

Architecture of Sanskrit Language with Special Reference to Sentence Generator

Dr. Haripada Mahapatra

Asst. Professor (Teacher-in-Charge) Department of Sanskrit
Sankrail Anil Biswas Smriti Mahavidyalaya, Jhargram

Abstract: The fundamental linguistic principles needed to create a phrase from its meaning construction are provided by Pini's grammar. The word significance is represented abstractly by the meaning structure. It is the intermediary depiction from which the intended phrase can be produced using Pini's principles without relying on general knowledge. The language of Pini offers a syntactic structure for production. While the word level of Sanskrit creation is where the intricacy resides, phrase level generation is fairly simple. Choosing the level of precision of the semantic relations suitable for both analysis and production posed the generator's only design challenge. In this article an attempt has been made to study the architecture of Sentence Generation in the Semantic Level, Karaka level and Vibhakti level. Information has been gathered for this purpose by consulting a variety of books, periodicals, and online sites.

Keywords: Sentence Generated, Pāṇini's Rules, Semantic Level, World Knowledge, Grammar.

Introduction: A person is unique among all other living things because of their use of language. A method of contact is language. In a two-way process called communication, the speaker converts his ideas into a language string, which the recipient then decodes into his or her own thoughts after hearing the language string as sound waves. The first procedure is known as production, while the second is known as research. The process of creating significant words and writing from information given in a rational meaning depiction is known as natural language generation. It could be considered the polar opposite of natural language understanding. According to Pini's grammar, which is an abstract depiction of word import (bdabodha), a phrase can be generated from its meaning structure. It is the intermediary depiction from which the intended phrase can be produced using Pini's principles. This meaning structure also serves as a representation of the sentence's interdependence parse at the same time.

There is a dearth of knowledge regarding the basic issues individuals face when words and writing are generated by machines. For instance, the user cannot obtain any kind of knowledge regarding the precise relationship between St and Rma in order to create a phrase like Rmeasaha Stvanagacchati. Similar to how one cannot expect a user to provide the name of the connection as sahrtha from the perspective of creation. Whether writing is generated by a person or a machine, the human element is inevitable. Concept dictionaries can be created once this issue has been reduced. The vocabulary terms of any language can then be transferred to these ideas. Any language's generator is tailored to work best for that language and is therefore closely related to its grammatical structure and lexicon.

Background of the Study: One language can be translated into another using a variety of techniques in the area of natural language processing. In the fields of linguistics and artificial intelligence, natural language processing is used to facilitate dialogue between machines and human languages. In order to make encounters with machines and technology as user-friendly as

feasible, NLP, which is concerned with human-computer interaction, works to allow computers to understand human English. For linguistic translation, numerous other methods have been created in languages like Marathi, Telugu, Kannada, Punjabi, Hindi, etc. The syntax in Sanskrit is the strongest and makes it easy to comprehend each phrase. Panini, who created 3,949 norms for the language, is credited with creating Sanskrit. In NLP, it's crucial to comprehend how words are connected to create meaning in addition to the individual words themselves.

Significance of the Study: This article discussed different approaches to natural language generation, input and output specifications, the architecture of the Sanskrit sentence generation system, which has four levels from semantics to surface level and three mappings, as well as an internal representation of the input in the required format and output produced by the system. The grammatical extraction module, the central component of Sanskrit sentence generation, is the foundation for the second transfer from karaka-level to vibhakti-level, which is explained in this chapter. The current research is important enough in this respect.

Objectives: The present study has been carried out with the following objectives-

- To study the architecture of Sentence Generation in the Semantic Level
- To study the architecture of Sentence Generation in the Karaka Level
- To study the architecture of Sentence Generation in the Vibhakti Level

Difference between a Morphological Generator and a Sentence Generator

Why is a sentence generator necessary when a grammatical generator already exists? It is a simple inquiry about the Sanskrit Sentence Generator that needs to be addressed. It is common knowledge that the foundation of the sentence generator is the grammatical generator. But each serves a unique purpose. A morphological generator is restricted to the word level and lacks the relationship data found in sentences about other things. However, a sentence generator must pay attention to the following elements—

Number and person: The verb form reflects the number (vacana) and person (puruṣa) of the noun with which it agrees.

- ✓ saḥgrāmaṃ gacchati
- ✓ taugrāmaṃ gacchataḥ
- ✓ tegrāmaṃ gacchanti

According to person and the use of conjuncts in a sentence the declension of the verb changes.

