Development of a General-Purpose Sanskrit Parser

A Thesis

Submitted for the Degree of

Master of Science (Engineering)

in the Faculty of Engineering

By

P. Ramanujan

Department of Computer Science and Automation

INDIAN INSTITUTE OF SCIENCE

BANGALORE-560 012, INDIA

MARCH 1999

Development of a General-purpose Sanskrit Parser

A Thesis submitted
for the award of the degree

Master of Science (Engg.)
in the
Engineering Faculty

By

P. Ramanujan
C-DAC, Bangalore

to

Department of Computer Science and Automation
Indian Institute of Science
Bangalore

March 1999

श्रीः

मच्छक्तिमात्रगणने किमिहास्ति शक्यं ?

शक्येन वा तव करीश ! किमस्ति साध्यम् ? ।

यद्यस्ति, साधय मया, तदपि त्वया वा,

किं वा भवेत् भवति किंचिदनीहमाने ? ।।

What is possible to achieve here, counting only my abilities ? Or even with what is possible, what is to be accomplished for you, O Lord (Śrī Devādhirāja Svāmī) of Elephant hill (at Kāñcīpuram) ? If any such thing exists, kindly get it done by me, for, even that would be BY YOU indeed, as What can happen if you were to be even slightly unwilling ?

- (Ācārya Sārvabhauma Śrī Vedānta Deśika, Varadarāja Pañcāśat, 5)

एकयैव गुरो र्दृष्ट्या द्वाभ्यां वापि लभेत यत् ।
न तत् तिसृभिरष्टाभिः सहस्त्रेणापि कस्यचित् ।।

That which is got by only a single sight (or two) of a Guru is not got by three, eight or thousand of somebody (else - meaning, respectively, Paśupati, Prajāpati or Śrīpati).

- (Ācārya Sārvabhauma Śrī Vedānta Deśika, Subhāṣita Nīvī)

DEDICATED TO

MY LATE FATHER

Salakṣaṇa-Ghanapāṭhī, Veda-Sāhitya-Vidvān,

Bhūloka-Nityasūri,

Śrīmadubhaya-Vedāntācārya

Śohattūr **PARĀṄKUŚĀCĀRYA** *Svāmī*

ACKNOWLEDGEMENTS

I thank the Almighty for providing me an opportunity to research in this area of Sanskrit and Computers. I pay my humble respects to my teachers in traditional as well as modern disciplines for equipping me with a modicum of competence to undertake the task. Centre for Development of Advanced Computing (C-DAC), Bangalore, deserves my wholehearted gratitude for enabling me to work on Pāṇinian Grammar and Vedic texts under KBCS and further TDIL Project funded by Dept. of Electronics, Govt. of India.

I am grateful to Dr. BB Prahlada Rao, my initial Organisational Guide and Dr. U Nagaraj Shenoy, the current Organisational Guide for their encouragement and useful technical interactions. I feel guilty of not living upto the latter's expectations despite his highly intellectual inputs and genuine concern for improving my grasp of CS (it was he who originally persuaded me to register for this research). I am also beholden to Prof. M Narasimha Murty, Dept. of Computer Science and Automation (CSA), Indian Institute of Science (IISc), my Research Supervisor for his being kind enough to consent to be my guide and for the unstinted support. I'm indebted to CSA Dept. to have acceded to my request to have Natural Language Processing (NLP) activities in the Indian context to be brought to focus in my work.

The faculty and students of M.Sc. (Engg.) batch of '96 at CSA, colleagues and project students at C-DAC and many academics have all made this research activity pleasant and rewarding. If at all there is any success in this endeavour, I owe it to every one of these.

Project students Sabarinath, Saji David, Tarun Sharma, Ajay Seth, Arun Mattikalli, Naushad Kapasi, Prakash Koth, Vineet Jain, Shanmukhappa Angadi, Manikandan, Thirunavukkarasu, Sajeev, Harshavardhana, Krishna Kumar and Archana; Project staff Vidula Chinchwadkar, Sridharan, Satakopan, Hari Prasad have all contributed to various parts of implementation of *DEŚIKA*, the prototype Natural Language Understanding (NLU) System for Sanskrit, wherein, the ideas presented here have been tried out. This research was partially funded by Dept. of Electronics, Govt. of India, under the scheme, Technology Development for Indian Languages.

I fully realise my limitations on the formal Computer Science (CS) Theory side and found it difficult to master any major topic to a reasonable extent. However, the support and encouragement from many well-wishers (particularly, Prof. PS Sastry, Dept. of Elec. Engg., IISc) saw me through to this stage of presenting this thesis.

TABLE OF CONTENTS

Part I
Chapter 1 : Introduction 01
1.1 Organisation of the Thesis 01
1.2 Abstract 01
1.3 Aim of the Project 02
1.4 Motivation 03
Part II
Chapter 2 : Problem Definition 09
2.1 Problem Overview 09
2.1.1 Subanta Recognition 09
2.1.2 Tiṅanta Recognition 10
2.1.3 Sentence level program 12
2.2 Steps of Computation 13
2.3 Constraints 14
2.4 Conceptual basis 15
Part III
Chapter 3 : Uniqueness of Accented inputs 25
3.1 Distinction between classical and Vedic Sanskrit 25
3.2 Vedic characters and Accent markers 25
3.3 Specialities of Vedic Sanskrit 26
3.4 Accented word-types 28
Chapter 4 : Morphological Analysis 32
4.1 Case of Accented inputs 32
4.2 Accent determination using rules 32
4.3 Examples of Vikṛ̥ti-s (patterns) 33
Chapter 5 : Language structure 36
5.1 Definition of a word 36
5.2 Word Structure 36
5.2.1 Non-compound words 37
5.2.2 Compound words 41
5.3 Word Analysis 45
5.4 Word-Splitting - multiple splits 48
Chapter 6 : Sentence Analysis 56
6.1 Kāraka - Vibhakti mapping : Functional assignment to words 56
6.2 Syntactic Analysis 56
Chapter 7 : Functional Semantics 58
7.1 Root Meanings (Activities) 58
7.2 Activity classification 59

7.3 Scheme of specification 59
7.4 Grouping 60
7.5 Characterisation 62
7.6 Conflict Resolution & disambiguation 64
Chapter 8 : Śābda Bodha (Verbal import) 68
8.1 Concept of Śābda Bodha 68
8.2 Fundamental Metaphysical Categories 68
8.3 Case relations used 69
8.4 Examples 70
Part IV
Chapter 9 : Design and expected outputs 80
9.1 System details of General Purpose Sanskrit Parser 80
9.2 Algorithm for Morphological Analysis 80
9.3 Śābdabodha Flow-chart explanation 82
9.4 Expected outputs 83
Part V
Chapter 10 : Remarks 138
10.1 Conclusion 138
10.2 Future Scope 138
10.3 References 138
Annexures
Annexure - A - Paradigm Analysis of Nominal stems in Sanskrit 140
Annexure - B - Vedic font key maps, samples of various accented Vedic texts 149
Annexure - C - List of verbal meanings (activities) 158
Annexure - D - Pāṇinian rules and Kāraka Theory 167
Annexure - E - Activity categories and specification 173
Annexure - F - Preliminary semantic lexicon 181

PART I
Chapter 1 : Introduction

1.1 Organisation of the thesis

The report is organised into five parts, *viz.* Prologue, Problem definition, Theoretical details, System Design and Epilogue. The first part has the Introductory Chapter 1 covering synoptical details of the project, aims and motivation. The second part has the Chapter 2 giving the overview in the light of an earlier similar effort elsewhere and our perception and contrast. The third part has Chapters 3 through' 8 covering a vast range of info. required and auxiliaries created for the design. Chapter 3 stresses the phonological aspect of accents in Vedic texts and their special features. Chapter 4 follows this up with morphology-related details with illustrations. Vedic specialities, particularly, accents, the essence of Vedic texts, are treated at length to understand their links to semantics. Chapter 5 visits the Sanskrit language structure as required for computational processing, characterising it for splitting, analysis etc. An interesting case of multiple splits is described as a basis for hoping to meet the challenges of inexhaustible possibilities. Sentential analysis with the well-known Pāṇinian Kāraka Theory with the K-V Mappings is detailed in chapter 6. We move to functional analysis of semantics in Chapter 7, where, we have looked at verbal root meanings comprehensively from the Pāṇini's Dhātupāṭha to come up with classification, grouping scheme, specification and characterisation of a few activity groups to link semantics ontologically to syntax. A resultant ontology is annexed. Chapter 8 discusses the Śābda-Bodha, conceptually and illustratively. The Fourth Part has Chapter 9 enumerating the design of the parser through the algorithms, K-V Mappings, activity groupings, ontology etc. and has numerous figures of design and outputs expected. Fifth Part containing Chapter 10 formally closes the work with concluding remarks and customary references.

1.2 Abstract

Sanskrit is an inflectional language and words in sentence carry information about entities in terms of stem, ending, gender, case, number and case-relation, while verbs denote activity, functor relation, mode, voice, tense, person, number etc. Extracting and organising these info. is the first step towards understanding the language.

The problem is that the same Rūpa (finished wordform) may generate multiple answers. It is frequently the case that a subanta has multiple vibhakti-vacana pair. Sometimes, the Prātipadika (stem) also may be different. A Rūpa may also be Subanta and Tiṅanta, Subanta and Kṛdanta, Kṛdanta and Tiṅanta, Avyaya and Subanta or Tiṅanta etc. It is the task of the word-level program to give the exhaustive analysis of a Rūpa.

It is intended to design a General purpose Sanskrit parser, which generates a parse tree for a given sentence if it can be derived from the grammar. To arrive at the desired result the problems of language as that of human communication, various aspects like syntax, semantics and pragmatics are used integrally.

This thesis concerns the design (and implementation) of a system that provides help in understanding Sanskrit. Any such system has to address the problem at different levels like word, sentence and passage. The accented lexicon based on Sage Yāska's Nighaṇṭu is also built computationally. This report covers the word and sentence levels fully and passages at design level.

The word-level program (morphological analyser) extracts information from various categories of words like Subanta, Tiṅanta, Kṛdanta, Taddhita, Sasvara (accented), Avyaya, Sandhi and Samāsa. The sentence level program uses these information, does parsing and generates a paraphrase of the input sentence or provides output in many other forms.

The sentence level program solves problems like ambiguity etc. by parsing and produces the paraphrase etc. extracting word information from the word level program. The organisation and structuring of the words are decided using the decision tree based on the Vibhakti-Kāraka Mapping given in Pāṇinian grammar rules.

The sentence analysis as per the grammatical is basically an expectancy driven approach, (traditionally called Śābda-bodha) similar to Semantic nets, but has addl. constraints from the many-to-many mappings between case-markers and case-relations. Words of the sentence are grouped into source words and demand words. If the assignment of words to these classes are not unique, due to multiple identifications into nominal and/or verbal categories, each possible case is considered and the procedure iteratively applied. Avyaya-s and other particles are also handled and assignment of functional roles upto sub-Kāraka level is done using the semantic lexicon, based on the ontology developed.

It is felt that the general-purpose parser with provisions to handle even accented texts will be a good candidate for integrating speech processing as well and since languages are human communication media, this help would be desirable. The refinement of ontology and specification of entities, attributes, activities and relations in tune with Śāstra-ic inputs would assist even classical literature and language leaning/teaching/research.

1.3 Aim of the Project

The research reported in this thesis has the main objective of bringing to fore the salient features and specific requirements of parsing Sanskrit language to be usable for Computer-based processing of any variety of its literature. **While** most western NLP developments are : I) with English or similar positional languages as the focus; ii) assuming implicitly that (even natural) language = text, *i.e*, written; iii) for languages which have linear scripts and lack euphonic combinations and major compounding (as they are not phonetic-based); iv) having speech processing delinked from NLP and **while** the Indian effort at Indian Institute of Technology, Kanpur is with a fixed (target) language among the language-pair considered for Machine Translation (MT) (as Hindi, a language akin to English in structure due to absence of Sandhi and presence of phrase or local word group structure) and **both** these are predominantly oriented towards Machine Translation applications, the current research has been launched to design a general-purpose parser for Sanskrit with all the theoretical, *i.e*, linguistic details covered. However, it draws upon generously from these researches.

We intend developing an algorithm for Morphological analysis of plain and accented Sanskrit inputs, to examine exhaustively the syntactico-semantic features of Kāraka-Vibhakti mappings, devise an ontology for word-sense determination and a detailed activity categorisation for semantic processing. Survey of the relevant Śāstra-ic literature and building upon the basis of IIT, Kanpur efforts is also part of the exercise. Phonological aspects are also provided for while analysing bases/affixes.

It is intended to answer the question, whether the sentence given as input can be derived from the grammar. If so the derivation or the parse tree is drawn highlighting the role of words in the sentence.

The first step is word-level program (morphological analyser) which extracts information from various categories of words like Subanta, Tiṅanta, Kṛdanta, Taddhita, Sasvara (accented), Avyaya, Sandhi and Samāsa for the word tokens in the sentence. The sentence level program uses these information, does parsing and generates a paraphrase of the input.

The sentence analysis as per the grammatical tradition is set forth further. This is basically an expectancy driven approach, (traditionally called Śābda-bodha) similar to Semantic nets, but has addl. constraints from the many-to-many mappings between case-markers and case-relations. Avyaya-s and other particles are also handled and assignment of functional roles upto sub-Kāraka level is done using the semantic lexicon, based on the ontology.

The information regarding the word categories is difficult to organise and represent. This involves Knowledge Representation issues and techniques. We start from the labels meant in the 106 Sūtras of grammar aforesaid to build the first level ontology, which of course, can be extended further with activity categorisation. These would result in the numerous categories at various layers. Instances are to be listed from lexicons like Amarakośa wherein classes are also used. Adjectives, modifiers and qualifiers are treated as another category called Viśeṣya Nighna as per Amarakośa.

At the level of passages or poetry where a single verse can have more than one sentence, the semantic extraction involves the connection between the various parts and handling 'reverse sannidhi' conditions. Accented words have sentence boundary demarcation also to be carried out using Mīmāṃsā principles and Vedic section of grammar rules. Accent-markers are considered for fixing senses and connections between sentences or clauses.

1.4 Motivation

The `Can A Sanskrit Parser Do ..... ?` series of questions at etymological, lexical, morphological, phonetic, metric, syntactic, semantic, contextual, pragmatic (allegoric), figurative and socio-religious levels. *[If it can, we have some use for one.]*

1. Can we get Pada-pāṭha from Saṃhitā-pāṭha computationally ? Given Pada-pāṭha of a Vedic hymn, can we generate the various Vikṛti-pāṭha-s ? Even using Vyākaraṇa, Śikṣā, Prātiśākhya-s etc. ?

2. Given a Vedic hymn or word (accented), can we get morphological, syntactic, semantic and contextual analysis with the help of the six Vedāṅga-s ? This would help objectively study various traditional commentaries etc. and work like a *`Logic Analyser`*, say, in philosophy.

3. There are symbolic meanings for all Vedic expressions (it is called **Trayī** because of *Ādhibhautika, Ādhidaivika* , and *Ādhyātmika* meanings). Can we obtain these details from a parser ?

4. Can Pramāṇa, Prameya Vibhāga, Śakti Graha (arising out of Vyākaraṇa, Upamāna, Kośa, Āpta-vākya, Vyavahāra, Vākya-śeṣa, Vivṛti and Siddha-pada-sānnidhya), Śābda-bodha Prakriyā etc. be simulated/ represented in a Computational system ? Will it work at all the three levels of denotation, *viz.* Abhidhā, Lakṣaṇā and Vyañjanā ? Is Inferencing possible ?

5. Can we automatically process text, say, by lemmatising, categorising or tagging using the parser for word analysis and classification ? Can it support lexicographic operations of creating wordlist, table of contents, word index, concordance, thesaurus, dictionary (vocabulary), spell-check etc. ?

Let us consider **certain cases** and see what gets involved.

1. Svasthā Bhavantu Kuravaḥ ? - Svasthā Bhavantu Kuravaḥ .

If we observe that the question and answer are identical, something has to be amiss. [*Accent* is the answer]. For the importance of this factor and further examples/details, ref. [1]. {Svasthāḥ = healthy, cheerful; Svaḥ-sthāḥ = residents of Heaven, *i.e* , dead}.

2. Mahāpadopasthitayaḥ Sadārāmāviyoginaḥ . Alaṅkārāśrayā Rāma Bhavanti Bhavatohitāḥ .. (Sāhitya Darpaṇam)

This verse speaks of the condition of those disposed favourably or otherwise towards Rāma.

Mahā-Pada-Upasthitayaḥ, Sad-Ārāma-A-viyoginaḥ, Alaṅkāra-Āśrayāḥ, Rāma, Bhavanti, Bhavataḥ, Hitāḥ

Trans. O Rāma ! Those who are favourably disposed towards you will occupy prominent positions, will not lose company of good gardenly surroundings and will be well-decorated.

Mahā-Āpadaḥ, Apa-sthitayaḥ, Sadā, Rāmā-Viyoginaḥ, Alam, Kārā-Āśrayāḥ, Rāma, Bhavanti, Bhavataḥ, Ahitāḥ .

Trans. O Rāma ! Those who are not favourably disposed towards you will have great adversities, will always lose company of wives and would eventually end up in prison.

3. Sovyādiṣṭabhujaṅgahāravalayastvāṃ Sarvadomādhavaḥ .

This is an invocation that can address either Lord Viṣṇu or Śiva. Saḥ, Avyāt, Iṣṭa-bhujaṅga-hāra-valayaḥ, Tvām, {Sarvadā, Umā-dhavaḥ} OR {Sarva-daḥ, Mādhavaḥ}.

4. Pātrākalitavedānāṃ Śāstramārgānusāriṇām . Tadetadarikālasya Karikālasya Śāsanam ..

From the stone inscription in an ancient South Indian temple. Describes a certain valiant ruler called Karikāla`s injunction. Pātṛ+Ākalita-Vedānām, Śāstṛ+A-Mārga-Anusāriṇām, Tat+Etat, Ari-Kālasya, Kari-Kālasya, Śāsanam, is the admissible word split, with Śāsanam being the Viśeṣya (qualified). The word-split, occurring at once, might be **:** Pātṛa-Ākalita-Vedānām, Śāstṛa-Mārga-Anusāriṇām, which is not correct.

5. Kṛṣṇaḥ Phalāni Khādati -

Here, Phalāni can be inflected from the Dhātu Phala Viśaraṇe/Niṣpattau in Loṭ Lakāra, Uttama Puruṣa, Ekavacanam and Khādati could be in locative case (Sati Saptamī).

6. Rāmaḥ Vanam Gacchati -

Here, Rāmaḥ can be inflected from the Dhātu Rā Dāne in Laṭ Lakāra, Uttama Puruṣa, Bahuvacanam, besides as a noun inflected from the stem Rāma, Vanam could be nominative or accusative and Gacchati could be in locative case (Sati Saptamī).

7. Mātrā Khādyate -

This sentence is in Passive Voice. [*lit.* `being eaten by mother` or `tablet is being eaten`]. Accordingly, this admits of two Active forms. `Mātā Khādati` (`Mother eats`) and `Mātrām Khādati` (`eats the tablet`).

Mātrā - Mātrāsparśāstu Kaunteya Śītoṣṇasukhaduḥkhadāḥ . (Bhagavadgītā, .)

Mātrā-sparśāḥ, Tu, Kaunteya, Śīta-uṣṇa-sukha-duḥkha-dāḥ. Here, Mātrā is inflected from the feminine stem Mātrā (trifle or part; also, syllabic instant).

Mātāraṃ Mātaraṃ Ca Lokānām . (Lakṣmīsahasram, 2). Here, Mātrā is inflectable either from the

feminine or masculine stem Mātṛ.

8. Rājā Janān Rakṣati -

This sentence is in Active Voice. In passive, the word Rājā, while inflected from the stem Rāj, being in instrumental case, becomes incompatible for Active Voice. If it is inflected from the stem Rājan, it`s in nominative case and fits the sentence and Rājñā Janāḥ Rakṣyante would be the Passive form.

9. Yajeta - Yajeta Svargakāmaḥ (Tai. Saṃ. ..)

Yajeta is from the root `Yaja`, which is Ubhaya-padī. Inflected in Vidhiliṅ, it is Prathama Puruṣa Ekavacanam in Ātmane Padī, while Madhyama Puruṣa Bahuvacanam in Parasmai Padī. The latter does not agree with the number in Svargakāmaḥ.

10. Naraḥ - Manuṣyā Mānuṣā Martyā Manujā Mānavā Narāḥ . Syuḥ Pumāṃsaḥ Pañcajanāḥ Puruṣāḥ Pūruṣā Naraḥ .. (Amarakośaḥ - Dvi.kā. Manuṣyavargaḥ, 1)

Naraḥ and Narāḥ are both inflected in nominative plural !. This is obvious if we observe all the other nouns in the verse, as they are all nominative plurals and hence, the catch is in the stem. Nṛ and Nara respectively.

11. Yatra Pārthodhanurdharaḥ . (Bhagavadgītā, 18.79)

This might take two splits as : Yatra Pārthaḥ Dhanur-dharaḥ and Yatra Pārthaḥ A-dhanur-dharaḥ. The latter is said to refer to Bhīma as the only Pārtha (*lit*. son of Pṛthā = Kuntī) who didn`t take a bow, by some interpreters.

12. I) Naḥ, - inflected from stems Asmad (acc., dat. and gen. plural) or Nṛ (Sambodhana Ekavacanam).

ii) Aśvaḥ - Noun and Verb - Stem Aśva and Dhātu Ṭu O Śvi Gati-vṛddhyoḥ, Laṅ Lakāraḥ, Madhyama Puruṣaḥ, Ekavacanam, meaning you have gone/grown

iii) Aruṇaḥ - Noun and Verb - Stem Aruṇa and Dhātu Rudhir Āvaraṇe, Laṅ Lakāraḥ, Madhyama Puruṣaḥ, Ekavacanam, meaning you have prevented

iv) Abibhaḥ - Tāṃ Saṃvatsaramabibhaḥ (Tai. Saṃ. 2.5.1.2) Dhātu Bhṛñ Bharaṇe, Laṅ Lakāraḥ, Prathama Puruṣaḥ, Ekavacanam

v) Vṛkṣe - Noun and Verb - Stem Vṛkṣa and Dhātu Vṛkṣa Varaṇe, Laṭ Lakāraḥ, Uttama Puruṣaḥ, Ekavacanam

vi) Parvataḥ - Noun and Verb - Stem Parvata and Dhātu Parva Pūraṇe, Laṭ Lakāraḥ, Prathama Puruṣaḥ, Dvivacanam.

13. Sakṛdeva Prapannāya Tavāsmīti Ca Yācate . (Rāmāyaṇam, Yu. Kā. 17..)

Here, Yācate is intended as a noun in Dative Case, to agree with the qualifier Prapannāya, though it also has an identification as a verb inflected from the root Ṭu Yācṛ Yācane.

14. Dṛṣṭe Bhavati Prabhavati Na Bhavati Kiṃ Bhavatiraskāraḥ ?

Here, the first Bhavati is intended as a noun (with stem Bhavat) in Locative Case, though it can also be a verb inflected from the root Bhū Sattāyām, like the second.

15. Nameyam - Nameyamuccairitarā Gatirna Me . (Lakṣmīsahasram)

This can be illustrated to be a benchmark case for a parser, as it admits of over a dozen different valid identifications including a single, two or three word combinations.

1. Nameyam - Dhātu Ṇama Prahvatve Śabde Ca, Vidhiliṅ Lakāraḥ, Uttama Puruṣaḥ, Ekavacanam, 2. Na, Meyam, 3. Nama + Iyam, 4. Na, Mā + Iyam, 5. Na, Me + Ayam, 6. Na, Me, Yam,

Here, Mā, Me can be inflected from stems Mā, (meaning goddess Lakṣmī) or Asmad, while Mā can also an indeclinable (privative particle).

16. Mātaḥ - Mātaḥ Parasmin Hṛdayaṃ Prasāṅkṣīt . (Lakṣmīsahasram)

Here, Mātaḥ can be a noun with stem Mātṛ (fem./mas.) in Vocative singular, a verb with root Mā Māne, Laṭ Lakāraḥ, Prathama Puruṣaḥ Dvivacanam or split as Mā+Ataḥ, where both Mā and Ataḥ could be indeclinable or Mā has stem Mā or Asmad and Ataḥ is indeclinable. In the verse referred, both feminine vocative of `mother` and the split Mā+Ataḥ, both being Avyaya-s are intended to strengthen the conviction through Anvaya and Vyatireka.

