DOI: https://doi.org/10.46632/cllrm/4/4/2

Contemporaneity of Language and Literature in the Robotized Millennium
Vol: 4(4), 2022
REST Publisher; ISBN: 978-81-936097-3-6
Website: http://restpublisher.com/book-series/cllrm/

Computational linguistics in Sanskrit Grammar

Dr. Samsrutha Devi. A
Assistant Professor, Dept. Of Sanskrit
Sacred Heart College, Thevara, Kerala, India.

Sanskrit Language and Linguistics

The grammar of the Sanskrit language has a complex verbal system, rich nominal declension, and extensive use of compound nouns. It was studied and codified by Sanskrit grammarians from the later Vedic period. The study of language in the Indian tradition began with the desire to maintain the purity of the dominant language and also to maintain the authenticity of scripture and ascertain the authenticity of its transmission. The Vedas were the primary scriptures and they were originally transferred orally. The Vedas also contained hymns that were part of rituals, methods to maintain accuracy of speech were necessary, and a sophisticated system of phonetics was employed to codify the language of ritual to keep it free from change. The earliest attempts of this are seen in 'Padapatha', in around 1000–700 BCE, which extracted individual words from the continuous recitation of the hymns—'Samhitapatha'—and also morphologically analyzed the words into roots and affixes. Vyakarana, which is grammar; and Nirukta, which is etymology and lexicography. The 'vedangas' were developed to help the student to correctly interpret the scriptures. This includes creating teaching aids, learning tools, online resources, along with lectures and textbooks. These tools include language teaching for sanskrit and Indian languages. The educational tools make use of the vast amount of research done by scholars in various fields of linguistics and computational linguistics.

While the study of phonetics and metrics helped in the development of language science, the major impetus came from the study of vyakrana and nirukta (Deshpande 2016). Nirukta is a treatise by Yasaka on the etymology of rare and ambiguous Vedic words, and the morphological analysis of words in terms of roots and suffixes, and grammatical description of a set of words in terms of parts of speech was already available in the Padapathas.

Panini's 'Ashtadhyayi' is a treatise in the vyakarana domain that contains 4000 sutras that provide an intricate system of rules that interpret a confounding array of linguistic matters, like the composition of nouns and case relations, the transformation of roots and nouns using suffixes, accent changes in word-formation and sentence construction.

Patanjali's 'Mahabhashya' discusses Panini's grammar and other commentaries on it like Katyayana's Vrittikas, and deals with technical as well as philosophical issues of Panini's grammar. 'Siddhanta-Kaumudi' is a commentary by Bhattoji Dikshit, who reordered the sutras in a way that makes them easy to follow for a student but distorts the architecture of the system designed by Panini.

Panini's Ashtadhyayi

Panini was a Sanskrit grammarian who gave a comprehensive and scientific theory of phonetics, phonology, and morphology. Sanskrit was the classical literary language.Panini is considered the founder of the language and literature.Panini wrote Sanskrit Grammar.Pāṇini is known for his text Aṣṭādhyāyī, a sutra-style treatise on Sanskrit grammar, 3,996 verses or rules on linguistics, syntax and semantics in "eight chapters" which is the foundational text of the Vyākaraṇa branch of the Vedanga.

The Ashtadhyayi and the tradition of commentaries that has developed over the ages provides us with perhaps the most comprehensive grammar ever developed for any language. So, though knowledge of Sanskrit is necessary to understand and apply this 'grammar', the framework developed is general enough to be applicable to many modern Indian languages as well as languages of other families. The other area of research is the format in which the Ashtadhyayi is presented and not just the use of coded language in the concise 'sutra' format but also the arrangement of rules by Panini has much to add to our knowledge of how algorithms, particularly for language processing and generation, are written.

The Paninian model is concerned with extracting information out of language and deals with language as a self-contained phenomenon, as much as possible, and therefore makes little appeal to knowledge of the world. This aspect brings it closer to the goals of natural language processing because it is difficult to encode 'common sense' in a machine. Concepts like 'the verbal root' and 'kaaraka relations' form the foundations of the Paninian model of language. The action represented by a verbal root, or 'dhatu', is claimed to be composed of 'vyapara' (an activity) and 'phala' (a result) thus providing a simple system to analyze the information contained in 'verbal roots'. For example, the action of 'opening a lock' will entail a list of sub-actions like inserting the key, twisting the key and, the moving of the levers, which lead the final state, the 'phala'. Each action and sub-action is carried out by a different participant, 'kaaraka', which is related to the objects in specific ways and the most independent of these participants is called a 'swatantrakarta'. In the model there are about six kaaraka relations that attempt to define the almost infinite relations possible between states and objects (verbs and nouns). The concept of 'vivaksha' in the Paninian model deals with the point-of-view or the attitude of the speaker and it defines the choice of action and how it is related to the participants because it is understood that no linguistic utterance states only facts but always includes the attitude and intention of the speaker.