- saḥcaahaṃ cagrāmaṃ gacchāvaḥ
- ahaṃcatvaṃ cagrāmaṃ gacchāvaḥ
- saḥcatvaṃ cagrāmaṃ gacchathaḥ

Voice: “According to kartaripra-yoga (active voice), karmaṇipra-yoga (passive voice), and bhāvepra-yoga (passive impersonal) the case markers of a noun and hence the declension of the verbal root changes and the verbal root mutually gets different case markers. Sanskrit also allows sentences where the non-finite verbal forms (kṛdanta) act as a finite verb as in the following sentences”.

- manuṣyaḥ kukkuraṃ vyāghraṃ kṛtavān. (ktavatu)
- manuṣyeṇa kukkuraḥ vyāghraḥ kṛtaḥ. (kta)

“In such cases also, the inflectional suffix of the kṛdanta should agree in number and person with that of the noun in the nominative case”.

Word order: “Most of the Sanskrit sentences are flexible in word order. In the case of viśeṣaṇa viśeṣya bhāva, , and special indeclinable like kim word order is important”.

1. saḥkim khādati? (kim is a kāraka)
2. kimsaḥ khādati? (kim may be a kāraka or yes/no)
3. saḥ khādati kim? (kim only a yes/no)

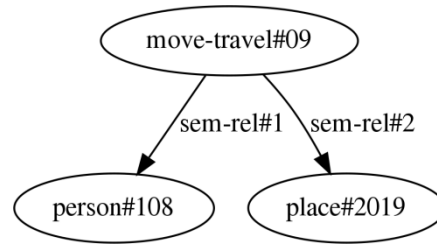
“Similarly another special usage is use of tatra before bhavat for respect as in”—

- kvatatra bhavatī śakuntalā?
- ādiṣṭaḥ asmitatra bhavatā kāśyapena.

Concord of qualifier and qualified: “The qualifier (adjective) gets the same gender, number and case marker as that of the qualified”-

Semantic Level:

This stage reflects the speaker's internal feelings. The speaker has recognized the concepts and has given them concrete meaning in his thoughts, so the material is still at the basic level. Let's say the speaker saw an occurrence where someone was departing a location and travelling to a particular location. Let's presume for the purposes of our conversation that the speaker has indicated the traveller as person (1), the location as place (2), and the activity as move-travel.(3). Additionally, the speaker has chosen to emphasize the portion of the activity of travelling in which the person (1) acts independently and that the objective of this activity is location.(2). This demonstrates the conceptual connections between individual (1), move journey (3), and location (2), as well as between move-travel.(3).



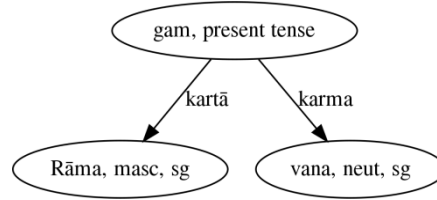
Kāraka-Level: The speaker now selects from among all the alternatives that reflect each of these ideas the vocabulary items that are suitable in the situation in order to communicate this. For instance, the speaker selects a vocabulary phrase from the synonyms Rma for the individual. {ayodhyā-pati, daśaratha nandana, sītā-pati, kausalyā nandana, jānakī-pati, daśa-ratha-putra, Rāma, ...}. The speaker selects the vocabulary words vana and gam to correlate to the other two ideas in a similar manner. The knowledge about pada and ga'a, as well as its meaning, is linked with the word gam.

The speaker now selects suitable kraka-labels that correlate to the meanings linked with the chosen relations after selecting the vocabulary items to represent the ideas. He selects the accent in which to deliver the statement as well. Let's presume that the speaker in our example chooses to use the active voice to describe the incident. Now, the Adhyya stras are put into use.

The semantic roles are mapped to kartā and karma, following the Pāṇinian sūtras—

- svatantraḥ kartā; which assigns a kartā role to Rāma.
- karturīpsitatamaṃ karma; which assigns a karma role to vana.

Let us further assume that the speaker wants to convey the information as it is happening i.e., in the present tense (vartamāna-kāla)



This information is alternate lyre presented in simple text format as shown below

wordindex	stem	features	role
1	Rāmapuṃ	eka	kartā3
2	vananapuṃ	eka	karma3
3	Gamparasmai pada bhvādi	vartamāna	kartari

The word index, which is represented by the first column, is used to identify a word when designating positions. “The second field is the stem (with gender in case of names), the third field provides grammatical characteristics such as number, mood, etc. and the fourth field provides the role designation and the index of the word with regard to which the role is noted. If this knowledge were to be verbally expressed, what we get the traditional śābdabodha such as eka-vana-abhinna-karmaka,-eka-rāma-abhinna-kartṛka,-uttaradeśa-samyogānukūla-vartamānakālika-vyāpārah”.