17. Rāmaḥ - Vane Carāmo Vasu Cāharāmo Nadīs Tarāmo Na Bhayaṃ Smarāmaḥ .

Here, Rāmaḥ can be split from the words Carāmaḥ, Āharāmaḥ and Smarāmaḥ. as Ca, Āha and Sma are valid words. Construction of sentences with Rāmaḥ as verb can be an exercise.

18. Viśvasya - Viśvasya Jananīmeva Viśvasya Śaraṇaṃ Vṛṇe . (Lakṣmīsahasram)

Here, Viśvasya can be a noun with the pronomial stem Viśva in gen. sing. or a participial gerund with Vi as preverb and Śvas as root (*lit*. having faith in). The former above is a noun while the latter is a gerund.

19. Marutām - Marutāmayeta Marutāmapi Sthalī . (Lakṣmīsahasram)

Here, the word Marutām is used twice in two different senses, once as a state of being desert, *i.e,* Maru-tā as stem (Maru = desert) and then as gen. plural of Marut-s, the wind Gods (stem is Marut). Even Heavens would be desertified (if your grace were not to be there).

20. Nābhijātaḥ - Nābhijātaḥ Sanneva Abhijātatamaḥ . (Hastigiri Māhātmyam)

Here, Brahman (mas.) is said to be born from the navel of Lord Viṣṇu. Nābhi-jātaḥ could also be split as Na + Abhijātaḥ. Abhijātaḥ refers to a high-born. Hence, the pun.

21. Bhajanti Vibudhā Mudhā Hyahaha Kukṣitaḥ Kukṣitaḥ . (Vairāgya Pañcakam, 3)

Bhajanti, Vibudhāḥ, Mudhā, Hi, Ahaha, Ku-kṣitaḥ, Kukṣi-taḥ . The word Ku-kṣitaḥ is inflected from the stem Ku-kṣit in acc. plural (serve bad kings). The word Kukṣi-taḥ is a Taddhita Avyaya composed of the stem Kukṣi with Tasil suffix (for the sake of stomach).

22. Na Kaścit Samaste Namaste Punaste Namaḥ . (Garuḍadaṇḍakam, 3)

Samaste can be a single word with stem as Samasta in loc. sing. or two words Samaḥ with stem Sama in nom. sing. and Te with stem Yuṣmad or Tad, or split as Sam and Aste.

23. Dhīmahi - is Uttama Puruṣaḥ Bahuvacanam - what are its other forms ?

This is an irregular form (Vedic usage) derived from three possible roots, *viz.* Dhyai Cintāyām, Dhīṅ Ādhāre, LaṄ Lakāra or *Ḍu Dhāñ Dhāraṇa Poṣaṇayoḥ* Dhāne Ca, *(Āśīr) Liṅ* Lakāraḥ, all with some or other form of Vyatyaya (Bāhulaka) - *Śablopa* , Vikaraṇalopa, Samprasāraṇa, Ātmanepada, Abhyāsalopa, Ākāralopa etc., ..., (Da)*dhīmahi Cf.* Manasi *Dadhīmahi* Satataṃ Mātāraṃ Mātaraṃ Ca Lokānām - (Lakṣmīsahasram, 2)

24. Kā Pūrvaḥ ? - Kāñcī Pūrṇaḥ .

Kā, Pūḥ, Vaḥ ? - Kāñcī, Pūḥ, Naḥ is the intended split, though Kā, Pūrvaḥ ? - Kāñcī, Pūrṇaḥ or

Kā, Apūrvaḥ ? Kāñcī-Pūrṇaḥ are also (incompatible) possibilities. Literally, the conversation means : Which is your (pl.) city ? Kāñcī is our city.

25. Alambusānāṃ Yātā

Alambusānāṃ, Yātā or Alam, Busānāṃ, Yātā are the two splits. In the first split, the destination is specified as Alambusā. In the latter, busa (*lit*. straw) is the noun stem.

26. Surabhimāṃsaṃ Bhakṣayati

Surabhi, māṃsaṃ, Bhakṣayati or Surabhi-māṃsaṃ, Bhakṣayati are the two splits. In the first split, the meat is qualified as Surabhi, *i.e,* fragrant. In the latter, it is compound word meaning cow-meat. (*lit*. eats fragrant meat (of any animal)/eats cow-meat).

27. Tad Vijijñāsasva . Vi Upasarga, Dhātu Jñā Avabodhane, Sannanta (Desiderative), Vidhiliṅ Lakāraḥ, Madhyama Puruṣaḥ, Ekavacanam.

28. Virāṭanagare Ramye Kīcakādupakīcakam .

Viḥ, Āṭa, Nagare, Ramye, Kīcakād, Upa-kīcakam . More readily occurring split as *Virāṭa-nagare* (in the city of Virāṭa) is incompatible for a verb is expected. Viḥ (a bird) Āṭa (wandered) supplies the missing information.

29. Rāmastu Sītayā Sārdhaṃ Sasnehaṃsavilāsayā .

Rāmaḥ, Tu, Sītayā, Sārdhaṃ, Sasne, Haṃsa-vilāsayā . The more easily occurring *Sasneham* is not admissible for want of verb.

30. Varā Varāharūpiṇī Carācarāntarasthitā Surāsurābhivanditā Dharādharādidevatā .
Sadā Sadāvalistutā Mudāmudāraśevadhiḥ Hitā Hi Tārkṣyaketanā Natānatāpatāṃ Nayet ..
Sadāvadātanimnagā Taṭīkuṭīravāsinī Kaṭisphuṭībhavatkarā Kiṭīśvarī Paṭīyasī .
Rasārasādihoditā Rasālasālasaṃvṛte Vanevaneṣu Dīkṣitā Madīkṣitādidevatā ..
This is an example for alliteration, prosody and double entendre etc. through innovative splits.

31. Vidhāvudeṣyatyapayāmi Bhāmini .. Udantam ... Nedantatvamajānata ..

Vidhau, Udeṣyati, Apayāmi, Bhāmini, Udantam, Na Idantatvam/Na Idam Tatvam, Ajānata . Here, the two meanings are : It-anta, *i.e,* Ikārānta, *i.e,* Vidhi Śabda, *i.e,* when luck clicks.

Idam Tatvam = this truth or fact, Udanta, *i.e,* Vārtā or Vṛttānta, OR Ut-anta, *i.e,* Ukārānta, *i.e,* Vidhu Śabda, *i.e,* when moon rises (by Candra-Udaya).

32. (Muhuḥ Pravṛttaṃ Navanīta-caurye Vatsān Vimuñcantam Adoha-kāle . Ulūkhale Kutracit Āttа-puṇye) Bandhum Satāṃ Bandhum Iyeṣa Mātā . (Yādavābhyudaya Mahākāvye)

Anvaya :- Satāṃ Bandhuṃ (= Bāndhavaṃ, Kṛṣṇaṃ) Mātā (Yaśodā) Bandhuṃ (= Bandhayituṃ Ulūkhale Iti Śeṣaḥ) Iyeṣa .

33. Gātuṃ Yajñāya . Gātuṃ Yajñapataye .

Here, Gātuṃ has been used as a noun in masc. acc. sing. Gātuṃ could be mistaken as an infinitive from the root Gā, to sing. Since this is a Vedic expression, the antodātta denotes the nominal base by Phiṭ Sūtra.

34. *Pātuṃ* Pātakino Janān A-śaraṇān *Pātuṃ* Tamaḥ-sāgaraṃ ..... (Viśvaguṇādarśa-campūḥ)

Here, both the roots Pā Rakṣaṇe and Pā Pāne are used with the suffix Tumun , respectively.

To decide this, we need to look at the Ākāṅkṣā (expectancy) and its fulfilment.

35. Śobhāṃ Vakṣyatyadhika-subhagāṃ Śobhamānāmatīndoḥ .. (Hamsa Sandeśam)

Both the roots Vaha Prāpaṇe and Vaca Paribhāṣaṇe have the same form Vakṣyati in Lṛṭ Lakāra future tense. Here it is Vaha Prāpaṇe. The next example also illustrates the same, in Veda-s.

36. Na Vo'bhāgāni Havyaṃ Vakṣyāma Iti . (Taittirīya Saṃhitā, 5.1.1)

Vakṣyāmaḥ has two possible identifications (both as verbs in future tense, first person, plural) with *Vaha* Prāpaṇe, *i.e,* 'to carry' and *Vaca* Paribhāṣaṇe, *i.e,* 'to speak'. But, as the object here is Havyam (sacrificial oblation), it can only be carried and not spoken of. Hence, by examination of Ākāṅkṣā, *i.e,* expectancy for a particular type of object, we can disambiguate.

Additionally, cases where context is to be considered like माङ् योग - as in मा भूत्, मा शुचः, etc, the split has to cover the previous word also. This may or may not involve Sandhi as illustrated. Certain special usages requiring semantic considerations like distinction between शूद्रा and शूद्री, पतिते पतौ where the form being ungrammatical in the sense of 'husband' has to be interpreted ingenuously, pose challenges to morphological analysis and parsing. Vedic examples are given in the Chapter 3.

Prerequisites

Processing Indian languages in Computers through Indian scripts, requires interfaces to keyboard (for data entry), display (for visualisation) and printers (for hard copies). It can also be enhanced to have automatic transliteration among scripts exploiting the commonality in the character set, if a proper scheme based on phonetics is worked out. Sanskrit processing can also use these features provided certain additional requirements are met. With these in view, Indian Standard Code for Infromation Interchange (ISCII) IS 13194:1991 has been evolved. Additional requirements for handling Vedic and Sanskrit-specific, phonetic-related aspects were provided for in Annex - G of this standard.

While attempting computational study of linguistic processes, 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). There is also a need for conceptual 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 "Padārtha Vibhāgaḥ".

Background

The current thesis covers about 9 years of research in the field of Sanskrit processing including the development of the Vedic extension, keyboard overlay, morphological, syntactic and semantic analysis modules etc. of a preliminary Natural Language Understanding (NLU) system for Sanskrit, DESIKA, at C-DAC. It also covers the computational rendering of Pāṇinian grammar rules, creation of databases of Ṛgveda Saṃhitā, Sāmaveda Saṃhitā etc. with all the Śāstra-ic treatises as an integrated system for analysis of Vedic texts. Annexure - B gives the details such as Excerpts of Annex - G of ISCII Standard, Vedic character font key map in Devanāgarī and Grantha, Samples of Vedic accented texts of Ṛgveda, Yajurveda - Saṃhitā and Brāhmaṇa and Sāmaveda.

PART II
Chapter 2 : Problem definition

2.1 The Problem Overview

Sanskrit is a language with inflexion of words. The words comprising a sentence of the language carry with them lot of information like person, number, case, tense etc. Extracting and organizing these information is the first step towards understanding the language.

[Pushpak, 86] describes an expert system for understanding Sanskrit. Its salient features are given below to illustrate the need for morphological analysis in an iterative manner. This system employs Constraint equations and integer programming solution for a possible implementation. In the accented input case also, after word level identifications, for syntactic and semantic processing, this approach may be utilised with the enhanced knowledgebase built up and consulting the ontology evolved.

The problem at this level of word analysis is that the analysis of a (Rūpa i.e word form) may generate multiple answers. It is the task of the word level program to give the exhaustive analysis of a Rūpa. At the sentence level, the problem is choosing the right information whenever there are multiple answers provided by the word level program. Another problem is that Sanskrit allows relatively free word ordering. Associating words to form meaningful groups, *i.e,* parsing, is important for understanding.

2.1.1 Subanta recognition

Given a Subanta pada the objective is to obtain : (a) Prātipadika ; (b) Vibhakti ; (c) Vacana ; (d) Gender ; (e) End type ; (f) The Paradigm Type information of the pada.

Paradigm Type indicates whether the Prātipadika may act as Viśeṣya, Viśeṣaṇa or both as Viśeṣya and Viśeṣaṇa or as Sarvanāma. Conventionally, Prātipadka-s in Sanskrit are classified according to their last letter and Liṅga. However, in a class of Prātipadika-s with the same last letter and Liṅga also, there may be variations due to certain suffix-specific morphological and/or phonetic processes resulting in different Rūpa-s.

The concept of Sanskrit Paradigm Type (SPT) is used with the view that Prātipadika-s with same Sanskrit type and same ending letter will undergo identical transformation to give Rūpa in the corresponding Vibhakti and Vacana. The transformation is composed of two steps :

1. Drop zero or more characters from the end of the Prātipadika string. (Let the number of characters dropped = n).
2. Add a string (say s) to the result of step 1 (s may be null).

Then the transformation operator can be a represented by an ordered pair (n, s).

Now the transformation operator (T_{op}) is a function of just four factors.

T_{op}= f(l, t, vi, va) where

- l = Last letter of Prātipadika
- t = Sanskrit paradigm type of Prātipadika
- vi = Vibhakti of Rūpa
- va =Vacana of Rūpa. [Liṅga appears here indirectly in type.]

Recognition of subanta pada may, therefore, be carried out if we :
1) Get all possible (n, s) operators after analysing various Rūpa-s and store them in a database.
2) Store all the Prātipadika-s of the language in a lexicon.

Thus, the database of operators are lexicon of Prātipadika-s and the essential suffix information structures for the program.

Subanta analysis is done scanning the Rūpa. At any point, the Rūpa string is divided into two parts. The left half is the possible Prātipadika part (PP) and the right half is a possible attached string AS. AS is searched in the database. If it is found, then PP is searched in the lexicon. If PP occurs as left half of a Prātipadika, then the last letter of Prātipadika (l1) and Sanskrit Type (t1) are obtained from the lexicon. The number of letters to be deleted (d1) is calculated from the length of PP and the length of Prātipadika in the lexicon. Then the database of AS is scanned for matching (d1,l1,t1) combination with a (d,l,t) combination in the lexicon. If a match is found, then the Rūpa is a valid one and the Vibhakti Vacana answer is given using the triple. All possible Prātipadika-s and Vibhakti, vacana-s are obtained because the above operation is carries out for every Prātipadika in the lexicon that has PP as left part.

The whole process is repeated by generating new PP and new AS. The algorithm terminates when 1) the Rūpa has been scanned completely *i.e,* PP = Rūpa and AS = null or
2) PP is not the left part of any Prātipadika in the lexicon.

Example: RŪPA = रामेण

PP's are र् रा राम् रामे रामेण्

AS's are आमेण मेण एण ण अ

एण is found in the database

राम् occurs as left substring of entry

राम 0 in lexicon (0 is the Sanskrit Type)

So, db1 = 1, l1 = A, t1 = 0. This matches with a database associated with एण.

So, Vibhakti = 3, Vacana = 1, Prātipadika = राम

The occurrence of PP as the left substring of a Prātipadika in the lexicon is looked for using binary search. For this, the lexicon is kept in sorted order (lexical). Once such a Prātipadika is found, we identify the whole range of Prātipadika-s for which PP is the left part.

This is done for two reasons :

1) The new PP formed always has old PP as left substring. So search for new PP can take place in a similar part of the lexicon, *viz.* the range of Prātipadika-s for old PP.

2) We are interested in all the Prātipadika-s for which PP is left substring.

This method ensures that whenever search for a PP fails, we need not scan the Rūpa any further.

2.1.2 Tiṅanta recognition:

Given a tiṅanta pada, the information provided here are : 1) Dhātu ; 2) Lakāra ; 3) Puruṣa ; 4) Vacana ; 5) Dhātu type, gives transitivity and other information required at sentence level.

The basic idea here is to recognise the DHĀTU-AṄGA dependent string (the string obtained by the transformation of the dhātu), the Lakāra specific string and the Tiṅanta pratyaya dependent string. Let us call them first operator (FO), second operator (SO), and third operator (TO) respectively. So, for the working of the program, we need to have a database storing the FOs, SOs and TOs. With each FO,

it is needed to keep the information of all possible SOs followed by the TOs for a particular FO-SO combination. The conjugational behaviour of most of the Dhātu's are like those of Dhātu-s. So for each Lakāra, it is required to construct TO strings of nine elements each (for three Puruṣa's x three Vacana-s, the position of each element gives the puruṣa and Vacana for which it is valid).

Example : For laṭ Lakāra, For TO string in parasmai padī, bhvādi gaṇa
अति अतः अन्ति असि अथः अथ आनि आवः आमः

Such lists of TOs (2 per Lakāra), the list of SO's and FOs are the important components of the database required for the program. The relationship among the three is established by storing the addresses of all possible SOs and TOs with each FO. A bit mask along with it indicates which element of the nine-information also are stored with each FO. Since each such bit mask, Lakāra, Dhātu address, SO address and TO address combination occupies six bits it is also called a hextuple. Again generating such an information structure is a voluminous and repetitive task. So, a Tiṅanta database generation program can be written to automate it. The program takes as input the Lakāra, Rūpas of the Lakāra, Dhātu and puruṣa, Vacana information. From these it constructs the data structure.

For Tiṅanta recognition, the Rūpa is scanned by moving a pointer over this characters, thus dividing it at any state into two halves (LH & RH); the left half (LH) is a possible candidate for FO, if the left half is found in the database, then for each hextuple associated with it we identify the second OPs. If a second OP occurs as such at the front of RH then it is extracted. Then using the bit mask the TO string is scanned. If the part left after FO, SO extraction matches an element of TO string and corresponding bit in the mask is 1 then the Rūpa is a valid one. So Lakāra, Dhātu and puruṣa, Vacana information are provided using the hextuple.

Example : RŪPA = भविष्यति

LH's = भ् भ भव् भवि etc. search for भव् succeeds

RH = इष्यति ; इष्य occurs as such in the list of SOs for भव् . So extract it.

Part left = अति

So, Puruṣa = 1 (3rd person); Vacana = 1 (singular number); Dhātu = भू ; Lakāra = laṭ.

To take care of upasarga, the above procedure has been enhanced. The analysis when upasarga-s are present is complicated, by the fact that there may be any number of upasarga-s and they may undergo sandhi. Moreover, the presence of upasarga often changes the meaning of Dhātu. So the essential information to be provided for upasarga analysis are:

1) Some rules of sandhi.
2) Upasarga and changed meaning pair with each Dhātu.
3) the list of valid upasarga-s.

Example - Rūpa : अन्वभवत्

Sandhi in Upasarga : अनु + अभवत्

Sandhi rule : इको यणचि

meaning change: The verb becomes transitive with the preverb.

The essential idea behind the procedure is to keep on collecting upasarga from Rūpa - if necessary by applying inverse sandhi - until the Rūpa is simply treateable by the aforementioned procedure. Since there may be multiple sandhi-s the procedure calls itself recursively for the right half

after breaking the sandhi. The Rūpa is scanned by advancing a pointer over its characters. The left half (LH) is a possible candidate for a upasarga. Attempt is made to break the sandhi if the string at the front of the right half (RH) happens to be one from among the collection or through appending the left character to LH after breaking sandhi; or the left character obtained after sandhi breaking my itself be an upasarga.

Example - Rūpa = प्रत्यागच्छत्

LH's = प्, प्र्, प्र, प्रत्,

till this point no valid upasarga either through character collection or through sandhi breaking.

प्रत्य् - Now य is sandhi character. On breaking and appending,

Upasarga = प्रति

New string on which recursion is applied = आगच्छत्

By breaking sandhi, new upasarga and rūpa are (आ + अगच्छत्)

New collection of upasarga (प्रति + आ), is therefore, simply treatable. The scanning continues but fails to produce any more valid upasarga-s either through character collection or through applying sandhi. So, finally we get the answer as : प्रति + आ + अगच्छत् .

2.1.3 Sentence level Program

After the word level program has extracted the information from words, it is the task of the sentense level program to solve problems like ambiguity, parsing etc.

(1) Word level information organization

It is necessary to collect the information provided by the word level appropriately so as to facilitate subsequent tasks. Each word is looked upon as a structure with one or more of each of subanta and tiṅanta structure. The subanta structure has various slots like Prātipadika, vibhakti, Vacana, Liṅga etc. Similatrly, tiṅanta structure has slots like Dhātu, Dhātu type, Lakāra, puruṣa, Vacana etc..

Prātipadika type and Dhātu type play important roles in sentence analysis. If the Prātipadika type indicates viśeṣaṇa, we know it is a demand word. There are certain Prātipadika-s which may act both as viśeṣya and viśeṣaṇa. This is an important information for analysis. Recognising pronoun is also necessary.

Dhātu type indicates if the Dhātu is transitive or not. Dhātu type also indicates the Vibhakti requirement of the Dhātu. For example, in an active voice sentence where main verb is from root गम् (gaml - to go), the presence of atleast one prathamā Vibhakti (1st case ending) word is essential. The presense of second (destination), 3rd (vehicle of requirement) and 5th (source) is also quite likely.

Avyaya-s and special words are also like subanta structure. However, for them a vibhakti of a special type is given which serves to indicate them.

After organising the word level information, the words are classified into two categories *viz.*, DEMAND WORDS and SOURCE WORDS. However, in case the assignment of the words into classes is not unique (e.g, for Prātipadika s which may act both as viśeṣya and viśeṣaṇa), each possible case is considered seperately and the procedure is repeated.

At any stage the DEMAND WORDS are those that require the presence of words with certain specific properties. In general, demand words can be grouped in four broad categories:

1) Tiṅanta words
2) Kṛdanta words
3) Karmapravacanīya words and special words and
4) Adjectives and 6th vibhakti words.

For the first two categories the demand depends on the sense of the Dhātu from which such words were derived. The demand of the third category is a purely syntactic requirement. For the last category, the demand arises from the fact that the adjective must qualify something while 6th Vibhakti needs a word with which it has SAMBANDHA (of whom /what) relation.

2.2 Steps Of Computation:

1) Whenever possible, group the words into unbreakable "chains". The words in a chain satisfy the following criteria".

a) All of them must have same liṅga, Vibhakti and Vacana.
b) They should be contiguous in the sentence.
c) At most one of them can be pure viśeṣya.

2) Set up constraint equations reflecting the demand of category-1 category-2 words.
$D_{[n][p1, p2,....]}$ = demand of nth word for words with property p^1, p^2 .. etc.
$X^{[i,j]\ [(n)]\ (p1,\ p2)]}$ .... = Assignment variable which is 'I' if the jth option of ith word/chain is assigned for demand $D^{[n]}$ [p^1, p^2....] otherwise it is 0.
Then the constraint equations due to demand words will be of the form

b) Similarly, set up constraint equations reflecting the fact that if a word/chain has multiple options, then at most one of them can satisfy the demand. These equations are of the following form :-
For each demand group *i*, for each of its mandatory Kāraka-s *k*, the following equalities must hold :

$$M_{i,k} : \quad \sum_{j} x_{i,k,j} = 1 ;$$

where j' s are options of the ith word. $M_{i,k}$ stands for the eqn. formed given a demand word i and k.
3) Setup cost functions of the form
$M_A,k_1 : x_A,k_{1,a} + x_{A,k1,c} = 1$ for mandatory Karakas k_1, k_2
$M_A,k_1 : x_A,k_{2,a} + x_A,k_{2,c} = 1$
where, X's are assignment variables and c' s are costs of such assignment. The cost is determined by the likely appropriateness of the assignment in view of the available information about word-type, word-position etc. (idea of आकाङ्क्षा, योग्यता, सन्निधि etc.)
Now the problem can be easily put into standard linear 0,1 integer programming of the form,
Minimize, $c_1x_1 + A_{12}x_2 + ... + A_{1n}x_n > b_1$
$A_{n1}x_1 + A_{n2}x_2 + ... + A_{nn}x_n > b_n$

4) Find the optimum solution.

5) For the words not assigned yet, set up constraint equations with type-4 words with appropriate cost function and again find the solution.

The following guidelines are used for assigning cost:
1) For kṛdanta immediately preceding word should have the least cost.
2) Nearer the source word to the demand word, lesser the cost.
3) In adjectives, first word is more likely to be the modifier.

