above noted points.

H. Information Retrieval/Information Extraction:

1. Text mining
2. Web mining

Information to be collected.

I. Search Engines/Web Technologies:

Information to be collected.

J. Speech Technology:

1. Signal processing
2. Text to Speech (TTS)
3. Speech to Text (STT)
4. Speech Recognition/Understanding
 - (a) Language Recognition
 - (b) Speaker Identification

Information to be collected on this from All India Institute of Speech and Hearing, Mysore.

K. Standardization Issues:

1. Character level standards. ISCII/UNICODE
2. Glyph Standardization
3. Keyboard Layout
4. Rendering engines
5. Operating system level support
6. Browser level support.

Information to be collected on this.

REFERENCES FOR TECHNOLOGY

1.Garg, Ganga Ram 1982 An Encyclopedia of Indian Literature (Sanskrit, Pāli, Prākṛit and Apabhramsa) Mittal publishers, 1857 Trinagar New Delhi. First published in 1982

2.Vanitha Ramaswamy 'Computer compatibility of pāṇini's Grammar and its utility' In "Indian Alternatives in Linguistics" professor B.N.Chandraiah Felicitation volume. Editor-in-chief: Prof.D.Javere Gowda Editors: M.R.Ranganatha, V.Gnanasundaram, K.P.Acharya. Published by Vishwamaithri Institute of Research and Rural Development (R) 3. East of B.Ed College, T.K.Layout, Kuvempunagar, Mysore – 570 009.

Copyright CIIL-India Mysore