Kaaraka relations express semantic as well as syntactic information encoded in a sentence as they map nominal and verbal elements in a sentence but this is not the level that describes how a sentence is uttered. Before the rules of phonetics are applied, the model creates 'vhibhakti'. The 'vhibhakti' for a verb is composed of the markers of tense, aspect and modality and that of a noun is composed of information like person, gender and, number.

Computational linguistics

Computational linguistics is the scientific and engineering discipline concerned with understanding written and spoken language from a computational perspective, and building artifacts that usefully process and produce language, either in bulk or in a dialogue setting. To the extent that language is a mirror of mind, a computational understanding of language also provides insight into thinking and intelligence. And since language is our most natural and most versatile means of communication, linguistically competent computers would greatly facilitate our interaction with machines and software of all sorts, and put at our fingertips, in ways that truly meet our needs, the vast textual and other resources of the internet.

Computational Linguistics in Sanskrit Grammar

The first major area is of course Machine Translation, for Sanskrit and for Indian Languages. Other areas of research in computational linguistics include designing of natural language processing tools and building lexical resources for Sanskrit.

A possible solution to the problem of machine translation is provided in a system called Anusaraka. This is not a machine translation system but what is termed as a 'language accessor'. This system uses a parser based on Paninian grammar. The output of the system is not grammatical but is comprehensible ans thus it can be used by people who have basic understanding of the target language but need to translate text from one language to another. Consider for example a scholar who needs to access

a certain text in Sanskrit but has only basic knowledge of Sanskrit. In this case Anusaraka is a tool that this scholar can use to help him/her in the process of translation.

An important task of computational linguistics research is building text to speech engines, which are based on research in phonetics and the tools created as a result are useful not only as stand alone projects but also as input for larger systems. Another research task is optical character recognition, which is vital in building our database of digital texts. Apart from these research areas, an important task for researchers is making sure that the next generation is prepared to continue the research and for that building of educational tools is of utmost importance. This includes creating teaching aids, learning tools, online resources, along with lectures and textbooks. These tools include language teaching for Indian languages and Sanskrit. The educational tools make use of the vast amount of research done by scholars in various fields of linguistics and computational linguistics, so that information and knowledge is not only accessible but also engaging.

Conclusion

Sutras are like mathematical formulae. So, a lot of information can be given using a few words. And since Panini uses sutras, the Ashtadhyayi is crisp. Panini also lays down rules to resolve conflicts between sutras

Computational linguistics is used in tools like instant machine translation, speech recognition systems, text-to-speech synthesizers, interactive voice response systems, search engines, text editors and language instruction materials. Machine translation systems are a wide range of commercial systems that automatically translate one natural language into another available today. However, the quality of the translations produced by these systems is still limited, and machine translation remains one of the most important research areas of computational linguistics.

Natural-language access systems are programs that provide natural language interfaces for databases or expert systems instead of using formal query languages that are hard to learn. Speech recognition and speech synthesis systems: Speech-driven operation of devices, dictating machines, text-to-speech systems, telephonic dialogue and information systems.

Systems for accessing and managing texts: Document retrieval (e.g. for extensive technical documentation, legal statutes), text filtering (sorting messages from electronic news services customized electronic newspapers), text understanding (partial comprehension of the content of texts and answering questions about it), automatic abstraction (compiling summaries of texts with respect to a specific profile or a specific question). Systems for generating written texts: Generating technical documentation from a formal description of the respective devices, as well as generating stock market reports, weather reports, etc. from raw data. Intelligent text critique systems: Grammar and style correction programs.

Intelligent computer-aided language learning systems: Common contemporary systems simply compare the student's input with fixed sequences of words and thus judge them as "correct" or "wrong". These systems can be improved by determining the grammatical structure of a sentence entered by a student, evaluating its correctness and locating potential errors as precisely as possible. For computationl linguistics, the paninian tradition has insights to offer in many areas of research.

Reference

Cardona, George. 2007. Panini: A Survey of Research. Delhi: Motilal Banarsidass Publishers.

The Aṣṭādhyāyī of Pāṇini with Translation and Explanatory Notes

Volume 8, By Pāṇini, S. D. Joshi, J. A. F. Roodbergen · 1991

Craggs, Laura. 2011. 'Indian Language Traditions and their Influence on Modern Linguistics'.Vernaculum.

Orara, E. de Guzman. 1967. 'An Account of Ancienct Indian Grammatical Studies Down to Patanjali's Mahabhashya: Two Traditions'. Asian Studies: 369–76.

Deshpande, Madhav. 2016. Language and Testimony in Classical Indian Philosophy, Stanford Encyclopedia of Philosophy. Online at https://plato.stanford.edu/archives/fall2016/entries/language-india (viewed on January 10, 2018).