2.3 Constraints

A parse is a sub-graph of the constraint graph containing all the nodes of the constraint graph and satisfying the following conditions
C1. For each of the mandatory kāraka-s in a kāraka chart for each demand group, there should be exactly one outgoing edge labelled by the kāraka from the demand group.
C2. For each of the desireable or optional kāraka-s in a kāraka chart for each demand group, there should be at most one outgoing edge labelled by the kāraka from the demand group.
C3. There should be exactly one incoming arc into each source group.

Integer Programming

A parse can be obtained from from the constraint graph using integer programming. A constraint graph is converted into an integer programming problem by introducing a variable x for an arc from node i to j labelled by kāraka k in the constraint graph such that for every arc there is variable. The variable take their values as 0 or 1. A parse is an assignment of 1 to those variables whose corresponding arcs are in the parse sub-graph, and 0 to those that are not. Equality and inequality constraints in the integer programmming problem can be obtained from the conditions (C1, C2 and C3) mentioned earlier.

Our Proposal

Parsing Sanskrit has to cater to various types of inputs like Vedic, Classical, Prose/text, poetry, Sūtra, polemic, narrative, conversation etc. each with speciality of its own in terms of features and characteristics. To achieve this, various factors like phonological, morphological, syntactic, semantic and pragmatic aspects are to be considered.

While western efforts for parsing of English-like languages have followed certain paradigms in AI and NLP, the requirements for Sanskrit Parsing are distinct due to the aforesaid features. Also, euphonic combinations (Sandhi), compounding (Samāsa), special modes denoted by specific suffixes, like Causative (Ṇijanta), Desiderative (Sananta), Intensive (Yaṅanta), Frequentative (Yaṅluganta), Reflexive (Karma-kartari), voices like active (Kartari), passive (Karmaṇi), Impersonal (Bhāve, usages like Parasmai Padī, Ātmane Padī and ditransitive (Dvikarmaka) roots, preverbs (Upasarga), Indeclinable particles like Avyaya, Nipāta, Karmapravacanīya, Gati etc. demand a different approach.

Current Research

Among contemporary researchers, Kak [2] points out the distinction as : The greatest sucess a constructive approach to the description of a natural language has ever had was when Dākṣiputra Pāṇini devised his grammar for the Sanskrit language, an achievement termed by the famous linguist Leonard Bloomfield [[2]] as "One of the greatest monuments of human intelligence". We would expect that the insights of Pāṇini, "The greatest linguist of antiquity, if not of all time" (Staal [[3]]) could be exploited to help answer questions regarding limits to computers as well as to define an approach that could yield

powerful text and speech processing systems. The knowledge representation methodology in the grammar of Pāṇini and his successors is in many wys equivalent to the more powerful, currently researched artificial intelligence (AI) schemes. Furthermore, it includes rules about rules, analogs of which are not known for any other language (or for a flexible enough subset of, say the English language), which would help in the writing of efficient AI software". He goes on to define the Pāṇinian approach as using :

1. Root and affix analysis.
2. Linear strings of rules and analysis by rule sequence.
3. Analysis by functional structure and
4. An exhaustive description.

Briggs [3,4] has noted the similarity between Semantic Netwrks in AI and Kanthan [5] has described the formal language system of Pāṇini. IIT, Kanpur team has used Pāṇinian framework for Machine Translation among Indian language-pairs in their Anusaaraka system [6-10]. However, our effort is for a Sanskrit parser, *per se* , and a general-purpose one at that. This includes accented inputs as well.

Approach of Sanskrit morphological analysis effort at Leiden University, Netherlands [13], involves a word-parser which produces preposition(s) --> stem --> infix(es) --> ending, prefix(es), augment, reduplication. It recognises patterns in context-free mode. Use of Pāṇinian rule encoding technique in SIMURG Project in Germany, on Pāṇini's grammar (auto-semantics) as cognitive knowledge structure [11], Navya-nyāya formalisation by a KRL [15] AV Hudli, Univ. of Nebraska, Pāṇinian database at Upenn [14], Two-level morpho-phonology of Sanskrit [12], Oslo University's Dr. Fosse's work on lemmatising/tagging Sanskrit texts with Prof. Peter Schreiner's e-text corpus and their statistical and other analyses, OCP, LEXA and such other software systems etc. have all been sources of inspiration.

2.4 Conceptual Basis

As stated earlier, various factors like phonological, morphological, syntactic, semantic and pragmatic aspects are to be considered here.

Phonological aspect includes accent recognition of input syllables in their Pada- Pāṭha or 'pausa' form, determining semantics related to accents, using them for sense-disambiguation when multiple possibilities exist, splitting euphonic combinations and dissolving compound words. These require rule-base derived from Vedic section of Pāṇini's Aśṭādhyāyi, various Prātiśākhya, Śikṣā, Varṇa-krama and Lakṣaṇa-grantha-s.

Morphological analysis requires the various Sandhi rules, at word, syllable, character, accent and phonetic levels from the works cited earlier. An algorithm derived for an exhaustive and comprehensive analysis for Sanskrit is the main emphasis of this report. This is designed at three levels. *Viz* . Plain single words, accented single words and plain and accented compound words.

The **syntactic** aspect requires Pāṇini's Kāraka-Vibhakti and Upapada-Vibhakti mapping rule-base, Nyāya-Śāstra's Ākāṅkṣā, Āsatti, Yogyatā and Mīmāṃsā principles, to determine word-senses, assign functional roles to constituent words of a sentence and extract/explain semantics of input. Various sentence types are characterised on parametric basis to group them on syntactic patterns for ease in extracting semantics.

Semantic aspects involve the communication (or expression) content or knowledge representation and evolving powerful ontologies with well-defined hierarchies. Padārtha-vibhāga of Nyāya Śāstra, Pada-śakti, Śakti-graha, Pada-padārtha Sambandha etc. expounded in Mīmāmsa Śāstra etc. are to be considered here. Discourses, passages of text, conversation, speech acts, narrations etc. are covered here as accent-related semantics are emphasised. The process of semantic extraction with word-meaning relation at the core is termed '**Śābda-bodha**'.

Pragmatics include world knowledge/common-sense issues to incude situation-specific factors to be taken into account where again, Laukika Nyāya-s derived from Mīmāmsā Nyāya-s and techniques for discourse coherence are applied. Vedic examples are also considered.

An illustration to highlight these levels is given in Fig. 2.1 through 2.5. Let us consider a sentence : "**Āpagāyāṃ Mahārājo Ghaṭāv Apūrayat**". ['In the river, the great king filled two jars'].

Figure 2.1 shows the parse-tree expected from a scheme that employs phonology, morphology, syntax and semantics integrally. Figures 2.2-2.5 are clearer versions for words 1 to 4 respectively. Initially, the individual words are separated by 'undoing' sandhi-s, if any, (as shown by the rule 8.3.23, for instance). A brief note on Pāṇinian rules and Kāraka Theory is given in Annexure - D.

Consider the first word ' Āpagāyāṃ'. Here, the Anusvāra becomes 'm' in 'pausa' form. Then its analysis yields the base-suffix combination for the locative singular. The rules associated are shown. Pada, Kāraka and Adhikaraṇa definitions are applicable here. The base-affix analysis proceeds further and Āpagā and yāṃ get further split into their constituents to extract the meaning added at every stage of the derivation. Finally, the roots Āpḷ and Gamḷ happen to be the roots for the word Āpagāyāṃ. This also takes accents into account as shown at different steps in Fig. 2.2. Thus, the example covers a Adhikaraṇa Kāraka (Locus), Samānādhikaraṇa (co-referential compound) in Samāsa, Karma Kāraka (object) and a verb.

The various grammatical rules are also illustrated to appreciate the advantage of a program to generate such 'parse trees'. A comparison with English deep structure tree is shown in Fig. 2.6, taking the case of the sentence 'Rāmaḥ Vālinaṃ Araṇye Bāṇena Jaghāna' ('Rāma killed Vālin in the forest with an arrow').

It may be noted that the Vedic syllable syntax and the way the inputs are stored is to be considered while 'exploding' words into LHS and RHS. Vowels and accent-markers are to be retained towards the same side, while accounting for accent-changes due to sandhi-splitting. These are covered in Chapter 4.

Even with respect to databases, lexicon should be accented and hence, Yāska's Nighaṇṭu is used for nominal bases. Verbal roots and all suffixes have accents specified by Pāṇinian grammar rules and these are invoked to split the sub-strings suitably.

We deal with morphological analysis in more detail with an exhaustive 'algorithm' presentation, description of a comprehensive paradigm analysis of nominal inflections (of archetypical suffixes), an all new Vedic processing extension and include detailed case suffix-relation mappings, syntactic check procedures, verbal activity classification and mapping to expected functional roles, creation of an ontology based semantic lexicon and conflict resolution methods from the three Śāstra-s.

Sanskrit word-types

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 qualifying) the base. Affix covers 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.

Words in Sanskrit are classified into two major types based on the affixes that get attached to the bases, as nominal class or Noun (literally those which end with सुप् 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 320 types of Noun based on the declensions and cover all nominal word forms including actival class or primary derivatives (कृदन्त), secondary derivatives (तद्धित) and compounds (समास). Annexure - A gives the details of paradigm type analysis of nominal stems in Sanskrit. The latter type decline to indicate various tenses/moods, voices and other aspects like causal, desiderative, frequentative, intensive, reflexive, abstract forms etc. alongwith person and number. There are about 2000 verbal roots in धातुपाठ.

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 a 100 types covering a list of about 700 words and procedures for deriving many more.

Word-type classification and databases

Word type classification and coding on parametric and declensional basis is useful in analysis of word senses to a great extent. The codes are devised for the ending, gender, declension paradigm type, case and number for Nouns; गण, पदी, इत्, कर्म, इट्, अन्त, स्वर 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. To begin with, a vocabulary of about 3000 words has been chosen, which represents all the possible word types, so that while growing the dictionary also it is done in a categorised manner. The धातुपाठ has been codified as described above and is the verbal root lexicon for the system. A list of the various types of अव्ययs with codes aforesaid is the lexicon for indeclinables.

The database is then prepared which, in the case of Nouns and verbs is in the form of list of suffixes for each distinct type of base declension (modified सुप् and तिङ् suffixes) with the code cross-reference which is unique for any base type.

Method of Analysis

The word level analysis then proceeds *'ab ultimo'* as follows :

The given word is 'exploded' into each of its component vowel and consonant parts. Accents are also separated and attached to respective vowels if the input is accented. Starting from the right end, by a process of iterative recombination, we produce sets of LHS and RHS strings. Matching the RHS part with the suffixes' database and the LHS part against the base database - lexicon for nouns and root list for verbs - and carrying out paradigm type specification match, we get valid identifications. The process is continued till the left end of the word.

For compound words, the chopping process at the second level would yield options wherein more than one nominal stem will be there in the LHS part etc. Thus, applying the characteristic conditions, we can determine the compounding information. All possible identifications which are grammatically valid are thus got for sense determination/ disambiguation. Such an exhaustive list of valid possibilities helps linguistic research.

The process is repeated for possibility of identification as noun(s), verb(s) and indeclinable(s) for all words. Also, check for preverbs is done starting with the left end, the necessary rules being invoked. By increasing the level of complexity, literary constructions and compound words can also be analysed.

SENTENCE STRUCTURE ACCORDING TO PĀṆINI

The sentence : Āpagāyām Mahārājo Ghaṭāv Apūrayat - "In the river the great king filled two jars"

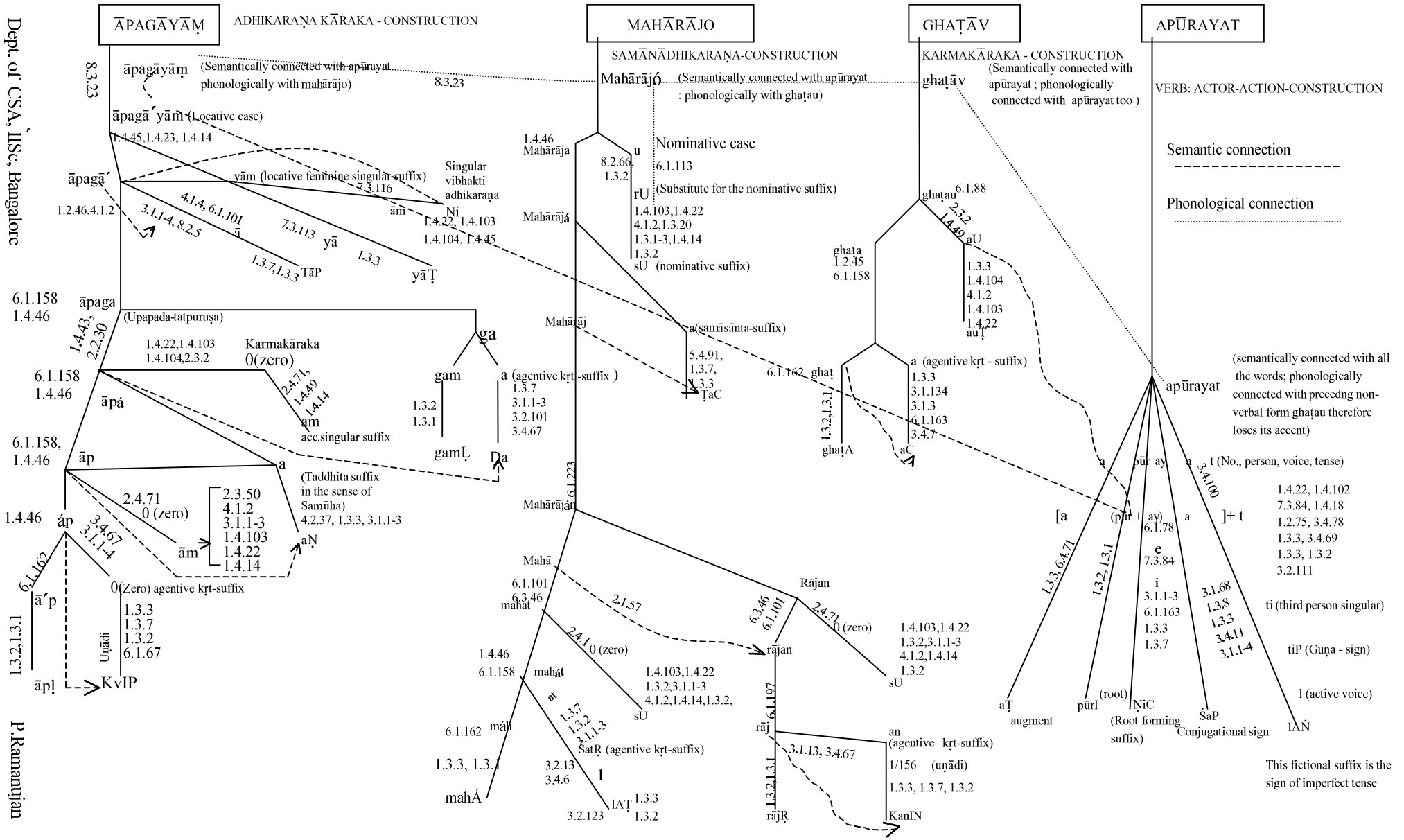


Figure 2.1

Dept. of CSA, IISc, Bangalore
P.Ramanujan

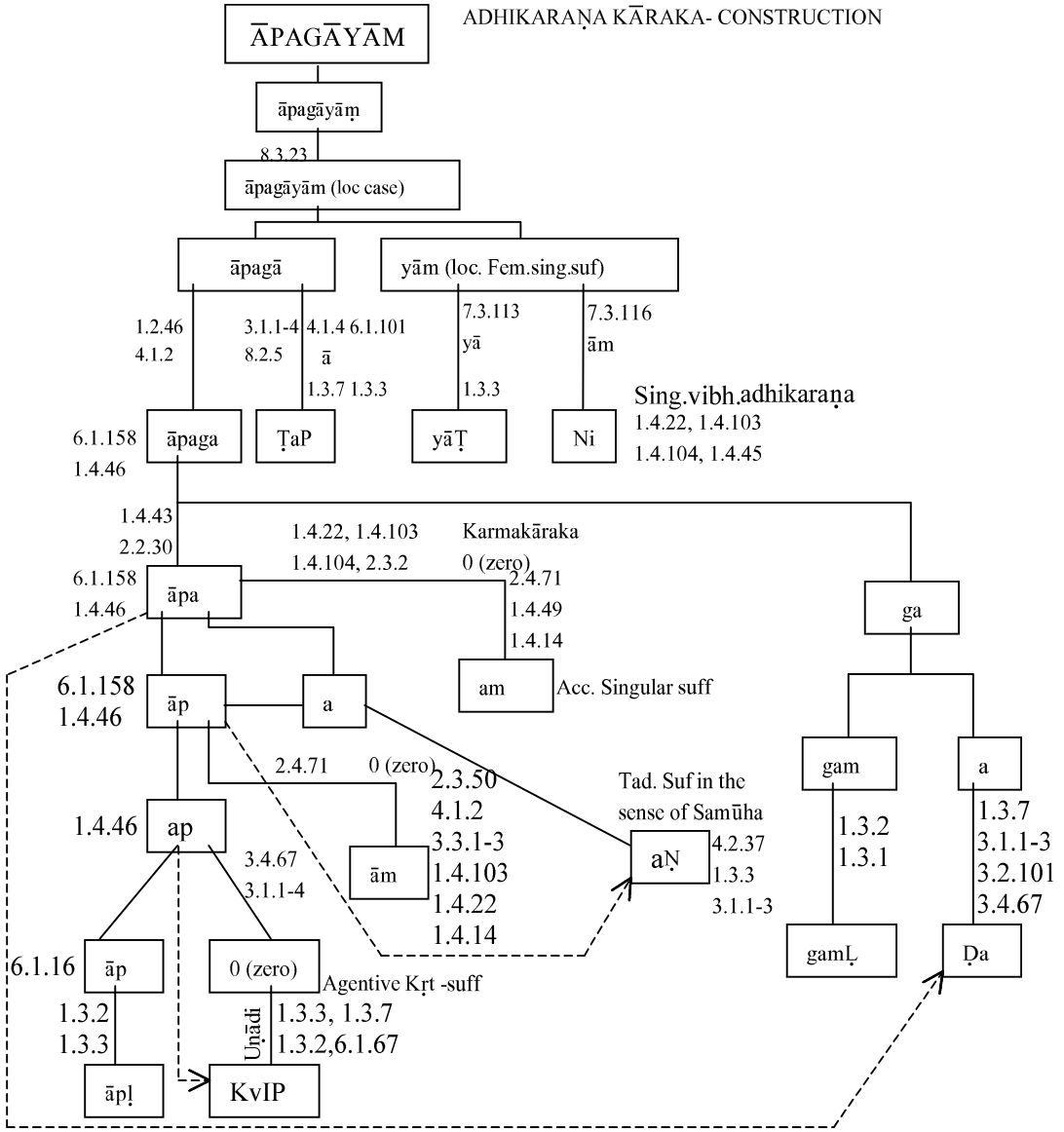

ĀPAGĀYĀM
ADHIKARAṆA KĀRAKA- CONSTRUCTION
āpagāyāṃ
8.3.23
āpagāyām (loc case)
āpagā
yām (loc. Fem.sing.suf)
1.2.46
4.1.2
3.1.1-4
8.2.5
4.1.4 6.1.101
ā
1.3.7 1.3.3
7.3.113
yā
1.3.3
7.3.116
ām
Sing.vibh.adhikaraṇa
1.4.22, 1.4.103
1.4.104, 1.4.45
6.1.158
1.4.46
āpaga
ṬaP
yāṬ
Ni
1.4.43
2.2.30
1.4.22, 1.4.103
1.4.104, 2.3.2
Karmakāraka
0 (zero)
6.1.158
1.4.46
āpa
2.4.71
1.4.49
1.4.14
ga
am
Acc. Singular suff
6.1.158
1.4.46
āp
a
2.4.71
0 (zero)
2.3.50
4.1.2
3.3.1-3
1.4.103
1.4.22
1.4.14
Tad. Suf in the
sense of Samūha
gam
a
1.4.46
ap
3.4.67
3.1.1-4
ām
aṆ
4.2.37
1.3.3
3.1.1-3
1.3.2
1.3.1
1.3.7
3.1.1-3
3.2.101
3.4.67
6.1.16
āp
0 (zero)
Agentive Kṛt -suff
gamḶ
Ḍa
1.3.2
1.3.3
Uṇādi
1.3.3, 1.3.7
1.3.2,6.1.67
āpḷ
KvIP

Figure 2.2

Parse tree showing word analysis- through Pāṇinian derivations with rulebase references

Sentence structure according to Pāṇini

The sentence Āpāgāyām ***Mahārājo*** Ghaṭāv Apūrayat " In the river the great king filled two jars"

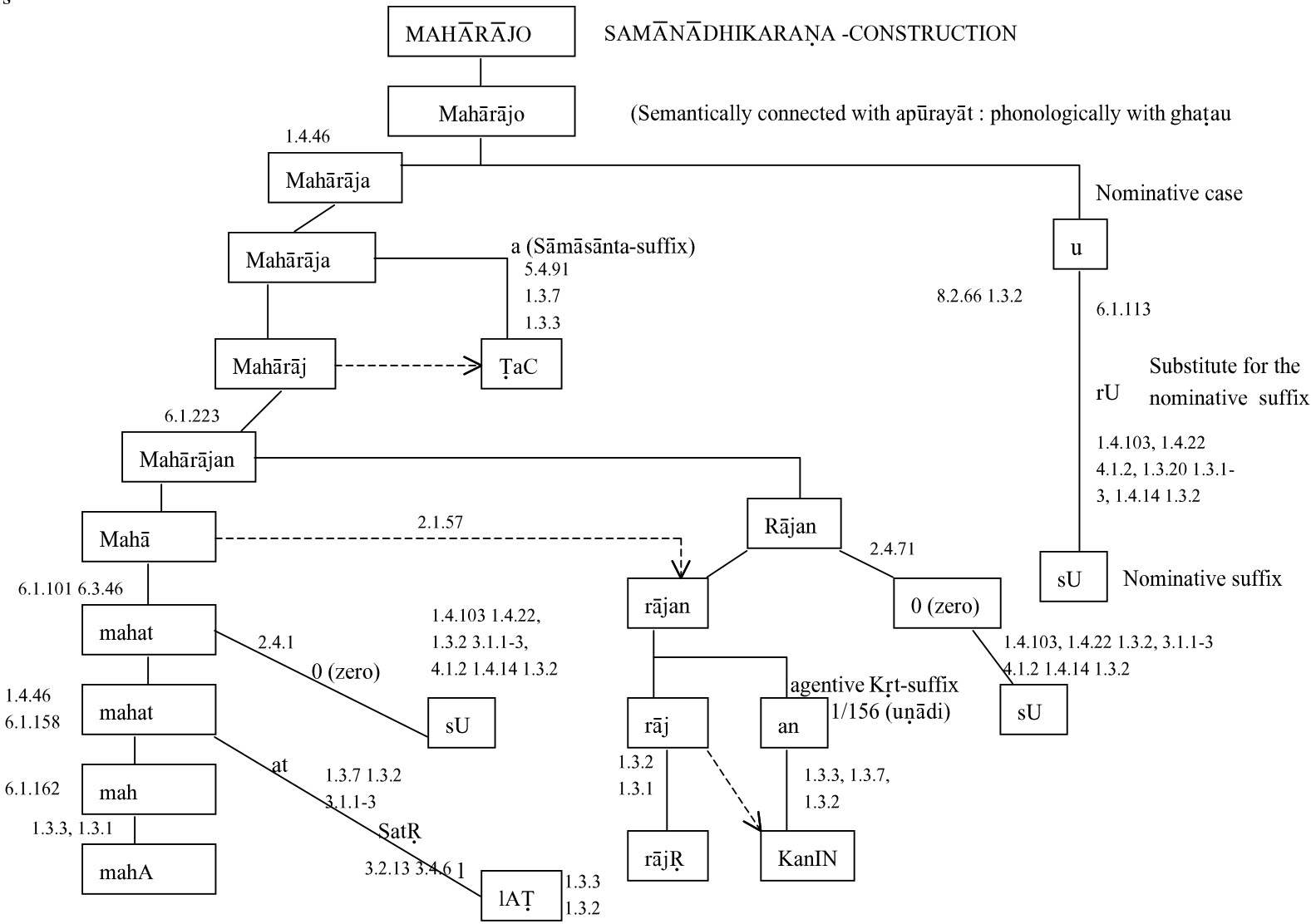


Figure 2.3

Parse tree showing word analysis- through Pāṇinian derivations with rulebase references