Vibhakti-Level: “Now the sūtras from vibhakti-section of Pāṇini’s Aṣṭādhyāyī come into play. Vana which is karma gets accusative (dvitīyā) case marker due to the sūtraka-rmaṇid-vitīyā (anabhihite). Since the sentence is desired to be in active voice, kartā is abhihita (expressed), and hence it will get nominative (prathamā) case due to the sūtra – prāti-padikārtha-liṅga-parimāṇa-vacana-mātre-prathamā. The verb gets a laṭlakāra due to vartamāna-kāla (present tense) by the sūtra -vartamānelaṭ. It also inherits the puruṣa (person) and vacana (number) from the kartā Rāma, since the speaker has chosen an active voice”. Thus at this level, now, the information available for each word is as follows.

Word index	stem	Morphological features
1	Rāmapuṃ	ekaprathamā
2	vananapuṃ	ekadvitīyā
3	Gam parasmaipada bhvādi	laṭprathamaeka

SurfaceLevel: With this knowledge, the accessible word generator is now used to create each pada. At the phrase level, sandhi is discretionary. The sandhi-rules enter into action and a statement with sandhi is created if the speaker means. Thus, as an outcome, we receive either Rmavana gacchati or, potentially, Rmovana gacchati.

wordindex	stem	features	role
1	Rāmapuṃ	eka	kartā3
2	vananapuṃ	ekavartamāna	karma 3kartari
3	gam ₁		

Conclusion: Most Indian languages descended from Sanskrit, which is also a member of the Indo-European language family. Understanding Sanskrit may make it easier to comprehend other commonly spoken Indo-Aryan languages in the Indian subcontinent. A non-native speaker might find it difficult to produce inactive structures, which are primarily found in Sanskrit writing, with just one click. • The word sequence is not predetermined by the algorithm. So, using any combination of words, one can create a statement. It should also be feasible in the future to offer a generator that will assist the user in rendering the text in a specific prosodic metre.

References:

1. Becker, T. (2006). Natural Language Generation with Fully Specified Templates, pages 401–410. Springer Berlin Heidelberg, Berlin, Heidelberg.
2. Bharati, A. and Kulkarni, A. (2011). ‘Subject’ in English is abhihita. In Cardona, G.,
3. Aklujkar, A., and Ogawa, H., editors, Studies in Sanskrit Grammars. D.K. Printworld, New Delhi.
4. Cardona, G. (2007). On the Structure of Pāṇini’s System, volume 1&2, pages 1–31. Springer.
5. Goyal, P. and Kulkarni, A. (2014). Converting phrase structures to dependency structures in sanskrit. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 1834–1843, Dublin, Ireland.
6. Dublin City University and Association for Computational Linguistics.
7. Goyal, P., Kulkarni, A., and Behera, L. (2009). Computer Simulation of Aṣṭādhyāyī : Some insights. Springer Verlag LNCS 5402.
8. Houben, J. E. (2009). Pāṇini’s grammar and its computerization: A construction grammar approach. Sanskrit Computational Linguistic, 3:6–21.
9. Kiparsky, P. (2009). On the Architecture of Pāṇini’s Grammar, volume 1&2, pages 32–94. Springer, Berlin, Heidelberg.
10. Kulkarni, A. (2019). Appropriate Dependency Tagset for Sanskrit Analysis and Generation. In Proceedings of Sanskrit in China International Conference 2019: Sanskrit on Paths. Forthcoming.
11. Madhavan, M. and Reghuraj, P. (2011). Application of Karaka Relations in Natural Language Generation. Govt. Engg. College Sreekrishnapuram Palakkad, India-678633.
12. Pan, S. and Shaw, J. (2004). Segue: A hybrid case-based surface natural language generator. INLG 2004.
13. Ramakrishnamacharyulu, K. V. (2017). Vibhaktyarthavicāraḥ. In Sārasvatam, A Book of Proceedings of National Seminar, 2016-2017. Rashtriya Sanskrit Sansthan, Guru-vayoor Campus.
14. Shailaja, N. (2014). Comparison of Paṇinian Dhātuvṛttis. PhD thesis, Department of Sanskrit Studies, University of Hyderabad.
15. Subrahmanya, K. (2012). Mahāvākyārthavicāraḥ. Śrīlakṣmaṇārya Bhavyasmṛti Granthamālā, Bhāgyanagaram.