Sentence structure according to Pāṇini The sentence Āpāgāyām Mahārājo ***Ghaṭāv*** Apūrayat " In the river the great king filled two jars"

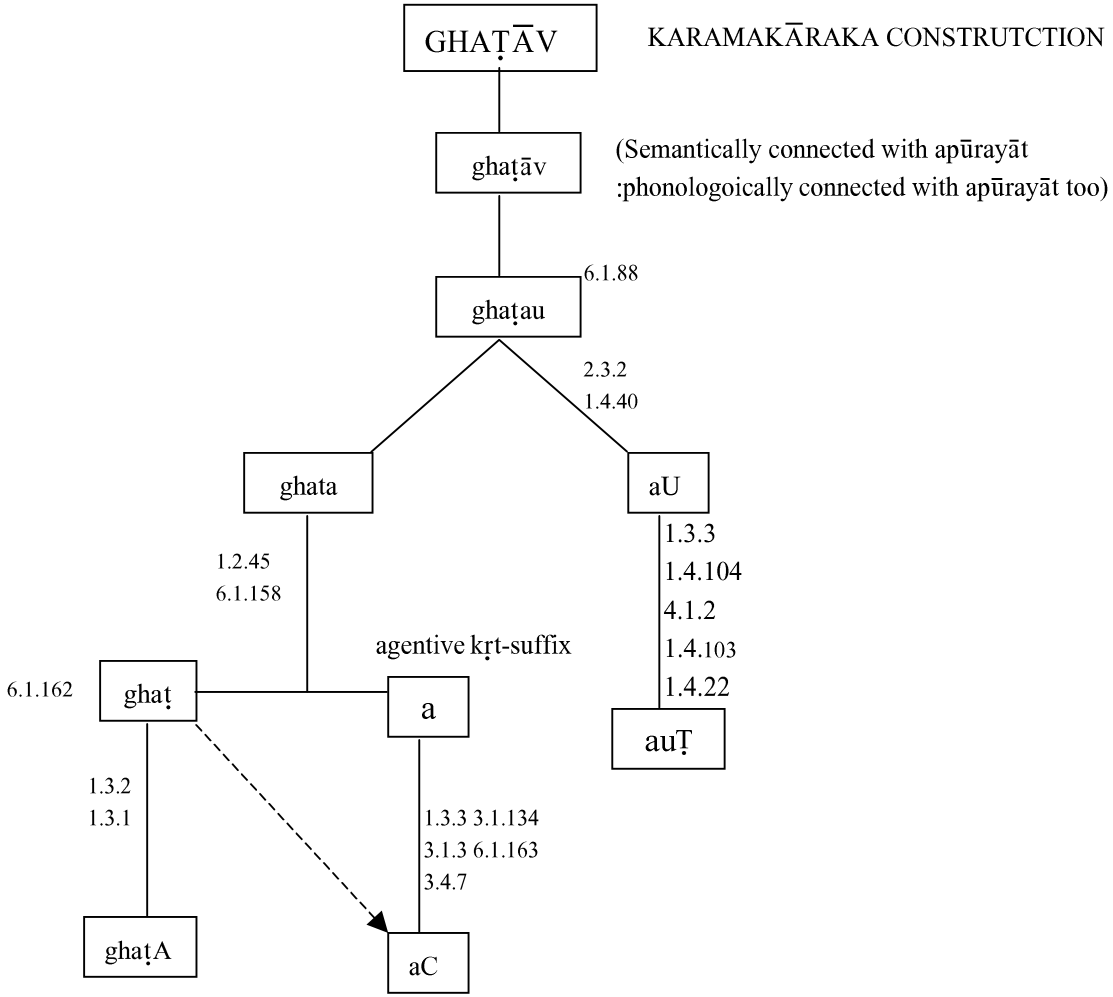


Figure 2.4

Parse tree showing word analysis- through Pāṇinian derivations with rulebase references

Sentence structure according to Pāṇini The sentence Āpāgāyām Mahārājo Ghaṭāv ***Apūrayat*** " In the river the great king filled two jars"

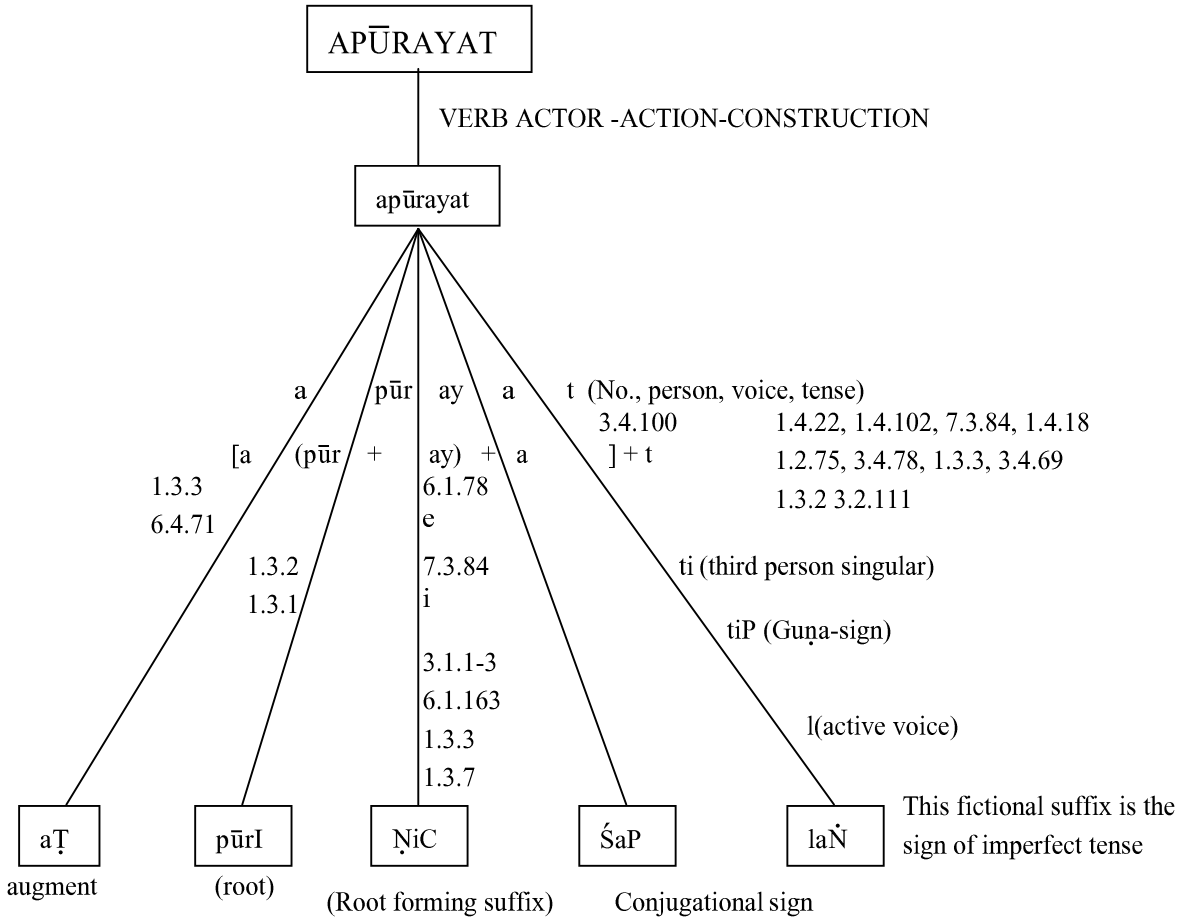
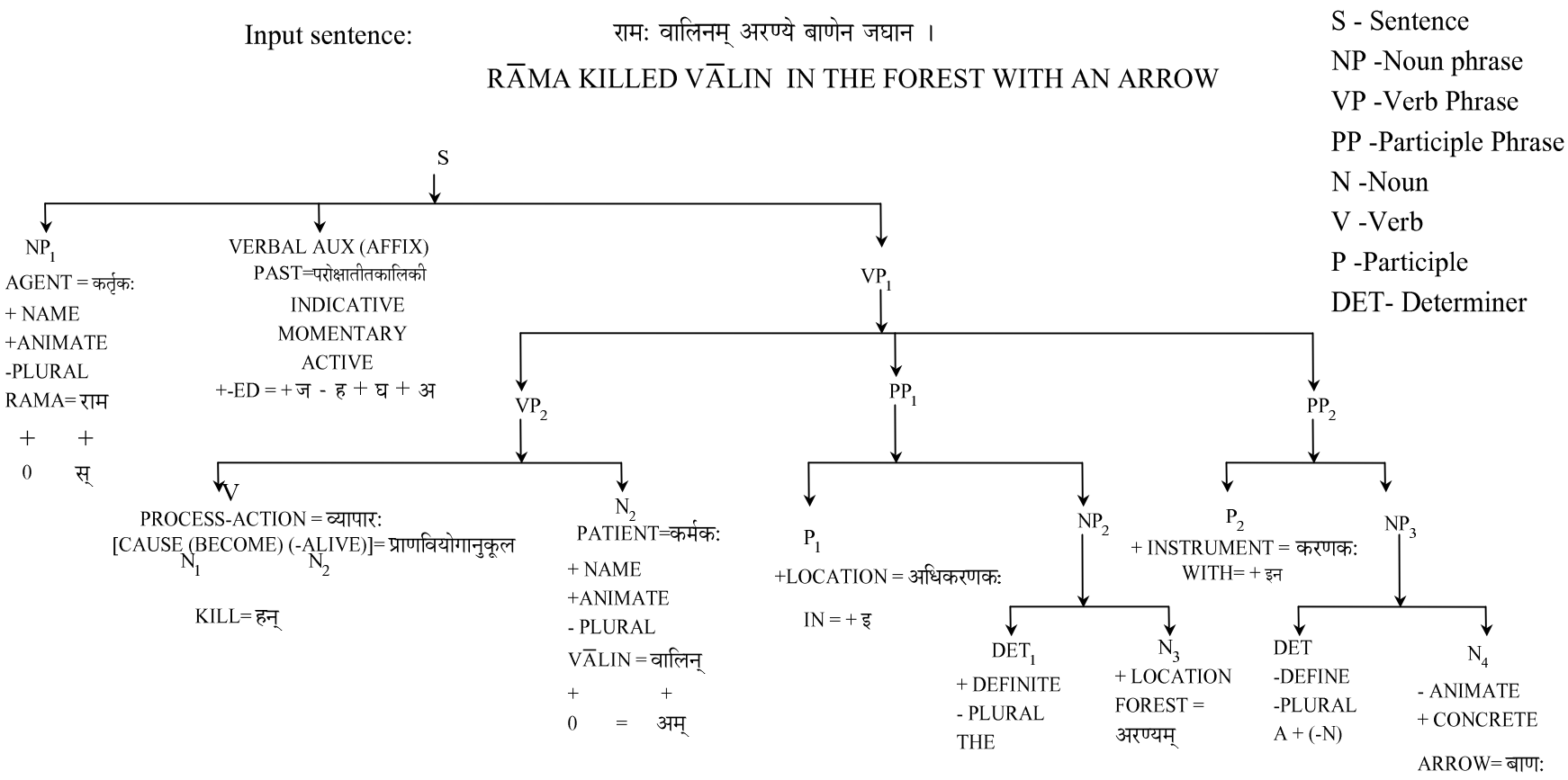


Figure 2.5

ŚĀBDABODHA AND DEEP STRUCTURE REPRESENTATION.



शाब्दबोधः - पुंस्त्वैकत्व -विशिष्ट-रामाभिन्नकर्तृक-पुंस्त्वैकत्व-विशिष्ट-वाल्यभिन्न-कर्मक-बाणकरणक-अरण्याधिकरणक-परोक्षातीतकालिकः प्राणवियोगानुकूलः व्यापारः ।

Representation of English Deep structure Tree and Sanskrit Śābdabodha .

Figure 2.6

PART III
Chapter 3 : Uniqueness of the Accented inputs

3.1 Distinction between Classical and Vedic Sanskrit

In Vedic Sanskrit, both the utterance and the content are sacrosanct, as मीमांसा शास्त्र establishes Veda-s to be eternal (नित्य) and authorless (अपौरुषेय), and hence infallible and unviolable. Also, the very utterance/recitation (पारायण) is of paramount importance without even the slightest variation in the intonation, for achieving lasting good. Dealing with Dharma and ensuring real benefits for all times for all living beings, Veda-s are the Universal source for righteous way of life and hence the extraordinary insistence on **'zero-distortion'** in its form and content. Also, over millennia, the Vedic texts have been propagated in an unbroken chain of oral tradition with elaborate, exhaustive and comprehensive system of devices for scientifically retaining their form and content. Thus, no other literary text shares the exalted place accorded to the Veda-s in Indian tradition.

Hence, Classical Sanskrit literature and the language, devoid of this feature of phonetic, phonological and prosodic intonation preservation in oral system and intonation marking scheme in the written variety, is only given the next place to Vedic Sanskrit. That written form of language fails to preserve a significant (and at times, valuable) portion of the content and context without such elaborate devices and conscious efforts is a telling factor in the rise of ambiguous and multiple/different semantic cognitions of different hearers of the same speech, in the natural language communication scenario.

Though requiring meticulous effort and rigorous discipline, Vedic texts remain sonorous, rhythmic and pleasing to hear and are quite effective in achieving various objectives. Sage पतञ्जलि in his महाभाष्य refers to याज्ञिक-s (sacrificial priests) being extra careful in Vedic utterances while not being so vigilant in colloquial usage of Sanskrit. He also expounds the virtues of Vedic text recitation and virtues obtained by reciters in detail, microscopically analysing the characteristics of Vedic texts.

3.2 Vedic Characters and Accent-markers

Based on pronunciation requirements, physiological constraints, euphonic combinations etc., besides the characteristics of sounds with respect to their place of origin, path followed, places of reflection, internal and external effort involved in their production etc., certain additional Vedic characters are in use in Vedic texts. Prolated duration of vowels (*supra* long) of different varieties (रङ्ग दीर्घ, प्लुत, रङ्गप्लुत, सार्धद्विमात्र, सार्धत्रिमात्र, चतुर्मात्र etc. are some of the characters found only in Vedic usage. Redoubled consonants called 'यम's, two distinct varieties of एकार, वकार, हकार etc. and many different forms of Vedic अनुस्वारs in different Vedic branches (Śākhā-s) like यजुर्वेद and सामवेद, varieties of विसर्ग, the trembling of the voice in specific situations called 'कम्प' of the ह्रस्व and दीर्घ and उदात्त and स्वरित variety, the character called 'स्वरभक्ति' with vowel and consonant-like properties, non-sandhi or 'विरळ' utterance, special elision and

non-elision of literals in euphonic combinations not covered in पाणिनि's अष्टाध्यायी (grammar treatise) etc. make up the additional repertoire of Vedic character set, over and above the Classical Sanskrit set (which itself has many additions over देवनागरी of the popular (Hindi) variety (like ॠकार, ऌकार, अवग्रह, half-visarga - जिह्वामूलीय and उपध्मानीय, various augments, substitutions/replacements and elisions and modifications of consonants, particularly). Vedic sandhis are also distinct in their two-fold nature of character and accent relation (वर्णसन्धि and स्वरसन्धि).

3.3 Specialities of Vedic Sanskrit

The various phonetic processes in the Veda-s include, character set actually used in particular Śākhā-s, exceptions to Pragṛhya-s, doubling, elision, elongation, augments, modification, cerebralization, aspiration, euphonic combinations, distinction and unitization (Anaikya and Aikya), sub-types of the accents (Anudātta as Dīrgha, Vikrama etc., Svarita as Nitya, Kṣaipra, Praśliṣṭa, Prātihata, Pādavṛtta, Tairovyañjana, Abhinihata etc.), hand-signs for accents, Kevala and Miśra Anusvāra-s of different kinds in different śākhā-s, Kampa, Raṅga-pluta, Svara-bhakti and its types and description of the various Varṇa-krama parameters.

In the Veda-s, there are three lengths for a vowel. These are short, long and extra long (hrasva, dīrgha, pluta (and Raṅga Pluta). The short and long vowels are denoted by the normal vowel signs used in Devanāgarī, while the extra long vowel is indicated by putting a 3 or 4 sign after a short or long vowel sign. For example: ओ३ङ्क्रतौ स्मर.

Vedic Accents

Following the rule of Pāṇini in the formation of a word from its rudimentary elements, the vowels acquire one of three basic pitch accents or svara:

(a) udātta, raised pitch [1.2.29],
(b) anudātta, not raised [1.1.30],
(c) svarita, a blend of the first two [1.2.31].

The pronunciation of the svarita is initially udātta for the period of half a short vowel and anudātta for the rest (*i.e.* , one and a half measures for a long vowel) [1.2.32].

In continuous speech or saṃhitā, all anudātta syllables following a svarita are called pracaya and are pronounced ekaśṛuti (monotone, between anudātta and udātta) [1.2.39], however, the anudātta immediately preceding a svarita or udātta is pronounced sannatara (lower than anudātta) [1.2.40]. The Veda-s may also be recited entirely in ekaśṛuti (with exceptions) [1.2.36ff.].

Nonetheless, the accent may affect the meaning of a word, particularly a compound [6.1.223, 6.2.1] (*e.g.* **sú-kṛta**, *indeclinable* . well done; **su-kṛtá**, *noun,* a good deed). With rare exceptions, a word standing alone has at most one vowel accented udātta or svarita, the rest being anudātta [6.1.158].

Udātta - The vowel that is percieved as having a high tone is called Udātta, or acutely accented. It is normally not marked. " ‿ " is used in śukla Yajurveda texts, at the end of a sentence.

Anudātta - The vowel that is perceived as having a low tone is called Anudātta or gravely accented. In writing, it is marked by a line underneath the vowel. it is also denotes Udātta in śatapatha brāhmaṇa. In Kāṭhaka text, Anudātta is shown as vertical line below the charatcer ˌ .

[Śatapatha Brāhmaṇa uses only two accents, udātta and anudātta : a horizontal line below the syllable indicates anudātta (as in the ṚgVeda) The relationship to the grammatical accents described by Pāṇini is rather indirect: the anudātta marked here are the syllables before those which would be marked svarita in the Ṛgveda.]

Svarita ˈ - The vowel that is a combination of Udātta and anudātta tones is said to be Svarita (स्वरित) or circumflexily accented. It is pronounced by combined raising and falling of the voice. It is marked by a vertical line above a character. This definition applies in ṛgveda. However, the same sign is defined as Udātta in Yajurveda's Maitrāyaṇīya text.

Long Svarita ˈˈ - Long Svarita is denoted by two vertical lines above a character", in some taittirīya texts.

Kampa ˈ‿ - Kampa is vibration in voice while reciting. The Kampa arises only when a Svarita is followed immediately by an उदात्त (or स्वरित). In such situations, a low pitched Mātrā is inserted between them. The recitation sequence consists of high pitch of स्वरित, followed by low pitch of कम्प, and then the high pitch of उदात्त or स्वरित, which causes the vibration. The मात्रा which is inserted between the स्वरित and उदात्त (or स्वरित) is indicated by figure १ if the svarita is short and by figure ३ if it is long. The figures are marked as both स्वरित and अनुदात्त, i.e, ˈ१̱ and ˈ३̱ .

Jihvāmūlīya (ᳵ) - This is like a half visarga sound, and can be come only before four consonants. before क and ख it is called Jihvāmūlīya, while before प and फ it is called upadhmānīya.

Example: अग्निः + खादति = अग्निᳵखादति ; रामः + पश्यति = रामᳵपश्यति

Puṣpikā (ᳶ) - This symbol is just a substitute for the spaces between words, and hence is not needed.

Yajurvedic अनुस्वार - अनुस्वार followed by non-Varga characters is pronounced with a ग् element mixed with उ vowel. This is also known as गुंकार.

ᳶ is used in śukla yajurveda.

ᳵ is used in Kṛṣṇa yajurveda. This is also used in śukla Yajurveda.

ᳶ is an elongated Kṛṣṇa Yajurvedic Anusvāra.

ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ - are the variants of Yajurvedic Anusvāra-s found in different editions and seem redundant.

Visarga - There are many variant is the normal visarga (ः) found in the vedic texts like ः (after उदात्त) ः (after स्वरित) ः , ः and ः . All of these seems to be redundant.

Others - ∫ - अथर्ववेदिक जात्य स्वरित is attached after a character. In some texts, it is shown preceding a Visarga. But it is better to show it after Visarga, since it is the logical order.

Svara-s in Sāmaveda are marked by 1, 2, 3, क , र and उ placed above the characters. These Svara-s, when required, can be placed in the corresponding positions of the previous row. As such, no special symbols are needed for them. Udātta, Svarita and Anudātta are marked with the numerals 1, 2, and 3 respectively; the figure 2 is also used to indicate an udātta when there is no following svarita; ekaśruti syllables are not marked. Special marking is used for particular accent sequences;

U - U -S => 1- marked - २र

U - U - A => @u - unmarked - ३

A - Si => ३क- २र [See Annexure - B.]

3.4 Accented word-types

There are nine types of accented words as (ending in Udātta) Antodātta, (beginning with Udātta) Ādyudātta, Anudātta, Nīca-svarita (Svarita preceded by Anudātta), Madhyodātta (middle Udātta), Svarita, words with two Udātta-s and three Udātta-s (Pāṇini Śikṣā. 45).

Their examples are : अग्निः, सोमः, प्र, वः, वीर्यम्, हविषाम्, स्वः, बृहस्पतिः, इन्द्राबृहस्पती respectively. (Pāṇini Śikṣā. 46-47). Places of accents are Hṛdaya (heart) for Anudātta, Mūrdhā (head) for Udātta, Karṇamūla (root of the ear) for Svarita and Sarvāsya (entire mouth) for Pracaya (Pāṇini Śikṣā. 48).

A difference in accent accounts for a difference of meaning. As an illustration, the word **'yaśas'** is neutral gender and means 'fame' when *ya* is acute, while the same word means 'famous' and becomes masculine when *śa* is acute.

Examples are : द्युम्नितमं यशः (ऋ. सं. ८.१९.६);

यशसं वीरवत्तमम् (ऋ. सं. १.१.३);

त्वमिन्द्र यशा असि (ऋ. सं. ८.९०.५); etc.

Similarly, the word '**brahman**' is neutral and means sacred utterance, the supreme being, true knowledge or penance (*Cf*. Amarakośa 3.3.114 - vedastattvaṃ tapo *brahma* , *brahmā* viprāḥ prajāpatiḥ) when *bra* is acute,

while it is masculine and means priest when *hma* is acute.

For example, **ब्रह्म ब्रह्मा**ऽभवत्स्वयम् (तै. का. ३.९.३).

Other examples of words identical in form but distinguished by accent could be :

1. **ते** - *they*, *those* (nom. pl. of the base तद्), as in **ते ते** पिबन्तु जिह्वया (ऋ. सं. १.१४.८) and **ते** - *to thee* (dat. sg. of the base युष्मद्), as illustrated in ये **ते** पन्थाः सवितः पूर्व्यासः (ऋ. सं. १.३५.११);

2. **दासः** - *servant*, *slave* (substantive), as in स्वयं **दास** उरो अंसावपि ग्ध (ऋ. सं. १.१५८.५), while **दासः** *servile* (adjective), illustrated in यो **दासं** वर्णमधरं गुहाकः (ऋ. सं. २.१२.४);

3. **धर्मन्** - neuter, meaning *ordinance, (solemnity, stay)* as in दिवो **धर्मन्** धरुणे सेदुषो नॄन् (ऋ. सं. ५.१५.२) as against **धर्मन्** - masc. *ordainer, upholder*, as in **धर्मा** भुवद्वृजन्यस्य राजा (ऋ. सं. ९.९७.२३);

4. **दामन्** - neuter, *cord, rope*, as in यद्वाजिनो **दाम** सन्दानमर्वतः (ऋ. सं. १.१६२.८) as against **दामन्** - masc. *bestower,* as in रायो **दामा** मतीनाम् (ऋ. सं. ६.४४.२);

5. **द्वादश** - *twelve*, illustrated by वेद मासो धृतव्रतो **द्वादश** प्रजावतः (ऋ. सं. १.२५.८) as against **द्वादुश** - *twelfth* as in देवहितिं जुगुपुर्द्वादुशस्य (ऋ. सं. ७.१०३.९);

6. **भूमन्** - neuter, *earth*, as in चर्मोवोदभिर्व्युन्दन्ति **भूम** (ऋ. सं. १.८५.५) as against **भूमन्** - masc. *abundance* as in सौधन्वनासश्चरितस्य **भूमनाऽऽगच्छत** ; अपां **भूमानमुप** नः सृजेह (ऋ. सं. १.११०.२, १०.९८.१२);

7. **युवभ्याम्** - *to you two* as in अरमिन्द्राग्नी मनसे **युवभ्याम्** (ऋ. सं. १.१०८.२) as against **युवभ्याम्** - *to the two youths*, from the stem *yuvan*, as in नमो युवभ्यो [illegible] ३.३१.७);

8. **रक्षः** - neuter, *demonhood*, *(injury ?)*, as in इन्द्राग्नी **रक्ष** उब्जतम् (ऋ. सं. १.२१.५) as against **रक्षः** - masc. *demon*, as in अपसेध**न्रक्षसो** यातुधानान् ; यो वा **रक्षाः** शुचिरस्मीत्याह (ऋ. सं. १.३५.१०, ७.१०४.१६);

9. **वरिमन्** - *neuter*, *expanse* as in मितज्ञवो **वरिमन्ना** पृथिव्याः (ऋ. सं. ३.५९.३) as against **वरिमन्** - *masc. expanse*, as in दिवश्चिदस्य **वरिमा** वि पप्रथे (ऋ. सं. १.५५.१) etc.

Accents as word-category denoters

It is noteworthy here that words denote different class of objects when accented differently, i.e, accents have a link to the objects referred to by them. As a generalisation, we see words ending in Udātta are substantives (*Cf*. Phiṭ Sūtra of *Śāntanava* - phiṣo'nta udāttaḥ), while those with initial accent are adjectives, as these amount to an implied Bahuvrīhi compound interpretation, i.e, 'one or that which has (*word stem meaning*)'. This point deserves deeper consideration as it has a good potential in guiding speech recognition, in terms of supplying 'expected' categories of word and lexical category search.

Compound words in Sanskrit have their meanings derived using the grammar rules wherein the constituent nominal stem meanings, suffix meanings and the compounding (or sambandha) meanings

are involved. While a same surface form of a derived word could have two or more different processes of derivation with their different semantics, the determination of the appropriate meaning in a given context may be possible by considering the accents.

Illustrations of meaning change due to accent-change

To illustrate, the word **'indraśatruḥ'** which is a compound word, has two constituents, viz., Indra and Śatruḥ, where śatru means a killer. In the case of *bahuvrīhi* compounding इन्द्रशत्रुः, the first word retains its original accent and hence *i* is acute and the word would mean 'one whose killer is Indra'. Alternately, if the second word is accented, i.e, *tru* is acute इन्द्रशत्रुः (this compounding is called *tatpuruṣa*), it would mean 'Indra's killer'. A Vedic episode explains how a mistake committed in reciting this word in a sacrifice by **tvaṣṭā** resulted in a counter-productive consequence यदब्रवीथ्स्वाहेन्द्रशत्रुर्वर्धस्वेति तस्मादस्येन्द्रः शत्रुरभवत् (तै. सं. २.५.२.१).

A similar example is quoted by sage Patañjali in Mahābhāṣya regarding the word **'sthūlapṛṣatī'** appearing as an adjective of a sacrificial cow स्थूलपृषतीमाग्निवारुणीमनड्वाहीमालभेत, which, when resolved as Tatpuruṣa, स्थूलपृषती (sthūlā ca sā pṛṣadvatī ca), would have *ṣa* acute and mean 'big (fat) and spotted', while as a Bahuvrīhi compound स्थूलपृषती (तै. सं. ५.६.१२) (sthūlāni pṛṣanti yasyāḥ sā) - as opposed to क्षुद्रपृषती (तै. सं. ५.६.१२), *sthū* would be acute and mean 'one whose spots are big'. Obviously, one has to be precise in choosing the proper type of the cow (cow is big and *has* spots or *cow* with big spots).

Range of accent meanings

In general, Pāṇini has referred to various aspects of speech like :
vidhi (optative),
anujñaiṣaṇā (seeking permission),
āmantraṇa (giving free choice),
nimantraṇa (invitation for a particular thing),
prārthanā (request), *āśiṣ* (benediction),
atisarga (seeking permission for a free choice),
praiṣa (order),
kāmapravedana (expressing one's purpose),
adhiṣṭa (polite),
sambāvanā (probability),
anavakḷpti (impossibility),
kriyāprabandha (continuity),
kriyātipatti (condition),
kriyāsamabihāra (iterative),
samuccaya (simultaneity),
samprašna (ascertaining by question),
hetumat (causative),

hetuhetumat (cause-effect),
prāptakāla (occasion), *icchā* (desire),
āśaṃsā (hope),
abhijñāvacana (recollection),
garhā and *bhāvagarhā* (derogatory and that affecting action only),
pṛṣṭaprativacana (reply to a question),
parīpsā (hurry),
amarṣa (bad feeling, ire),
citrīkaraṇa (wonder),
prahāsa (ridicule) etc. while prescribing verbal tense/mood markers and describing sentence accents etc. Thus, a same word to convey different aspect from among these is accented (stressed) variously and thus, the accents are helpful in determining the speaker's intentions. (*Cf*. *Bhartṛhari* in *Vākyapadīya* on word-meaning relation, .. kālo vyaktiḥ **svarā**dayaḥ ...).

These are the key link to research further to establish a correspondence between spoken and written (keyed-in) forms of Sanskrit language in NLP.

Chapter 4 : Morphological Analysis

4.1 Case of Accented Inputs

There is a need for morphological analyser to apply Accent-Sandhi (from Aṣṭādhyāyī - Vaidika Prakriyā, Śikṣā-s and Prātiśākhya-s etc.) rules and provide valid combinations of all possible nuclei and suffixes with their respective accents also considered. Many processes like compounding, elongation, cerebralisation etc. applicable due to specific accents are to be provided for. This could help in getting Pada-pāṭha from Saṃhitā and *vice versa*.

Pāṇinian rules dealing with Accent-Sandhi-s like 6.3.114-139, 8.4.66-67 etc. and specific Śikṣā-s and Prātiśākhya-s etc. are to be utilised to split input strings into two parts when accented inputs are encountered. For example, the following splits should be possible :

माऽघशꣳसः = मा + अघशꣳसः (तै. सं. १.१.१)

नर्क्षीकाः = न + ऋक्षीकाः (तै. ब्रा. ३.९.१)

वृत्रहाऽसि = वृत्रहा + असि OR वृत्रहा + असि (तै. सं. २.४.१४)

उदगासीत् = उदक् + आसीत् OR उत् + अगासीत् (तै. ब्रा. ३.८.२१)

There are also several Vikṛti-pāṭha-s (combinatorial patterns) of Vedic hymns which help in distortion-free preservation in oral tradition of knowledge-propagation. Their study helps get a good command over accents of individual words and there are many works dealing with their subtleties and nuances (in phonetic utterance). For example, certain special character/accent sandhi-s cannot be known at all unless the Krama, Jaṭā or Ghana Pāṭha-s are learnt. We illustrate the importance of some of these features, to justify the effort undertaken to enhance the ISCII (IS 13194:1991) Standard for Vedic extension overlay, as found in its Annex - G.

4.2 Accent determination using rules

Routines to identify the proper accents of words from ISCII symbols of input are developed and word-splits including these are done. Determination of the particular type of accent, e.g, among the seven (or eight) sub-types of Svarita (circumflex accent), distinction between Udātta and Pracaya using special rules, As an illustration, in .. यावन्तो गृह्याः स्मस्तेभ्यः .. (तै. सं. १.८.६.१), because of the presence of the pronomial यावन्तः, by यद्वृत्तात् ... (८.१.६६) स्मः is Udātta and NOT Pracaya, even though it is preceded by a Nitya Svarita, *i.e* , गृह्याः.

In the hymn, सं ज्योतिषाऽभूवमेन्द्रीम् (तै. सं. १.६.६.२) । (पदकाले - अभूवम्), the word split is as shown in the brackets, while in .. ततो मा पराभूवमिति .. (तै. सं. २.५.१.२) । (पदकाले - भूवम्) it is different as seen in the brackets since the particle occurring before the preverb, *i.e,* मा invokes the rule of Pāṇini 6.4.74 and is justified. (also similar in मा परा भूमेति.. । पदकाले - भूम ।). But, to handle such cases in a morphological analyser, we need specific conditions to be checked which are to be devised carefully. In the earlier case

also, the preverb सं is separated from its verb part by a noun invoking the rule of Pāṇini 1.4.82. Hence, for accented and Vedic inputs, the extension of the morphological analyser has to be provided with these features.

4.3 Examples of Vikṛti-s (patterns)

Krama - संहितायाम् - .. य॒त ई॑श्व॒रो वा अश्वोऽय॒तोऽप्रतिष्ठितः .. (तै. सं. ५.४.१२.३) इति ।

पदपाठः - .. अश्वः॑ । अय॑तः । अप्रतिष्ठि॒त॒ इत्यप्रति - स्थि॒त॒ः । इति ।

क्रमपाठे तु - .. वा अश्व॑ः । अश्वोऽय॑तः । अय॒तोऽप्रतिष्ठितः । .. इत्यादि

Ghana - संहितायाम् - .. ब्र॒ह्म॒वा॒दिनो॑ वदन्ति॒ न्य॑ङ्ङ॒ग्निश्चे॑त॒व्या (३) .. (तै. सं. ५.५.३.३)

पदपाठः - .. ब्र॒ह्म॒वा॒दिन॒ इति॑ ब्रह्म-वा॒दिनः॑ । व॒द॒न्ति॒ । न्य॑ङ् । अ॒ग्निः । चे॒त॒व्य (३)॒ः । ..

घनपाठे तु - .. । .. व॒द॒न्ति॒ **न्य**॑ङ्न्य॑ङ्वदन्ति वदन्ति॒ न्य॑ङ्ङ॒ग्निर॒ग्निर्न्य॑ङ्वदन्ति वदन्ति॒ न्य॑ङ्ङ॒ग्निः । ..

It is to be noted that many phonetic processes like Kampa, the (physiologically necessary and inevitable) trembling of voice under certain accent collocations are best understood only by these Vikṛti Pāṭha-s. While splitting a Saṃhitā text into Pada, these are to be looked out for. (While processing a speech input, these could be suitably modelled and handled.)

There are cases of different subjective interpretations in Vedic texts, particularly in the Upaniṣad-s, where the morphological analysis can help clarify the rule position.

For example, in श्र॒द्धयादे॑वो देव॒त्वम॑श्नुते (तै. का. ३.३.१), it is interpreted as श्रद्धया and देवः by Sāyaṇa, while the word accent should have been श्र॒द्धया॑ and दे॒वः for such a division. (श्र॒द्धया॑ and अदे॑वः is the split, possibly**!**)

Similarly, in 'उ॒ताविद्व॒ान॒मुं लो॒कं प्रेत्य॑' (तै. आ. ५.१४.६), the word is split as 'अवि॑द्वान्'. But then, the accent should have been उ॒ताविद्वा॑न्. Hence, 'विद्वान्' looks suitable. See also in '.. लो॒कमाय॒न्नप्रसृ॑ते॒नासु॑रान् ..' (तै. आ. २.१) where the split is as 'नप्रसृ॑तेन', against the other possibility of 'अप्रसृतेन'.

Even in databases, the lexicon to be used for nominal stem check in the Vedic case is the Nighaṇṭu of Sage Yāska, besides the Vedic parts of Aṣṭādhyāyī.

Samples of Pada-pāṭha

In general, the procedure (pseudo-algorithm) for getting Pada- pāṭha from Saṃhitā or *vice versa* is given as below : (Ref. स्वरमञ्जरी, संहितापदपाठविमर्शः - त्र्यंबक बळवंत अभ्यंकर)

पदपाठः - पादो मानं संहितायां पदं मानं पदेषु च । एकत्र्यङ्कात्मकः कम्पः पदपाठे न चास्त्यतः ।। ६८

संहितापाठात् पदपाठः -

1 अधोरेखाङ्कितान् कुर्यात् स्वरितान् प्रचयानपि । 2 ततः पदानि विगृह्य सर्वाणि च पृथग् लिखेत् ।। ६९

3 उदात्तं च ध्रुवं कृत्वा लिखेत् प्रतिपदं स्वरान् । 4 यदि कश्चित्तत्र जात्यस्तं यथैव तथा लिखेत् ।। ७०

5 छांदसं यदशुद्धं स्यात् शुद्धं कृत्वा तु तल्लिखेत् । 6 वैशिष्ट्यानि च सर्वाणि पदपाठस्य दर्शयेत् ।। ७१

यथा - श्रुधी (संहितापाठे); श्रुधि (पदपाठः)

पदपाठात् संहितापाठः -

1 वैशिष्ट्यानि च सर्वाणि पदपाठस्य लोपयेत् । 2 अधोरेखाङ्कितान् कुर्यात् स्वरितान् प्रचयांस्तथा ।। ७२

3 कुर्याच्च पदसंयोगं संधिशास्त्रानुसारतः । 4 उदात्तकान् ध्रुवं कृत्वा रचयेत् स्वरपद्धतिम् ।। ७३

5 एकव्यङ्कात्मकं कम्पं योग्यस्थाने समालिखेत् । 6 छंदोयोग्यं ह्रस्वादीर्घादिकं सर्वं प्रकल्पयेत् ।। ७४

Given below are certain sample passages with specific features to be handled while parsing Vedic texts as input. We typically need **to convert from Prose form into word-split** ('pausa') form using rule-bases (besides Pāṇinian grammar) like Śikṣā, Prātiśākhya, Lakṣaṇa-grantha, Vikṛtipāṭha-niyama-s etc. and Varṇa-krama details. We have given the input as Saṃhitā and the Pada-pāṭha expected after split. As explained earlier, multiple possibilities, where exist, need resolution using these rulebases.

[Key : T.S. #.#.#.# = Taittirīya Saṃhitā of Kṛṣṇa Yajur Veda - Kāṇḍa, Praśna, Anuvāka, Pañcāśat]

1. *Saṃhitā pāṭha* from T.S. 5.5.3.2; example for **Pluta** in Saṃhitā

... ब्रह्मवादिनो वदन्ति न्यङ्ङग्निश्चेतव्या **(३)** उत्ताना **(३)** इति वयसाव्वाँ एष प्रतिमया चीयते ... ।

(*Pada-pāṭha*)

ब्रह्मवादिन इति ब्रह्म-वादिनः । वदन्ति । न्यङ् । अग्निः । चेतव्या (३) ः । उत्ताना (३) इत्युत्-ताना (३) ः । इति । वयसाम् । वै । एषः । प्रतिमयेति प्रति-मया । चीयते ।

2. *Saṃhitā pāṭha* from T.S. 5.6.21.1; example for **Pragṛhya** in Saṃhitā

सोमाय स्वराज्ञेऽनोवाहावनड्वाहाविन्द्राग्निभ्यामोजोदाभ्यामुष्टाराविन्द्राग्निभ्याम्बलदाभ्याꣳ सीरवाहाववी द्वे धेनू भौमी दिग्भ्यो वडबे द्वे धेनू भौमी वैराजी पुरुषी द्वे धेनू भौमी वायवे आरोहणवाहावनड्वाहौ वारुणी कृष्णे वशे अराड्यौ दिव्यावृषभौ परिमरौ ।।

(*Pada-pāṭha*)

सोमाय । स्वराज्ञे इति स्व - राज्ञे । अनोवाहावित्यनः - वाहौ । अनड्वाहौ । इन्द्राग्निभ्यामितीन्द्राग्नि - भ्याम् । ओजोदाभ्यामित्योजः - दाभ्याम् । उष्टारौ । इन्द्राग्निभ्यामितीन्द्राग्नि - भ्याम् । बलदाभ्यामिति बल - दाभ्याम् । सीरवाहाविति सीर - वाहौ । **अवी इति । द्वे इति । धेनू इति । भौमी इति** । दिग्भ्य इति दिक् - भ्यः । **वडबे इति । द्वे इति । धेनू इति । भौमी इति । वैराजी इति । पुरुषी इति । द्वे इति । धेनू इति । भौमी इति** । वायवे । आरोहणवाहावित्यारोहण - वाहौ । अनड्वाहौ । वारुणी इति । कृष्णे इति । वशे इति । अराड्यौ । दिव्यौ । ऋषभौ । परिमराविति परि - मरौ ।।

3. *Saṃhitā pāṭha* from T.S. 6.2.10.3; example for **kampa** in Saṃhitā

..... पितृदेवत्य ꣳ ह्येतद्यन्निखातय्यँद्बर्हिरनवस्तीर्य मिनुयात् ..... ।

(*Pada-pāṭha*)

पितृदेवत्यमिति पितृ-देवत्यम् ॥ हि । एतत् । यत् । निखातमिति नि-खातम् । यत् । बर्हिः । अनवस्तीर्येत्यनव-स्तीर्य । मिनुयात् ।

4. *Saṃhitā pāṭha* from T.S. 1.1.11.2; example for **two consecutive verbs accenting** in Saṃhitā

..... एता असदन्थ्सुकृतस्य लोके ता विष्णो **पाहि पाहि** यज्ञम्पाहि यज्ञपतिम्पाहि माय्यँज्ञनियँम् ।।

(*Pada-pāṭha*)

एताः । असदन् । सुकृतस्येति सु-कृतस्य । लोके । ताः । विष्णो इति । पाहि । पाहि । यज्ञम् । पाहि । यज्ञपतिमिति यज्ञ-पतिम् । पाहि । माम् । यज्ञनियमिति यज्ञ-नियम् ।।

5. *Saṃhitā pāṭha* from T.S. 4.6.2.5-6; example for **special rules of Prātiśākhya**

वाचस्पतिँविँश्वकर्माणमूतये मनोयुजव्वाँजे अद्या हुवेम ।।

(*Pada-pāṭha*)

वाचः । पतिम् । विश्वकर्माणमिति विश्व-कर्माणम् । ऊतये ॥ मनोयुजमिति मनः-युजम् । **वाजे** ॥ **अद्य** । **हुवेम**

6. *Saṃhitā-pāṭha* from T.S. 6.1.1.7; example for **double kampa** in Saṃhitā

इन्द्रो वृत्रमहन्थ्सो ऽ पो ऽ भ्यम्रियत ।।

(*Pada-pāṭha*)

इन्द्रः । वृत्रम् । अहन् । सः । अपः । अभीति । अम्रियत ।।

Chapter 5 : Language Structure

The syntactic analysis involves morphological analysis at the word level (to identify the words to their grammatical forms) and then assigning functional roles to each word in the sentence using the Kāraka-Vibhakti relations given in Pāṇini's Kāraka theory by an expectancy driven approach. The morphological analysis is the most important step in syntactic analysis. It involves identifying each of the words to their grammatical forms. Let us first have a look at what are the types of words that are possible in Sanskrit, then we will propose an algorithm for identifying each word to their grammatical forms (morphological analysis).

5.1 Definition of a word

A word, by definition, has शक्ति (s'akti) (denoting power or potency or significance), which is the relation of the word with an object, that always serves to revive the memory of that object whenever that word is spoken. This power is capable of denoting a particular object, the similarity of all the members of its class and distinguish from its dissimilar objects. *E.g,* The word "COW", can point out a particular cow, signify that the particular cow is similar to all the cows in the world and distinguish it from all other things such as stones, walls, trees etc.

Pāṇini has technically defined a word to be one with a nominal or verbal termination, सुप्तिङन्तम् पदम् (1.4.14). Hence a word can be denoted as follows:

W = B + T, where

W represents a word,

B represents the nucleus of a word and can take variety of forms depending on the type,

T represents the terminal suffix (either nominal declensional suffix or verbal conjugational suffix).

The meaning of the word is derived from the meaning of the nucleus and that of the termination.

5.2 Word structure

The Figs. 5.1 to 5.5 depict the word structure of Sanskrit. 5.1 gives the categories, 5.2 the chracteristic structural notations, 5.3 characteristic structural details, 5.4 compound word classification and 5.5 compound word categories.

As mentioned above, a word in Sanskrit has the basic structure, W = B + T. Depending on what form B and T take, the words can be classified into different types. According to Pāṇini, there are only two basic syntactic types of words, depending on the forms T takes. They are:

i. Nouns or Nominals : Here the terminal suffix T is a Nominal suffix *i.e,* $T = S_n$.

ii. Verbs : Here the terminal suffix T is a verbal suffix *i.e,* $T = S_V$.

But based on the form the nucleus of a word takes, the words in Sanskrit can be classified into two types as follows:

i. Non-Compound words & ii. Compound words.

The nucleus of a word B, may consist of one or more Nominal bases (Prātipadika-s), one or more Kṛt suffixes, one or more taddhita suffixes, a feminine suffix, a verbal root, one or more upasargas

etc (i.e the categories of compound and non-compound words are present in both nouns and verbs). Words which contain only one nominal base with zero or more suffixes (non - vibhakti), one verbal root with zero or more suffixes (non - terminal) along with terminal suffixes are termed as **Non-Compound** words. Words which contain more than one nominal base, one or more upasargas along with a verbal root etc are termed as **Compound** words.

NOTE : In the subsequent sections various notations used are as under :

W = finished word (Pada),
B = nucleus (Prakṛti),
b = base (nominal base - Prātipadika or verbal root - Dhātu),
T = Terminal suffix (Vibhakti),
S_n = Nominal suffix - Sup,
S_v = Verbal suffix - Tiṅ,
S_k = Kṛt suffix,
S_t = Taddhita suffix,
S_w = Vikaraṇa suffix, e.g, Śap,
S_m = Modal suffix, e.g, Yak,
S_{mc} = Causative mode suffix - Ṇic,
S_{md} = Desiderative mode suffix - San,
S_{mp} = Passive mode suffix - Yak,
S_s = Feminine (Strī) suffix,
N_b = Nominal base - Prātipadika,
V_r = Verbal root - Dhātu,

$S_|$ = non-terminal suffixes of different/same form such as taddhita suffix, Kṛt suffix etc,
A = represents a word from Gaṇapāṭha,
U_r = represents the rth upasarga (preverb).

5.2.1 Non - compound words

As defined above, Non-compound words are those which have only one nominal base or only one verbal root. Hence, we can represent a Non - Compound word in a very general form as follows:

$W = b + S_1 + S_2 + S_3 \ldots\ldots + S_| + T.$

The Non-Compound word structures can be further classified into:

i. Primary Structures ; ii. Secondary Structures

PRIMARY STRUCTURES

Primary structures are those which do not have any non-terminal suffixes. Hence, all the S's in the above general form are zeroes. We can represent a word with Primary Structure as follows:

$W = b + T.$

The Non-Compound words with primary structure are of three types and are detailed below.

NOUNS

Nouns with primary structures have nominal bases along with the terminal Sup suffixes (i.e Vibhakti pratyaya-s). The general form of a noun with primary structure can be represented as follows:

$W = N_b + S_n$

e.g, W = रामः ; = राम + सु ; = $N_b + S_n$

VERBS

Verbs with primary structures have verbal roots along with the terminal Tiṅ suffixes. The general form of a verb with primary structure can be represented as follows:

$W = V_r + S_v$

The verbal forms do not display even a single case of Primary structure of the form $V_r + S_v$, because in Pāṇinian System, all the Dhātu-s being divided into ten different conjugations (Gaṇa-s) take a Vikaraṇa, which is inserted in between the root, *i.e,* V_r and the terminal suffix, S_v. Every verbal structure therefore, is composed of three-morpheme-structure of the type $V_r + S_w + S_v$, where S_w is a conjugational suffix called Vikaraṇa pratyaya.

INDECLINABLES (AVYAYA-S)

Indeclinables or Avyaya-s of Primary structure have zero terminal suffix and the nuclei is one among the list given by Pāṇini as Gaṇa Pāṭha. The general form of this type of word can be represented as follows:

$W = A + 0$

Ex: W = च ; = च + 0 ; = A + 0

SECONDARY OR DERIVED STRUCTURES

Secondary or derived structures are those words which generally have non-terminal suffixes along with the terminal suffix. The number of non-terminal suffixes can be one or many depending on the type and compatibility. Hence, the general form of a word of secondary or derived structure is,

$W = b + S_1 + S_2 + S_3 + ... + S_l + T.$

SECONDARY STRUCTURES WITH ONE NON-TERMINAL SUFFIX

Secondary structures with only one non-terminal suffix are present in all the types of the words like verbs, Kṛdantas, taddhitas, feminine words etc. Here the base along with the Non-terminal suffix forms the nucleus for the word. The general form of these types of words is:

$W = b + S_l + T$, where, $b + S_l = B$ is the Nucleus of the word and T is the Terminal suffix.

KṚDANTAS

Kṛdantas are a subtype of nouns. The words of this type have the following general structure,

$W = V_r + S_k + S_n$, where, $V_r + S_k = B$, is the nucleus of the word taking nominal suffix.

The Kṛdantas represent active-substances, with stress on activity performing substances. (Nouns represent substances and the verbs represent activities). Kṛdantas are similar to Gerunds/Participials in English grammar.

e.g, W = हरः ; = हर् + अ + सु ; = $V_r + S_k + S_n$

The Nominal Singular forms of some of the root-nouns formed by adding Kṛt suffixes like क्विप्, क्विन्, ण्वि etc. present a still more interesting case in which both the non-terminal and terminal suffix amount to zero. Take, for example, the form स्पृक्, from the root स्पृश्, with the suffix क्विन्. The grammatical procedure is as follows:

W = स्पृक्

= स्पृश् + क्विन् + सु ($N_b + S_k + S_n$)

= स्पृश् + 0क्विन् + सु (क् is deleted according to, 1.3.8)

= स्पृश् + 0वि0 + सु (न् is deleted according to, 1.3.3)

= स्पृश् + 000 + सु (व् & इ are deleted according to 6.1.67 and 1.3.1; 1.3.9)

= स्पृश् + 000 + 0 (सु is made zero according to 6.1.68)

= $N_b + S_k + S_n$, here, both S_n & S_k are zeroes.

TADDHITA-S

The Taddhitas take taddhita suffix along with the terminal Sup suffix. They are also a subtype of nouns. The general form of the Taddhitas is as given below:

$W = N_b + S_t + S_n$, where, $N_b + S_t = B$ is the Nucleus taking Nominal suffix.

e.g, W = शैवः ; = शिव + अ + सु ; = $N_b + S_t + S_n$

FEMININES

Some feminine bases take Feminine or Strī suffix along with terminal Sup suffix. They are also a subtype of nouns. The general form of Feminines is as given below:

$W = N_b + S_s + S_n$, where, $N_b + S_s = B$ is the nucleus which takes a Nominal suffix.

e.g, W = अजाम् ; = अज + आ + अम् ; = $N_b + S_s + S_n$

VERBS

The simplest form of verbs fall under this category. The only Non-Terminal suffix they take are the conjugational suffixes or the Vikaraṇa pratyayas. The general form of this type of words is as given below: $W = V_r + S_w + S_v$, where, $V_r + S_w = B$ is the nucleus which takes a Verbal suffix.

e.g, W = नश्यति ; = नश् + य + ति ; = $V_r + S_w + S_v$

AVYAYA-S

Avyaya-s or Indeclinables of derived form can be classified into two types depending on the Non-Terminal suffix they take, *i.e,* this type of words have T = 0 and the Non-Terminal suffix S can be either a Kṛt suffix S_k or a Taddhita suffix S_t. The general form of this type of words con be denoted as shown below:

1. $W = N_b + S_t + 0$
2. $W = V_r + S_k + 0$

e.g,
1. W = कृत्वा ; = कृ + त्वा + 0 ; = $V_r + S_k + 0$
2. W = पूर्वतः ; = पूर्व + तसिल् + 0 ; = $N_b + S_t + 0$
3. W = परत्र ; = पर + त्रल् + 0 ; = $N_b + S_t + 0$

As shown above, the base along with the Non-Terminal suffix forms the nucleus for the terminating suffix T, in the words with secondary structures having a single Non-Terminal suffix.

SECONDARY STRUCTURES WITH MORE THAN ONE NON-TERMINAL SUFFIX

There are many words in Sanskrit which take more than one Non-Terminal suffix. In such words, each base with the immediately succeeding suffix forms the base for the following suffix. The general form of such words will be as given below:

$W = b + S_1 + S_2 + S_3 + \ldots + S_l + T$

Here $b + S_1$ will be the base for the remaining part of th word, this can be represented as,

$W = b + S_2 + S_3 + \ldots + S_l + T$

Similarly $b + S_2$ will be the base for the remaining and so on. Hence, the Secondary structures can be reduced to the simple form, $W = B + T$.

The following are examples of Non-Compound words with more than one Non-Terminal suffix:

1. Verbal Form

W = चोरयति ; = चुर् > चोर् + अय + ति ; = चोर् + अय् + अ + ति ; = $V_r + S_{mc} + S_w + S_v$

2. Taddhita form

W = कर्तृत्वम् ; = कर् + तृ + त्व + अम् ; = $V_r + S_k + S_t + S_n$

The Secondary or derived structures of this type can be further divided into two main types as:

1. Homogeneous Structures & 2. Heterogeneous Structures.

HOMOGENEOUS STRUCTURES

The Homogeneous structures are those in which all the S's (Non-Terminal Suffixes) are from the same type. Only Taddhita suffixes generate such structures.

A good example of a such a type of word is the taddhita formation वासुदेवकः. The formation is obviously from the base वासुदेव with the suffix अक (Pāṇinian वुन्) in the sense of "One's devotion". The whole form वासुदेवकः would then mean "One whose devotion is towards वासुदेव, *i.e,* Lord Krishna". The form वासुदेव in turn is derived from the word वसुदेव with the suffix अण्. The declined form वासुदेवकः can then be analysed as follows :

W = वासुदेवकः ; = वसुदेव + अण् + अक + सु ; = $N_b + S_t + S_t + S_n$.

As both the Non-Terminal suffixes, viz, अण् and अक are Taddhita suffixes, the word वासुदेवकः is of Homogeneous nature.

The Kṛdanta suffixes, however, do not provide structures of Homogeneous nature because they do not follow those of their own class. No two immediately succeeding Kṛdanta suffixes can ever be seen to build a single structure of Homogeneous type. A Kṛdanta suffix, however, can be followed by a taddhita one, as in the example कर्तृत्व, but never by a Kṛdanta suffix. (Note : Pāṇinian technique and Sanskrit grammar does not allow a Kṛdanta suffix to follow another Kṛdanta suffix).

HOMOGENEOUS STRUCTURES WITH RECURSIVE SUFFIXES

Some Taddhita suffixes can be affixed to a word repeatedly in a recursive fashion. Such words can be infinitely expanded by recursively adding the suffixes. In such words, each additional suffix will modify the meaning of the word, and the whole word has a unique meaning different from the word without the additional suffix.

The general structure of such words is as given below:

W = $b + S_1 + S_2 + S_3 \ldots\ldots S_\infty + T$, where,

b will generally be a Nominal base N_b & S's will be Taddhita suffixes.

Consider the case of Taddhita suffixes applied in the sense of "His progeny/child"; the general suffix to be applied is अ, *ṭ.ङ्,* Pāṇinian अण् according to the sutra प्राग् दीव्यतोऽण्, 4.1.83. Thus, from the pratipadika उपगु a formation signifying "Upagu's child" will be :

उपगु + अण् > अ ; = औपगव् + अ (applying वृद्धि to the first उ, according to the sutra 7.2.115)

= $N_b + S_t$

Now the formation औपगव can again become the base for the next suffix signifying "The child of औपगव". The suffix in this case is not अण् but इ (*i.e,* the Pāṇinian इञ्, according to the sutra अतः इञ्, 4.1.95) because the base औपगव ends in अ. Hence, the next formation is औपगवि.

This formation, *viz.* औपगवि, can again form the base for the next formation meaning "Child of औपगवि", with the suffix आयन, *i.e,* originally Pāṇinian फक्, which is applied according to the sutra यञिञोश्च, 4.1.101. Thus

W = औपगवि + आयन ; = औपगवायन (इ is deleted according to 6.4.184)

= औपगव् + अ + इ + आयन ; = $N_b + S_t + S_t + S_t$

This formation, *viz.* औपगवायन, again ends in अ and is liable to get the suffix इ. The next formation with इ will then be औपगवायनि, which would again get the suffix आयन, giving out the formation औपगवायनायन. Thus alternating from this stage onwards, the suffix इञ् and फक् > आयन would give an infinitely expanded structure in which each succeeding S signifies "The child of the previous base". To form a complete word, all the above formations take Nominal suffixes (Sn).

HETEROGENEOUS STRUCTURES

Heterogeneous structures are those words in which all S's (Non-Terminal suffixes) are of different types, *i.e,* the succeeding suffixes are always of different types. The Kṛt suffixes, the verbal conjugational suffixes etc. all give rise to such structures.

NOMINAL STRUCTURES

The Kṛt suffixes along with taddhita suffixes generate such type of words. Also, these words may additionally take feminine (Strī) suffixes.

W = नादवतीम् ; = नादवती + अम् ; = नादवत् + ई + अम् ; = नाद + वत् + ई + अम्

= नाद् + अ + वत् + ई + अम् ; = $N_b + S_k + S_t + S_s + S_n$

VERBAL STRUCTURES

The Conjugational verbal suffixes also take other suffixes such as modal suffixes etc, thus generating Secondary or Derived structures.

W = चिकीर्ष्यते ; = चिकीर्ष्य + ते ; = चिकीर्ष + य + ते ; = कृ + स + य + ते

= $V_r + S_{md} + S_{mp} + S_v$. (Desiderative Passive Verbal structure)

As discussed with Homogeneous recursive suffixes, it is possible to expand the verbal formations to the logical extremity, *i.e,* infinity. These verbal formations will be necessarily Heterogeneous in nature. It is to be noted here that Pāṇini has neither stated explicitly the process of infinite formations for verbal forms (as he has done in the case of Taddhita formations detailed above) nor has he prohibited its application anywhere.

5.2.2 Compound words

The Compound words in Sanskrit are those that contain more than one Nominal bases or a verbal root with one or more upasarga-s along with zero or more Non-Terminal suffixes and a Terminal suffix. Hence, in general, a compound word can be represented as,

$W = b_1 + b_2 + b_3 \ldots + b_l + S_1 + S_2 + S_3 \ldots + S_l + T$, where,

b's are the bases and can be Nominal bases, Verbal roots or secondary bases.

The Compound words are classified into:

1. Primary Compound word structures, 2. Secondary or Derived Compound word structures.

PRIMARY COMPOUND WORD STRUCTURES

The Primary Compound word structures are those that do not have any Non-Terminal suffixes. Hence, a Primary Compound word can be denoted as follows:

$W = b_1 + b_2 + b_3 \ldots + b_l + T$

As there were no Verbal Primary Non-Compound forms, so also there are no Verbal Primary Compound forms (because Verbal roots generally take a conjugational suffix in the least). Hence, only nominal bases give rise to Primary Compound word structures.

COMPOUND NOUNS

In this type of words, more than one Nominal bases combine to form the Nucleus of the word, which combines with the terminal Nominal suffix to form a noun. The general form of such words can be represented as follows:

$W = N_{b1} + N_{b2} + N_{b3} \ldots + N_{bl} + T$

W = रामलक्ष्मणौ ; = रामलक्ष्मण + औ ; = राम + लक्ष्मण + औ ; = $N_{b1} + N_{b2} + S_n$

In such types of words, the compound nucleus can be expanded infinitely by combining the Nominal bases one after another. Such types of compounds are called Dvandva Compounds in the classical classification of Sanskrit Compounds (The classical works of Sanskrit classify the compounds into 4 main types. They are Avyayībhāva, Tatpuruṣa, Dvandva and Bahuvrīhi).

SECONDARY OR DERIVED COMPOUND WORD STRUCTURES

Those Compound word structures that have one or more Non-Terminal suffixes are called Secondary or Derived Compound word structures. The general structure of such a type of word is :

$W = b_1 + b_2 + b_3 \ldots + b_l + S_1 + S_2 + S_3 \ldots + S_l + T.$

Both Verbs and Nouns exhibit Derived Compound word structures.

NOUNS

There are two types in this type of structures.

i. One in which only one of all the nuclei is साधित (derived) and the rest underived.

ii. One in which all nuclei could be साधित (derived).

Ex1 : W = पक्वफलम् ; = पक्व + फल् + अम् ; = $N_{b1} + N_{b2} + S_n$

Here the first Nominal base पक्व is a derived or a secondary base.

N_{b1} = पच् + क्त (according to 1.1.26)

= पक्त . This changes to पक्व from the sutra पचो वः, 8,2,52.

Ex2 : W = तुष्टिकर्ता ; = तुष्टि + कर्तृ + सु ; = $N_{b1} + N_{b2} + S_n$

Here both the Nominal bases are derived or secondary bases.

N_{b1} = तुष् + टि ; N_{b2} = कृ + तृ

Most of the Compound words so formed are declinable and take non-zero terminal vibhakti pratyaya. But there are some which take only the Nominative case, Singular and Neuter gender form. Such Compound words are indeclinable. The Compound words which are indeclinable, are included in the Avyayībhāva Samāsa type, in the Pāṇinian Classification of Compound words.

VERBS : Verbal roots taking upasarga prefixes can be considered as compound words (This assumes that upasarga-s are independant constituents of Sanskrit grammar).

Hence, words like आहरति (आ + हरति), प्रतिगृह्णाति (प्रति + गृह्णाति), अभिगच्छति (अभि + गच्छति) can be considered as compound words. The general form of such words can be represented as, $W = U_1 + U_2 + U_3 \ldots + U_l + V_r + S_1 + S_2 + S_3 \ldots + S_l + T.$

Though Pāṇini does not explicitly term these verbal structures joined with upasarga-s as compounds, later commentators on Pāṇinian grammar such as Patañjali, Bhaṭṭojī Dīkṣita etc. have proved that they can be considered so. For proving this, they have resorted to a technique called योगविभाग, *i.e,* they split the Pāṇinian Sūtra सह सुपा, 2.2.4, which lays down the nature and pre-requisites of two words to form a compound.

OTHER CLASSIFICATIONS OF COMPOUND WORDS

Till now the Compound words in Sanskrit were classified based on their structure, but Sanskrit Compounds can be classified into the following seven types according to Pāṇini:

1. अव्ययीभाव समास (Avyayībhāva Compounds),
2. तत्पुरुष समास (Tatpuruṣa Compounds),
3. द्वन्द्व समास (Dvandva Compounds),
4. बहुव्रीहि समास (Bahuvrīhi Compounds),
5. कर्मधारय समास (Karmadhāraya Compounds),
6. द्विगु समास (Dvigu Compounds),
7. उपपद तत्पुरुष समास (Upapada Tatpuruṣa compounds).

The last three types of Compounds are sub-varieties of Tatpuruṣa Compounds. The following discusses briefly these different forms of compounds.

अव्ययीभाव समास : The Compound words belonging to this category are characterised by the fact that the meaning of the first member (in a two member Compound) is pre-dominant.

पूर्वपदार्थप्रधानः अव्ययीभावः ।

Generally, atleast one of bases joining to form a compound is an Avyaya, but sometimes even if all the words are not of the indeclinable type, they combine to form an indeclinable Compound. *e.g,*

1. W = पारेगङ्गम् ; = गङ्गायाः पारे (Vigraha Vākya)

Here, both the bases are not avyaya-s.

2. W = उपकृष्णम् ; = कृष्णस्य समीपम् (Vigraha Vākya)

Here, the first base is an avyaya.

3. W = शाकप्रति ; = शाकस्य लेशः (Vigraha Vākya)

Here, the second base is an avyaya.

In general, the compounds of this type are indeclinable, but in some cases like प्रातःकर्म, they take the vibhakti pratyaya-s. But, in such cases, Pāṇini states that the words are to take vibhakti pratyaya-s of Neuter gender only.

तत्पुरुष समास : Compound words in which the meaning of the second member (of a two member compound) is pre-dominant are grouped into this type of Compounds.

उत्तरपदार्थप्रधानः तत्पुरुषः ।

The Tatpuruṣa Compounds are of three main sub-types, the main types with examples are listed below,

i. द्विगु समास :

W = षाण्मातुरः ; = षण्णां मातॄणामपत्यं पुमान् (Vigraha Vākya)

ii. उपपद तत्पुरुष समास :

W = वज्रभृत् ; = वज्रम् बिभर्ति इति वज्रभृत् (Vigraha Vākya)

iii. कर्मधारय :

W = नीलघटः ; = नीलश्चासौ घटश्च (Vigraha Vākya)

There is a fourth minor variety of Tatpuruṣa called the Gati-Tatpuruṣa,

e.g, W = प्रगत ; = प्र + गत

There are other types of Tatpuruṣa compounds called व्यधिकरण तत्पुरुष, which is again further divided based on the the case.

द्वन्द्व समास : Compound words in which the meaning of all the members is equally significant are called Dvandva Compounds.

उभयपदार्थप्रधानः द्वन्द्वः ।

There are two types of Dvandva Compounds as listed below with examples.

i. समाहार द्वन्द्व :

W = पाणिपादम् ; = पाणी च पादौ च, एषां समाहारः (Vigraha Vākya)

ii. इतरेतर द्वन्द्व :

W = जम्पती ; = जाया च पतिश्च (Vigraha Vākya)

बहुव्रीहि समास : Compound words in which the members combine to signify the meaning of a third entity (not represented by any of the members) are called Bahuvrīhi Compounds.

अन्यपदार्थप्रधानः बहुव्रीहिः ।

e.g, W = प्राप्तानन्दः ; = प्राप्तः आनन्दः यं सः (Vigraha Vākya)

In Pāṇinian classification of Compounds, the chief characteristics of compounds are:

i. There are a minimum of two bases in the structure.

ii. All the constituent members lose their inflexional terminations while forming a compound.

But there is still a new variety of compounds in Sanskrit in which the inflexional terminations (vibhakti pratyaya-s) of the constituent members in general and of all the members except the last one are not lost. Such compounds go by the name "अलुक् समास" or "Non-Zero Compounds-Structures" in Pāṇini's grammar. *e.g,*

1. W = पुरन्ध्री (द्वितीयालुक् समास) ; = पुरं दधातीति (Vigraha Vākya)

2. W = आत्मनेपदम् (चतुर्थ्यलुक् समास) ; = आत्मने पद्यत इति (Vigraha Vākya) etc.

As can be understood, the above classification is based on the semantic grounds. But Bhaṭṭojī Dīkṣita in his "Siddhānta Kaumudī" not only classifies the Samāsa's (Compound Words) into the aforementioned four main types, but also gives another form of classification, which seems to have been the view of Sanskrit Grammarians before Pāṇini. This classification is stated by a Kārikā in Siddhānta Kaumudī, सुपाम् सुपा तिङा नाम्ना धातुनाथ तिङाम् तिङा । सुबन्तेनेति विज्ञेयः समासः षड्विधो बुधैः ।।

This means that Samāsa's in Sanskrit can be formed by the combinations of the words of the following types: 1. Subanta + Subanta ; 2. Subanta + Tiṅanta ; 3. Subanta + Nāma Dhātu

4. Subanta + Dhātu ; 5. Tiṅanta + Tiṅanta ; 6. Tiṅanta + Subanta

As we can see, this classification of Sanskrit compounds is based on the syntactic nature of the constituent words and is more nearer to the structure of them.

In sum, we can say that all words in Sanskrit and consequently in Pāṇinian grammar describing it, are made up of a nucleus and suffix. There are different types of words depending on the form the base and/or suffix take. Now, if we are to analyse a given word to its grammatically valid form, we have to identify all the constituents of the word.

Consider for example, the analysis of the word कर्तृत्वमयम्.

W = कर्तृत्वमयम् ; = कर्तृत्वमय् + अम् (b + S_n)

b = कर्तृत्व + मय् (b_1 + S_k) ; b_1 = कर्तृ + त्व (b_2 + S_t) ; b_2 = कर् (<कृ) + तृ (V_r + S_k)

Hence, the word कर्तृत्वमयम् is of the structure : V_r + S_k + S_t + S_k + S_n, *i.e,* given a word, we should divide it into a base and suffix until no further division is possible. The sum of identifications of all the parts of the word will give rise to the identification of the total word.

If we get two or more nuclei at any stage of analysis, the word being analysed is a compound word. If, however, we do not meet with two or more nuclei but only one and find one or more suffixes, the word is a non-compound word. Similarly, if the terminal suffix is a sup suffix, then the word being analysed is a Noun. The sup suffix also provides information about the case and number of the word. But, if the terminal suffix is a Tiṅ suffix, then the word being analysed is a verb. The Tiṅ suffix also gives the information about person and number of the word.

From the above discussion of the types of words, we can conclude that a Sanskrit word can be identified to its grammatical form by identifying all its constituents. To identify the constituents, we should have prior knowledge of the characteristics of each constituent or a list of entities that can be identified as a particular type of constituent (for example, a list of Nominal suffixes).

Pāṇini's grammar for Sanskrit has given lists of pre-classified entities, like a list of verbal roots (Dhātu Pāṭha), a list of suffixes, a list of indeclinables, a list of upasarga-s etc. The rules which state what constituents can combine with what, are given in Sūtra style.

If this information can be embedded in an algorithm, then generation of Sanskrit words can be automated. Once this is accomplished, the analysis of a Sanskrit word would be the reverse process of this. This process may involve coding the rules completely in a program or splitting up the information contained in the rules between the databases and the programs.

5.3 Word analysis

The process is as follows :

When a word is input in Devanagari, it is subjected to base-suffix analysis (Prakṛti-pratyaya Viśleṣaṇam). This step is necessaitated by the fact that any finished word must necessarily have both the parts. We split the given word into two parts iteratively and consider all the LHS-RHS combinations.

The LHS must contain a base or root while the RHS must contain the suffixes. The base portion could be a nominal stem (Prātipadika) or a verbal root (Dhātu), primary or derived. The suffix part could contain multiple suffixes, but we look at the right end of the word for the terminating suffix first, which could only be nominal or verbal.

The nominal terminating suffixes are 24 for nouns (8 cases and 3 numbers) and verbal terminating suffixes are 90 (9(3 persons and 3 numbers) per tense/mood which are 10 in all -1 present, 3 past, 2 future and 4 moods.

As the nominal terminations vary according to base endings, gender etc. we look up at a database of these modified nominal termnating suffixes to find the proper identification. A syntactic type check of base and the suffix type arrived at is carried out. When this is passed, we get a valid noun identification for the input word.

Similarly, if verbal terminating suffixes match, we look up at the modified verbal suffix database for possible match. Syntactic type match between the root on the LHS and suffix found on the RHS is done to ensure compatibility. Preverbs are also looked for in the LHS part, for verbal suffix matches of the RHS using database of preverbs.

Indeclinables are identified by string match of the whole (input) word against the database of Avyaya-s. These are done In any case for all words to ensure identification of all possible grammatical forms, *i.e,* to cover multiple identifications for each word.

As an illustration, let us take the input word: रामः

Example 1: रामः = र् + आ + म् + अ + ः

1. **राम ः**

2. राम् अः

3. **रा मः**

4. र् आमः

5. रामः The full word itself

The second level split :

1. राम ः

2. राम् अः

3a. र अमः

3b. र आमः

3c. रा अमः

3d. रा आमः

4. र् आमः

5. रामः The full word itself.

By beginning with RHS and checking for terminating suffixes, we find

#2 produces an identification as Noun - Nominative singular of mascular base राम, meaning 'Rāma'.

#3 produces an identification as Verb - active voice, present tense, first person plural of root 'रा' meaning 'we are giving'. Here, we get both answers at the first level itself.

Second level Splits like : Ṛ + Amaḥ, Ṛ + Āmaḥ, Ṝ + Amaḥ, Ṝ + Āmaḥ are also possible and a root Ṛ and Ṝ exist ; and Āmaḥ is a valid suffix, but the root Ṛ gets 'Ṛcch' Ādeśa and hence the verb should be Ṛcchāmaḥ. These are to be handled in conjunction with Grammar rules.

Example 2: मातः = म् + आ + त् + अ + ः

1. **मात ः**

2. मात् अः

3. **मा तः**

4. म् आतः

5. मातः

Second level split :

1. मात ः

2. मात् अः

3a. म अतः

3b. म आतः

3c. मा अतः

3d. मा आतः

4. म् आतः

5. मातः

By first level analysis, we find :

#2 produces an identification as Noun - vocative singular of feminine base, 'मातृ', meaning 'addressing mother'.

#3 produces an identification as Verb - active voice, present tense, third person dual of root 'मा' meaning two persons are measuring'.

At the second-level split, we get another identification of the word as a combination of two words at #3c, as मा + अतः meaning 'not from this', if both are indeclinables. The first word can also be fem. nom. sing., meaning 'Lakṣmī', pronoun 'asmad' in acc. sing. meaning 'me'.

This level works on the basis of breaking up each character (vowel or consonant) into its appllicable constituents as per sandhi rules. This is essential in splitting combined words and also identifying compound words.

Example 3: यदि॒मे. (Vedic example from .... यदि॒मे गर्भ॒मद॑धातां ..... Taittirīya Saṃhitā 3.4.3.2)

यदि॒मे = य् + अ + त् + इ +_ + म् + ए.

1. यदि॒मे

2. यदि॒म् ए

3. **यदि॑ मे** (3a. यदि॑ मे॒)

4. **यत् ड्इमे**

5. य दि॒मे

6. य् अदि॒मे

Here, we have to note that vowels and accents are to be together on either LHS or RHS though, for clarity, the accent (Anudātta) is shown separately in the first line.

However, at #3, the word ending Anudātta (of LHS) should be restored to its Svarita state as it is a Tairovyañjana Svarita (as per the definition, Udāttapūrvaḥ Tairovyañjanaḥ - Taittirīya Prātiśākhyaṃ). But, this needs to be confirmed against the accent of the following vowel, or from the word itself, as the Svarita on the LHS end could be the case if the RHS is Anudātta also.

Hence, if the string मे is a valid word, then as the LHS is a valid Avyaya (meaning 'if'), we get an identification. Here, मे cannot be from the root 'Asmad' (meaning 'my', or 'to me') since thjat word is Anudātta (shown as #3a in brackets).

#4 also produces an answer as both the sides here are valid words. Here, the Pada Pāṭha also corroborates the split. The Pada Pāṭha for the input string is : यत्। इमे इति । In fact, that the second word is an 'itiga', *i.e* , Pragṛhya.

In the case of Vedic texts, though for Saṃhitā part, we already have the Pada Pāṭha (word-split form) in recitational tradition and it is supreme even if a rule-based approaach would produce a different form of split, we need this algorithm and rulebases (even including empirical ones described in the Lakṣaṇa Grantha-s) for understanding meanings, and most importantly, for the Brāhmaṇa and Upaniṣad texts' analysis for which traditionally no Pada Pāṭha is available. In the traditional teaching, the importance of this issue is recognised prominently.

5.4 Word-Splitting - multiple splits

The problem of generating all the splits given a string of length l

The problem involves generating all the splits of a given string of length l. Consider a string of 4 characters in english, *viz.* HELP. Now if we are to generate all the possible splits of this string, it will give rise to 8 'string' combinations as given below.

1. H E L P
2. H E LP
3. H EL P
4. H ELP
5. HE L P
6. HE LP
7. HEL P
8. HELP

This process, as we can see, will give rise to many such string combinations as the length of the input string increases. It can be shown that if the length of the string is l, then the word combinations that can be generated are 2^{l-1}. We can see that this is an exponential problem and will be impossible to handle as the length of the string increases, if the straight forward method is used.

The problem is more complicated in Sanskrit language because a string of (meaningful) characters may have sandhi, internal or external, between characters which constitute the string. Now if we are to generate all the possible 'word combinations' of a Sanskrit character string of length say 'l', then, we should take into account any sandhi's between characters.

Eg : Consider the character string **रामः**

It is made of the following characters, र् + आ + म् + अ + ः

As the string contains 5 characters the total number of string combinations will be 2^{5-1} *i.e,* 16, the 16 string combinations are as follows:

1. र् + आ + म् + अ + ः
2. र् + आ + म् + अः
3. र् + आ + म + ः

4. र् + आ + मः
5. र् + आम् + अ + ◌ः
6. र् + आम् + अः
7. र् + आम + ◌ः
8. र् + आमः
9. रा + म् + अ + ◌ः
10. रा + म् + अः
11. रा + म + ◌ः
12. रा + मः
13. राम् + अ + ◌ः
14. राम् + अः
15. राम + ◌ः
16. रामः

These 16 combinations do not take into account the sandhi's (if they exist). For instance, in the above example, the character आ can be formed due to any of the following four character combinations,

i.e, आ = अ + अ ; आ = अ + आ ; आ = आ + अ ; आ = आ + आ

Similarly, र् can be replaced by any of the following characters ऋ, ॠ, ◌ः and the character म् can be replaced by ◌ँ (Anuswara) under certain conditions. So, theoretically, the word combinations to be generated using the above string of 5 characters is,

Word combinations = 2^4 *(4*(1+2*4)*2) = 16 *(4*9*2) = 16*72 = 1152.

Hence, to theoretically generate all the possible word combinations of the string रामः, 1152 word combinations are to be generated. In general, if the length of the string is L, and m characters in L have a further split as shown above and each have N_m different possibilities including the character in the original string, then the total word combinations to be generated is,

Word combinations= 2^{L-1}*{ $(X_1*N_1)*(X_2*N_2)*(X_3*N_3)*.....(X_m*N_m)$}

where $X_m = N_m$

= 1+ 2*(N_m-1) depending on whether the character is replaced by one char or two chars.

But most of the string combinations so generated do not make any linguistic sense. Hence, the problem can be reduced to a considerable extent by applying some heuristics (Sometimes this will result in loss of some answers, but otherwise there is no go). Some of the heuristics that can be applied are,

1. When characters have further expansion in the given string, take each character at a time (rather than all permutations). This will reduce the number of word combinations to,

Word combinations = 2^{L-1}*{$(X_1*N_1)+(X_2*N_2)+(X_3*N_3)+... (X_m*N_m)$}

2. A single character (a consonant or a vowel) does not generally make a word. Hence, instead of treating each character as an entity (which is the case in english), a syllable can be treated as an entity.

Here, we adopt the following strategy for forming a syllable,

i. Add a consonant to the following vowel (starting from the left end of the string). If a vowel occurs first, it is taken as a syllable, if two or more consonants occur consecutively, then take each consonant as a separate syllable, This step is necessary for getting some answers which involve conjunct consonants. For example, consider the word वनम्गच्छति. Here, म् and ग occur consecutively. Here, if we attach consonants to only the following or the preceding vowel, it would not be sufficient. Hence, to get the answer वनम् गच्छति, the splitting of conjunct consonants is necessary.

ii. Add a consonant to the preceding vowel (end at the first occurrence of consonant.)

This will reduce the problem to a large extent. An algorithm which generates the two kinds of syllables mentioned above, and combines them in the manner outlined is developed.

IMPORTANCE OF GENERATING ALL SPLITS

The algorithm developed for generating all the possible splits of a given input string is very useful for NLP in Indian languages in general and Sanskrit in particular. As is obvious, this will help in the Morphological analysis of all kinds of Sanskrit words like nouns, Compound Nouns, Verbs, Compound Verbs, Kṛdantas, Taddhitas etc.

The algorithm will also be useful in giving out all the possible `word' combinations of a given input string. That is, if all the words of a given sentence are combined without any word demarcation, then this algorithm can be used to generate all the possible `word' combinations.

For example, consider the string, नमेयम् (from Mahākavi Veṅkaṭādhvarin's Lakṣmī Sahasram '.. **नमेयम्** उच्चैरितरा गतिर्न मे ।'. This string may give different meanings by different word splits as below :

i. नमेयम् - one word, verb, meaning let me/I shall salute.
ii. नम इयम् - (you) salute, this (fem.).
iii. न मेयम् - not measurable/conceivable.
iv. न मे यम् - not, for me/mine, whom etc.
v. न मा इयम् - not, me/Lakshmi, this (fem.) etc.
vi. न मे अयम् - This (mas.) (is) not mine/for me etc.

Hence, this algorithm is useful in giving out all the possible sentences that can be formed out of a given set of characters. Sometimes, this may give out some surprising options (of a given set of characters) which would not have occurred to humans. Because, humans are taught to understand a given set of characters in a certain manner, and think it better to go with the conventional analysis of the sentence formed out of the given set of characters than exhaust all possibilities.

When all the sub-string combinations are generated, there is every possibility that most of the sub-strings in the generated combinations are same. Hence, a slightly different approach is adopted in the actual implementation of the above algorithm for Morphological analysis of Sanskrit words. Here, firstly all the unique sub-strings from the different combinations are collected and each is passed through all the databases and all the identifications of the sub-strings are stored. Then we take a sub-string combination containing all valid sub-strings and if the structure of the combination is one among the valid word structures, then the the identification of the word for that particular split is obtained.

SANSKRIT WORD STRUCTURE

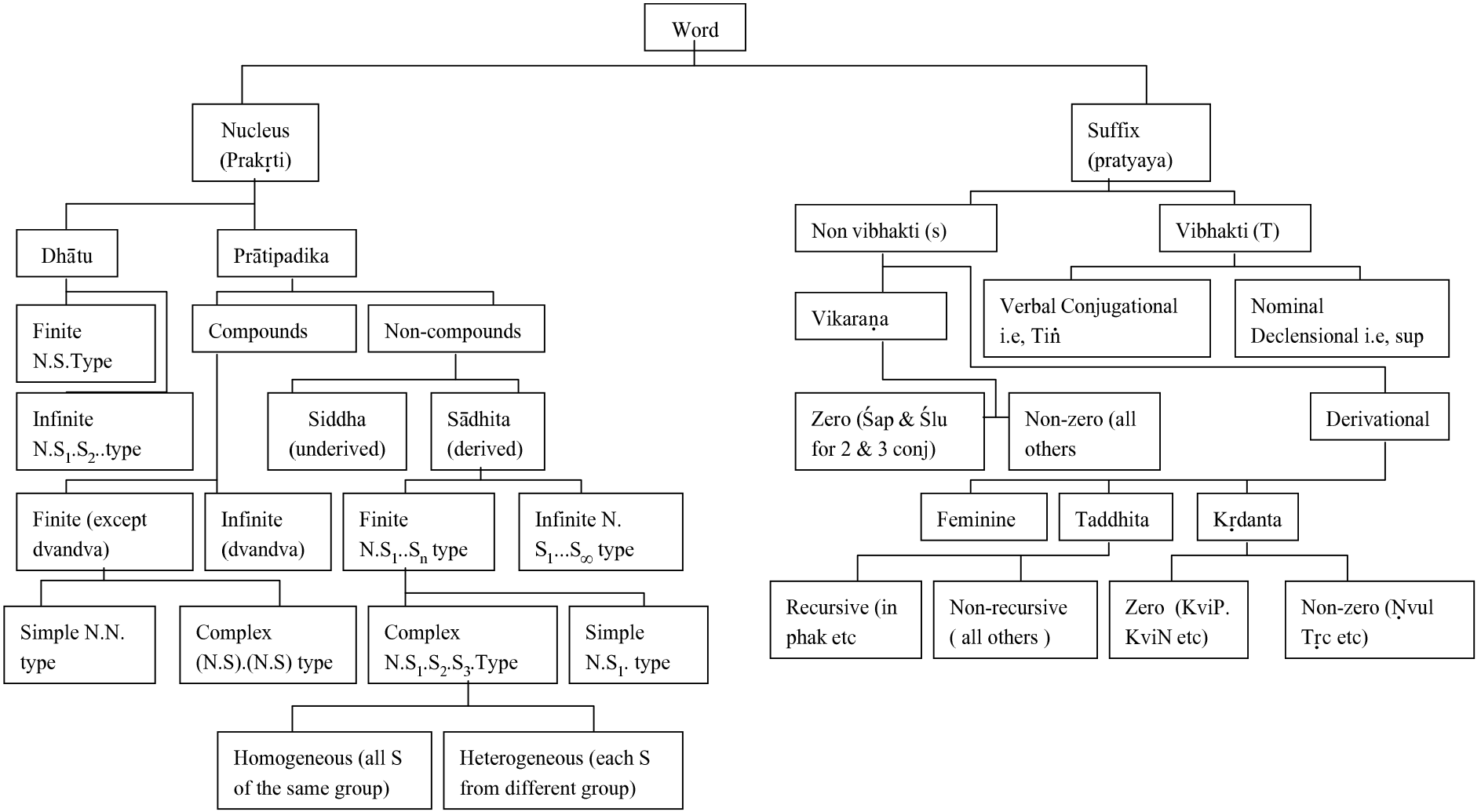
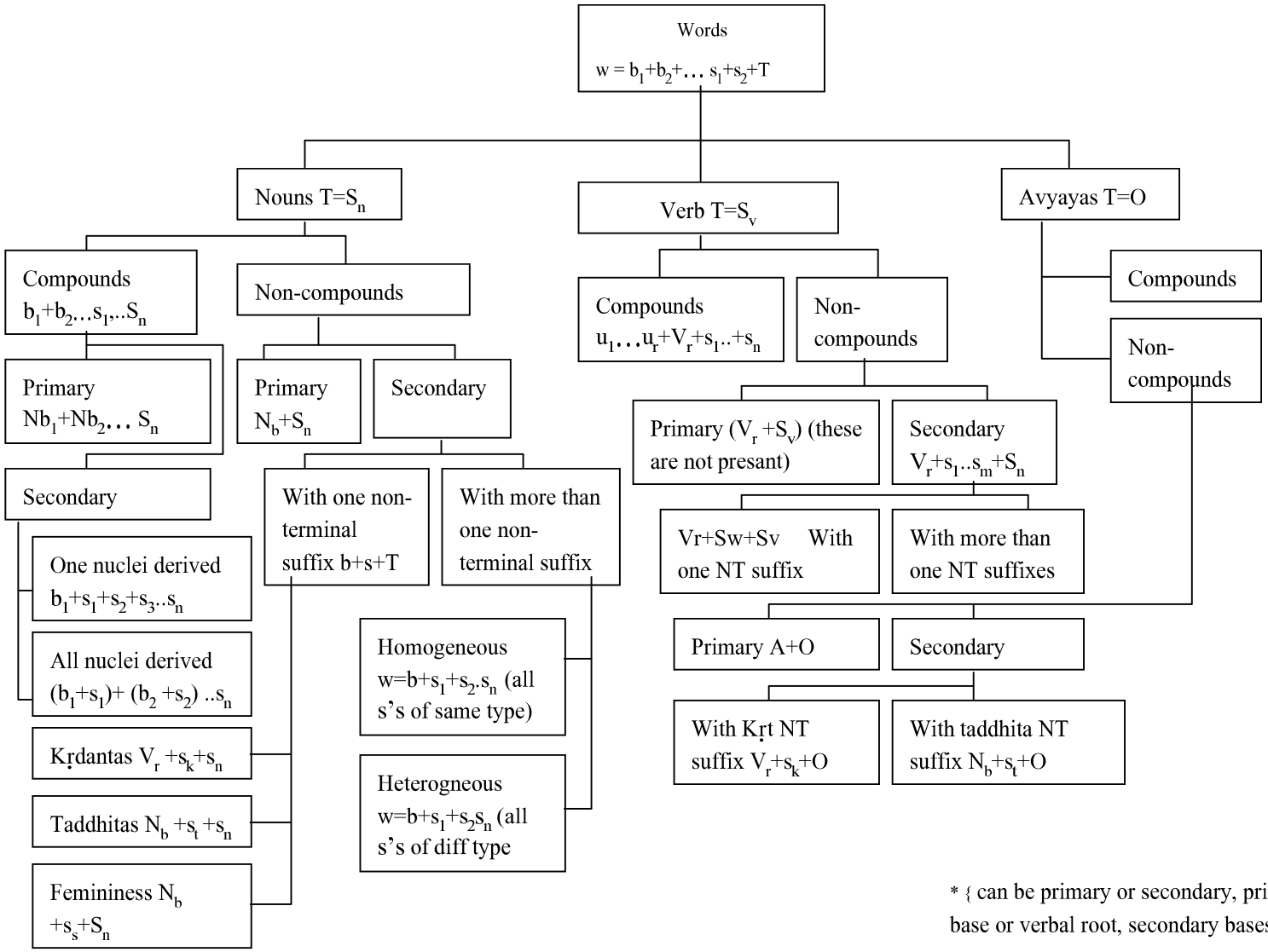


Figure 5.1

SANSKRIT WORD STRUCTURE



LEGEND

W=word;

b= base *

NT=non-terminal;

s=NT suffixes,

s_k=Kṛt suffix;

s_t=taddhita suffix;

s_s=Strī suffix;

s_m=mth suffix;

T =terminal suffix;

S_n = Nominal suffix;

S_v = Verbal suffix;

V_r=verbal root;

u=upasarga;

A=listed avyaya;

N_b=nominal base

N_{bl}=lth nominal base;

sw=Vikaraṇa suffix

* { can be primary or secondary, primary bases contain either Nominal base or verbal root, secondary bases consist NT suffixes}

Figure 5.2

SANSKRIT WORD STRUCTURE

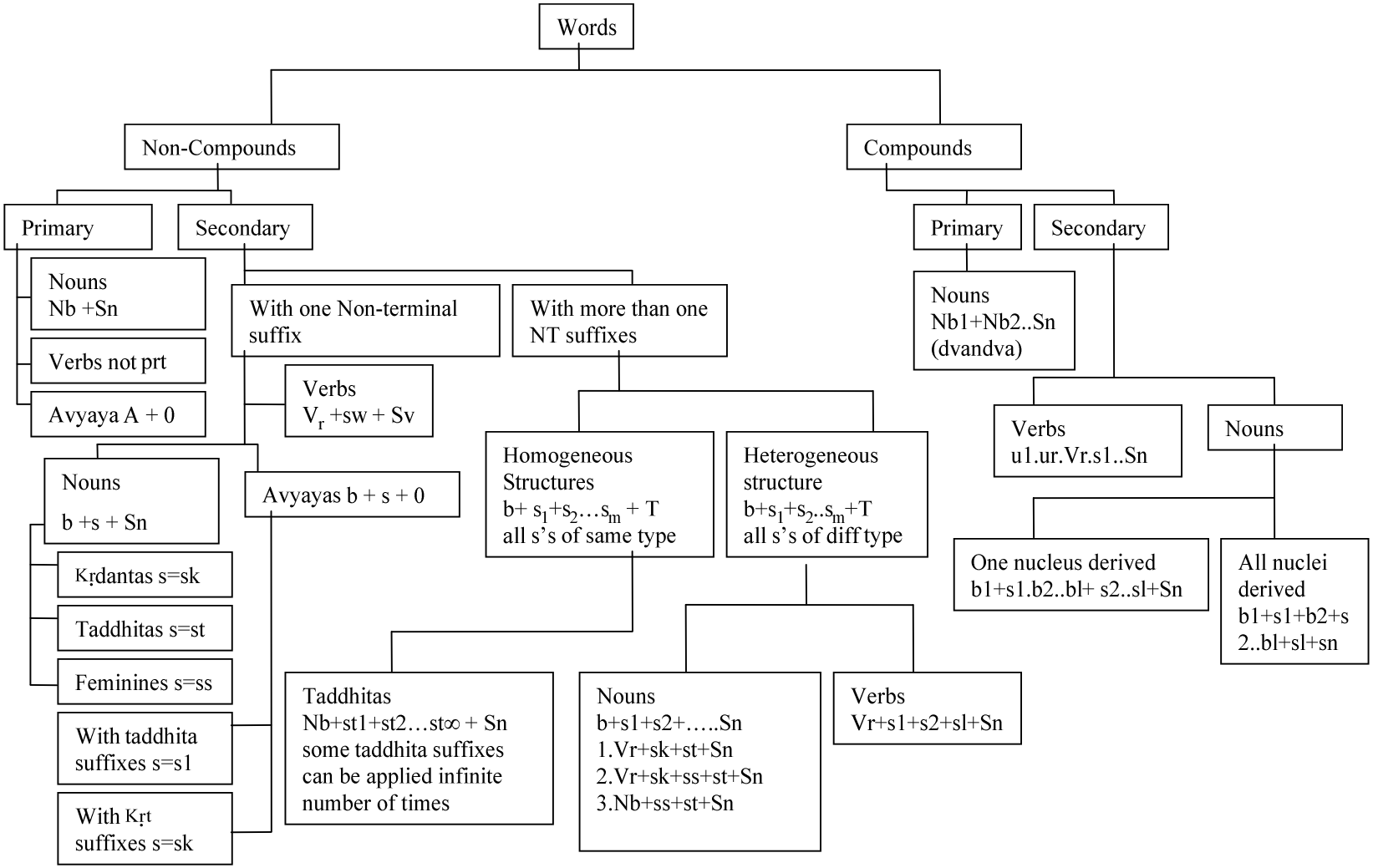
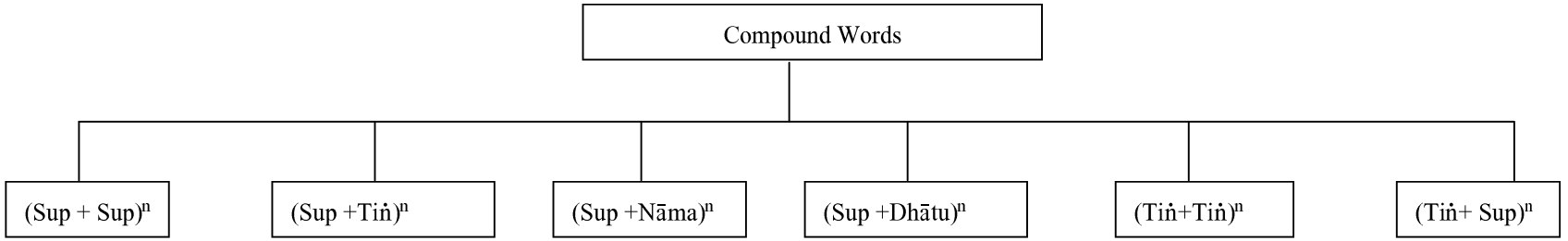


Figure 5.3

SANSKRIT WORD STRUCTURE



Supāṃ Supā Tiṅā Nāmnā Dhātunātha Tiṅāṃ Tiṅā .

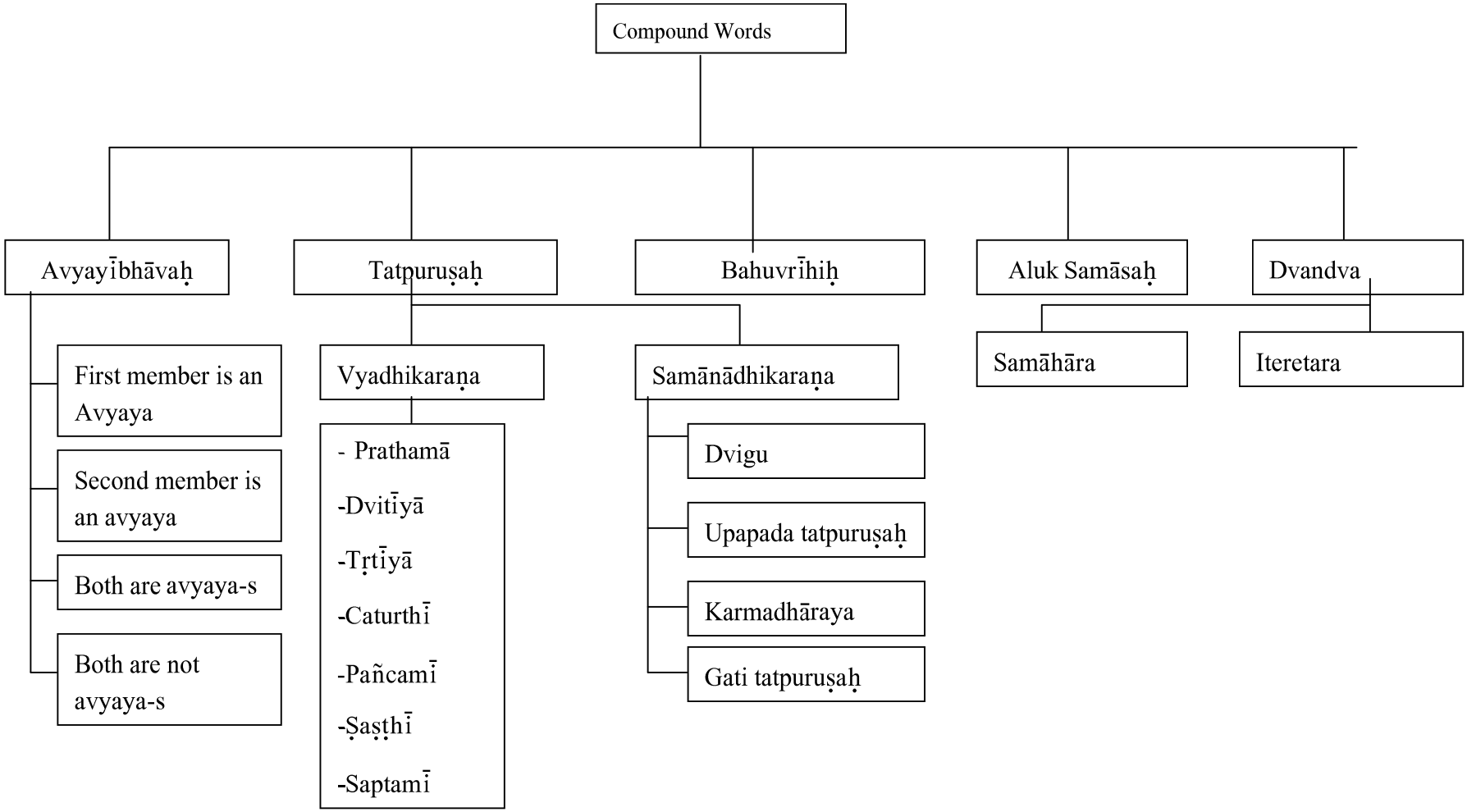
Subanteneti Vijñeyaḥ Samāsaḥ Ṣaḍ vidho Budhaiḥ ..

From Siddhānta Kaumudī by Bhaṭṭōjī Dīkṣita

COMPOUND WORD STRUCTURE -1

Figure 5.4

SANSKRIT WORD STRUCTURE



COMPOUND WORD STRUCTURE -2

Figure 5.5

Chapter 6 : Sentence Analysis

6.1 Kāraka-vibhakti mapping : function assignment to words

When a Sanskrit sentence is input to the system, then after Morphological analysis of each word, the Syntactic analysis further proceeds to assign functions for the word in the sentence. This is accomplished by Pāṇini's Kāraka theory. As we know, the meanings of the words consist of Base meanings and Suffix meanings. The base meanings are to be got from the lexicons, whereas the suffixal meanings are based on the case and number of the suffix (for nouns) and mode, voice, tense, person & number (for verbs) and assign a particular function to the words containing them. Once the Morphological analysis of all the words in a sentence is done, by an Expectancy driven approach, we try to determine the possible functional roles for each word in the sentence applying Kāraka Vibhakti mappings. [See. Figs. 9.8 - 9.21.]

6.2 Syntactic Analysis

For the case and number of the given noun, all possible functional relationships (Kāraka, Sub-Kāraka-s) are related in the Vibhakti Prakaraṇa of Pāṇini's Aṣṭādhyāyī (2.3.1 to 2.3.73). Hence, by checking against these rules (conditions), the possible functional relationships are assigned to the nouns. Next, from these possible Kāraka-s, the conditions contained in Kāraka Prakaraṇa (Aṣṭādhyāyī 1.4.23 to 1.4.56) are applied and possible specific functional relations are enumerated. These happen to be possible Semantic identifications for each noun in the sentence `in isolation'. Now, the verbal identifications are taken and the root meanings are referred to in the root meanings list. The root meaning is taken and its expectancy specification is checked with the available noun identificatons, one at a time. Multiple verb identifications are treated one by one.

Thus, Syntactic analysis not only analyses each word of a sentence morphologically, but also assigns a possible functional role for each word in the sentence.

For example, a sentence like **रामः वनम् गच्छति** (Rāma goes to Forest) is analyzed as under :
Each of the three words has two Morphological identifications.

They are :-
1. Word = **रामः**
i. रामः as a Noun, Masculine gender, Nominative case Singular with राम as base
ii. रामः as a Verb, meaning ` to give ', active voice, Adādi Gaṇa, present tense, Parasmai Padī, first person plural with रा as root.
2. Word = **वनम्**
i. वनम् as Noun, Neutral gender, Nominative case singular with वन as base.
ii. वनम् as Noun, Neutral gender, Accusative case singular with वन as base.
3. Word = **गच्छति**
i. गच्छति as Noun, Masculine gender, Locative case singular with गच्छत् as base.
ii. गच्छति as Verb, meaning ` to go ', active voice, Bhvādi Gaṇa, present tense, Parasmai Padī, third person singular with गम् as root.

Thus, in the above sentence, the first and the last words can have one of the two roles of verb and noun. The second word could be object or noun.

Now, we examine both the candidates for the role of the verb. The first word as verb requires an agent to be 'us' (since the verb is in first person plural). With this condition, the last word becomes superfluous, also as the verbal root then means ` to give ', there is a want of a recipient which is not found in the sentence. Hence the identification considering the first word as verb is not possible.

With the last word as verb, all the expectancies will be met. Here the root meaning is ` goes ' and the other two words fulfill the expectancies generated by this verb. The first word रामः forms the primary agent for the activity of going and the second word वनम् forms the object for the activity of going (an optional expectancy for the given class of verbs). Hence these identifications with the assigned functions for each word will be the output of syntactic analysis.

The sentences in Sanskrit can be divided basically into two types, *viz.* functional sentences and Non-functional sentences. The functional roles to each words of a sentence can be assigned only for sentences of functional type. The functional sentences denote some activity and these are characterized by activity markers. There are two types of activity markers, *viz.* the verbs and the Kṛdanta-s.

For a functional sentence with a single verb as the activity marker, the procedure outlined above applies. For sentences with more than one verb, firstly, the main verb indicating the main activity is chosen and the same procedure is applied. If the functional sentences do not contain any verbs, then the Kṛdanta-s are taken as the starting point and a similar procedure applies. But for the Non-functional sentence, the word with the Nominative case is taken as the starting point and a different procedure applies. [See Fig. 9.6.]

Chapter 7 : Functional Semantics

After syntactic analysis, we proceed to determine word-senses and disambiguate among options, if need be. This stage utilises an ontology as a discrimination net. The specification for activities in terms of noun functor categories is utilised for the purpose.

Semantic analysis considers the validity of a single sentence (in isolation) in terms of logical compatibility. For example, the sentence "He drinks banana", though syntactically correct, is not a semantically valid sentence, because banana is incompatible for a drinking object. Hence, at this stage, the meaning is assigned to a sentence by mapping it to a logical form. The semantic analysis needs us to determine the particular meaning, which is applicable by invoking semantic fitness criteria, which map grammatical functor categories to logical relations and conceptual classes based on activity expectancies and compatibility conditions. The Naiyāyika relations and ontology are used at this stage. By an expectancy driven approach, we fix up the possible functional role for each word in the sentence applying logical compatibility criteria.

To check for the meaningfulness of a sentence, the logicians have devised three essential criteria called Ākāṅkṣā (Expectancy), Yogyatā (Compatibility) and Āsatti (Proximity or collocation). These are obtained from the base and suffixal meanings of the word.

To check for the Expectancy, the Verb (or activity marker) forms the pivot (in functional sentences). Every activity has a given set of expectancies in terms of functional relationships (Kāraka-s) with certain Noun types. For example, every activity needs to have an agent necessarily.

7.1 Root meanings and activity categorisation

There are about 2000 verbal roots (monosyllabic) listed in Pāṇini's Dhātu-pāṭha. Dhātu-s have illustrative meanings, which have a many-to-many mapping. That is, some Dhātu-s have multiple meanings while many Dhātu-s also have a same/similar meaning. In all, about 690 meanings are used which are listed in Annexure - C.

However, root meanings are subject to modification with the usage of preverbs (Upasarga-s) which are about 20 in number, but may be used in combination. Upto 3 preverbs are frequent in prevalent literature (including Vedic). While their effect on phonetics, accents, morphology etc., are significant, it is most pronounced w.r.t. the root meanings. The well-known saying in this regard is :

उपसर्गेण धात्वर्थः बलादन्यत्र नीयते । प्रहाराहार-विहार-संहार-परिहारवत् ।।

धात्वर्थं बाधते कश्चित् कश्चित् तमनुवर्तते । तमेव विशिनष्ट्यन्यः उपसर्गगतिः त्रिधा ।।

Hence, it is not wrong if one were to say that determining factors for their effect on original (or base) root meanings is very difficult to exhaustively enumerate. Attestation in literature has been the usual approach in this case. Hence, more than heuristic/empirical formulations, precise decision is elusive. Similar is the case of transitivity getting affected by preverbs. While Yāska in Nirukta has broadly given hints on their semantic effects and grouping, it is not comprehensive.

7.2 Activity classification

To check for compatibility, "Conceptual type" information about the related functors is specified. For example, the agent has to be animate for primary agency, since the primary agent is defined as an "independant" doer. Besides agents, transitive activities need an object of specified type. Thus, the essential features (like agent, object, location etc) for an activity can be enlisted along with their type information. Proximity also plays a key role in grouping connected functors and their adjectives, non-functor relation, compound word-formations etc.

There are certain optional functional relations which may be mentioned in the sentence. These are instrument and location for most activities. Point of departure (or fixed reference) in activities other than separation type and the recipient or beneficiary in non-giving type of activities are optional factors. For these factors also, the type of Nouns is fixed in terms of their categories and these are also specified with activities. The residual relations, indeclinables, participles etc. are also to be handled.

7.3 Scheme of specification

We have arbitrarily attempted an *ad-hoc* grouping of root meanings to come up with a preliminary ontology to assist word-sense determination and resolution of conflicts where disambiguation at sentence level is needed. [This effort is prompted by Pāṇini himself using in his grammar rules, verbal stems with specific meanings. Some meanings referred are : गति, हिंसा, निगरण, चलन, भी, त्रा, वारण, रुचि, क्रुध, द्रुह, ईर्ष्या, असूया, बुद्धि, प्रत्यवसान, शब्द, अध्ययन, रुजा, मण्ड, मति, पूजा, अस्ति, इच्छा, ध्रौव्य, पश्यार्थाः. (See Annexure - C.)] The concept of intransitive activities (अकर्मक) also are referred to liberally, but without characterising or listing.

Activity types

Activity is looked upon as denotation of a root (verbal) which is of a continuous nature (even made up of distinct sub-ordinate activities) leading to a unique result which could be a state or an event. On this basis, we analyze the *Pāṇinīya Dhātupāṭha* and list out all activities denoted by verbal roots. After the listing of all the activities, we need to group them into classes which have similar functor requirements, under the heads of transitive, intransitive, and ditransitive activities. However, this is a tough task. Even Pāṇini has not mentioned any clear criteria on which verbal root meanings could be classified owing to the multiple senses possible, effects of preverbs and the natural way in which language gets used allowing for speaker's freedom to denote specific intention.

These specific functor requirements for activities includes the mandatory functor types that must be present, the optional ones that could be additionally there as well as certain functors which could be restricted or inhibited. Here we have the syntactic aspects on one hand like particular cases, specific words etc. and the conceptual types of the objects denoted by these functors (words). Thus, we have various combinations possible of the usage for a given activity. A type check would ensure compatibility at logical level and hence meaningfulness.

There are certain complex factors like an activity meant in some different sense, (E.g.: 'citizen lives' and 'king reigns' take the same root, but existence has different connotation for the two), the object and activity are co-existent (E.g.: 'he dances'), well-known acts, mostly natural (like 'river flows, cloud rains') and unintended situations where speaker just skips some essential details (by fancy or

rhetoric) which can cause *a transitive root to be used intransitively.* This detail is described in : धातोरर्थान्तरे वृत्तेः धात्वर्थेनोपसंग्रहात् । प्रसिद्धेरविवक्षातः कर्मणोऽकर्मिका क्रिया ।।

These factors bring to light an important point that while functional relations are effective in determining meanings unambiguously, they are by no means sufficient to cover all situations. Many activities are not resolvable into pure functor expectancies like emotive and knowledge-state related ones like hating, doubting, revering, etc. without taking recourse to metaphysical, transcendental or philosophical concepts. These are covered in depth under ontological classes in Śāstra-s.

There is a **list of intransitive actions or states** like growing, decaying, fearing, living, dying, shying, existing (or being), standing, keeping awake, sleeping, playing (sporting), liking, glowing (or shining) and their synonymous ones. वृद्धि-क्षय-भय-जीवित-मरणं लज्जा-सत्ता-स्थिति-जागरणम् । शयन-क्रीडा-रुचि-दीप्त्यर्थं धातुगणं तमकर्मकमाहुः ।। Also intransitive actions are : crookedness, eagerness, delusion, effort, langour, decrepitude, power, reviling, quiescence, bursting forth, bathing, crippling/deformity, yawn, deceit, crying, laughing. कौटिल्यौत्सुक्य-भ्रम-यत्न-ग्लानि-जरा-सामर्थ्य-क्षरणम् । शान्ति-स्फुट-मज्जन-वैकल्यं जृम्भण-दम्भन-रोदन-हसनम् ।।

Similarly, in Sanskrit, there are **16 ditransitive activities** (terms in brackets denote concept type of object or goal) such as milking (cow, milk), begging (donor, something), cooking (rice, bath), punishing (guilty, penalty), obstructing (confined, place), asking (someone, query), gathering (fruit, tree), speaking (hearer, topic), instructing (someone, query), conquering (loser, stake), churning (input, output), stealing (victim, item) and leading, carrying, drawing or dragging (thing, place). दुह्-याच्-पच्-दण्ड्-रुध्-प्रच्छ्-चि-ब्रू-शास्-जि-मन्थ्-मुषाम् । कर्मयुक् स्यात् अ**कथितं** तथा स्यात् नी-ह्व-कृष्-वहाम् ।। Activities other than intransitive and ditransitive ones are transitive.

As an illustration, let us consider an activity like *'going'*, which could be transitive or intransitive. All activities necessarily need an agent. The type of the agent differs between activities (E.g.; 'going' could take a sentient or non-sentient being as agent. In the former case, the literal meaning is always applicable, while in the latter case, it is not necessarily so (see with 'car', *'auto*mobile' etc.). For an activity like *'seeing'* the primary sense requires an object having a perceptible form, the agent to be sentient with ocular sense intact, presence of light, absence of obstruction, sense contact etc. Similar descriptions are to be defined for the basic activity categories. Thus, it is clear that more the precision with which we specify the verbal activities, we are able to disambiguate sentences sharply. This means deeper extensions to the ontology.

7.4 Activity Grouping

As the number of activities are quite large, if they can be categorised, then semantic analysis would not only become more systematic but also clearer. This has to be done taking all the roots of Pāṇini's Dhātu-Pāṭha. If all the activities are classified based on the similarities of expectancies, then using the principles outlined in Nyāya Śāstra (of Ancient Indian Sciences), the semantic analysis of an isolated sentence can be carried out. The semantic analysis considers a syntactically valid sentence and decides whether it is semantically valid.

However, for all original roots, the meaning expectancies in terms of functors required and their conceptual type can be defined in a preliminary way. Grouping root meanings to conceptual type expectancies, we have formed **11 major activity classes** like Movement, Perception (sensory

knowledge), Transaction, Displacement, Emotions, Getting, Changing, Hurting, Vocations, Intransitive and Giving. This grouping is arbitrary, but has only expected 'type' of functors as the common factor.

These groups are reasonably handled together with the ontology and functor mappings to precisely establish proper expectancies with ordinary sentences. The detailed specification of mandatory, optional and inhibited functor specification for each activity throws up necessary conceptual category labels which are then provided in the ontology for substances and attributes. While dealing with relations, a corresponding mapping is included under grammatical relations.

An Activity grouping List

Mapping Format : *map1*(#Group, #Group], [#root meaning(s)])
Grouping Format : *class*([#Group, #Group], [#mandatory Kāraka(s)], expected word-type for mand. Kāraka(s), [#optional Kāraka(s)], expected word-type for opt. Kāraka(s)])
Groups : 1 - गति - Movement; 2 - ज्ञान - Perception (sensory knowledge); 3 - भाषण - Transaction; 4 - स्थान - Displacement; 5 - भावना - Emotions; 6 - अर्जन - Getting; 7 - परिणाम - Changing; 8 - हिंसा - Hurting; 9 - क्रियासामान्य - Vocations; 10 - अकर्मक - Intransitive; 11 - प्रतिपादन - Giving;
Kāraka(s) : 1 - Karta; 2 - Karma; 3 - Karana; 4 - Sampradana; 5 - Apadana; 6 - Adhikarana.

map1([11], [6, 76, 263, 526])
map1([10, 8], [405, 524, 570, 590])
map1([10, 4], [447])
map1([10, 1], [188, 192, 255, 302, 307])
map1([10, 9], [98, 284, 329])
map1([10, 3], [558, 561])
map1([10, 7], [38, 563, 458])
map1([10, 2], [246, 371, 197, 168])
map1([10], [8, 10, 117, 253, 406, 12, 25, 359, 386, 448, 449, 26, 229, 495, 327, 639, 41, 46, 289, 54, 200, 303, 57, 58, 92, 163, 559, 59, 60, 317, 468, 75, 85, 86, 88, 202, 506, 103, 106, 668, 107, 630, 113, 120, 123, 176, 272, 273, 339, 669, 125, 127, 536, 131, 134, 133, 140, 143, 156, 148, 259, 260, 310, 429 , 430, 151, 265, 274, 416, 517, 518, 573, 154, 164, 165, 266, 167, 169, 171, 172, 269, 534, 177, 178, 372, 661, 189, 199, 210, 211, 212, 214, 215, 247, 248, 228, 350, 378, 379, 380, 420, 232, 376, 377, 233, 267, 271, 481, 356, 364, 285, 296, 651, 287, 291, 292, 311, 326, 334, 335, 358, 652, 360, 362, 370, 375 , 381, 398, 399, 401, 410, 418, 419, 489, 622, 435, 653, 450, 454, 542, 465, 475, 553, 482, 484, 490, 494, 520, 545, 560, 562, 662, 589, 656, 608, 629, 631, 665])
map1([10, 8], [519])
map1([9, 5], [610])
map1([9], [2, 7, 31, 185, 286, 403, 404, 426, 87, 96, 118, 119, 174, 642, 114, 483, 115, 116, 144, 170, 240, 304, 462, 551, 121, 320, 437, 586, 141, 309, 609, 663, 146, 300, 617, 158, 407, 222, 231, 239, 537, 252, 257, 301, 408, 308, 388, 597, 322, 331, 333, 459, 460, 336, 344, 396, 393, 394, 395, 400, 425, 5 01, 505, 521, 532, 566])
map1([8, 2], [315])
map1([8, 5], [237])
map1([8], [16, 49, 51, 52, 283, 27, 30, 35, 126, 184, 194, 195, 198, 250, 281, 352, 353, 361, 374, 397, 509, 510, 585, 625, 645, 670, 671, 672, 673, 48, 183, 226, 227, 277, 476, 502, 424, 67, 124, 213, 244, 313, 628, 82, 109, 175, 182, 145, 147, 457, 605, 159, 433, 595, 604, 201, 546, 230, 439, 242, 256, 262, 264, 411, 268, 463, 466, 591, 548, 541, 574, 667])
map1([7], [568])

map1([7, 2], [530, 620])
map1([7], [36, 37, 47, 136, 421, 422, 428, 61, 122, 129, 132, 453, 654, 655, 139, 152, 238, 249, 297, 254, 275, 314, 318, 623, 624, 529, 531, 321, 618, 328, 572, 575, 576, 441, 431, 442, 581, 443, 467, 540, 582, 584, 583, 600, 602, 603, 606, 621, 657, 658, 615, 616, 627])
map1([6], [34, 138, 423, 39, 56, 383, 473, 477, 71, 108, 592, 130, 417, 455, 456, 278, 599, 619])
map1([5, 3], [647, 648])
map1([5, 2], [644])
map1([5], [13, 29, 373, 63, 70, 89, 324, 94, 102, 270, 337, 338, 135, 607, 643, 149, 436, 569, 220, 480, 243, 251, 384, 385, 640, 290, 293, 445, 452, 552])
map1([4, 2], [664, 299])
map1([4], [19, 20, 69, 91, 206, 485, 535, 611, 21, 649, 650, 224, 33, 486, 508, 515, 53, 73, 280, 62, 110, 180, 181, 347, 348, 111, 112, 137, 209, 312, 492, 216, 279, 276, 547, 282, 295, 323, 538, 612, 325, 539, 340, 341, 342, 412, 511, 512, 351, 355, 357, 469, 646, 493, 363, 446, 513, 382, 496, 516, 528])
map1([3, 2], [160, 161])
map1([3], [14, 72, 294, 504, 564, 565, 15, 40, 28, 204, 43, 64, 223, 288, 507, 65, 81, 100, 142, 580, 205, 579, 235, 402, 596, 319, 345, 346, 349, 414, 415, 491, 543, 544, 632, 332, 343, 354, 523, 368, 409, 413, 626, 635, 636, 637, 638])
map1([2], [1, 4, 474, 525, 66, 155, 306, 613, 11, 17, 18, 440, 22, 23, 203, 32, 438, 42, 90, 219, 500, 522, 556, 577, 614, 44, 45, 234, 444, 499, 598, 50, 221, 503, 68, 196, 74, 128, 659, 660, 79, 80, 83, 84, 173, 470, 471, 93, 101, 104, 150, 153, 461, 478, 498, 99, 641, 666, 105, 179, 245, 369, 434, 633, 186, 31 6, 571, 557, 207, 208, 218, 464, 487, 488, 236, 497, 261, 387, 298, 305, 365, 451, 514, 549, 674, 427, 479, 527, 578])
map1([1], [3, 5, 9, 55, 24, 587, 588, 77, 78, 392, 389, 391, 554, 555, 95, 567, 97, 157, 258, 162, 166, 187, 190, 193, 191, 217, 593, 594, 225, 241, 390, 330, 366, 367, 432, 472, 533, 550, 601, 634])

class(1, [1], ["प्रत्यञ्"], [2], ["देश"])
class(2, [1, 2], ["प्रत्यञ्", "द्रव्य"], [], [])
class(3, [1, 2], ["प्रत्यञ्", "प्रत्यञ्"], [], [])
class(4, [1, 2], ["प्रत्यञ्", "द्रव्य"], [], [])
class(5, [1], ["प्रत्यञ्"], [2], ["प्रत्यञ्"])
class(6, [1, 2], ["प्रत्यञ्", "प्रमेय"], [], [])
class(7, [1, 2], ["प्रत्यञ्", "द्रव्य"], [], [])
class(8, [1, 2], ["प्रत्यञ्", "प्रत्यञ्"], [], [])
class(9, [1], ["प्रत्यञ्"], [2], ["प्रमेय"])
class(10, [1], ["प्रमेय"], [2], ["कल्प"])
class(11, [1, 2, 4], ["प्रत्यञ्", "प्रमेय", "प्रत्यञ्"], [], [])

7.5 Activity Characterisation

Group 1. Movement

Transitive : Direct, Voluntary, Uniform rate
Intransitive : Indirect, Reflexive, Variable rate

I. Agent : Must
Animate - Primary meaning Agency, Direct, Voluntary
Inanimate & Abstract - Secondary

ii. Object :
- Place denoter (destination) - Transitive
- Time denoter (if Intransitive only)

- Measure of time
- Distance denoters
- Place denoters

If Object - Direction - From reference point
- to reference is 'coming'
- From with prior to is return

Second Object in Causative mode
Meaning - Primary
Activity leading to physical displacement of agent in space - axis
Typically from source to destination, needing finite time.
Secondary - caused by 'Hetu'

Group 2. Knowing

Animate agent : Hunger, Sleep, Fear, Copulation
Sentient : Discrimination, States/Emotions

Sensory Perception : Agent awake
Sound - Hearing - Ears
Touch - Feeling - Skin
Form - Seeing - Eyes
Taste - Tasting - Tongue
Smell - Smelling - Nose
Others - Hands, Legs, Genital, Anus, Mouth.

Inference (Extra sensory)
Intellect : States - Learning, Mistaking, Thinking, Meditating, Considering
Mind + Information related Knowledge (Writing, Reading, Teaching)

i. Agent :
- Animate (in fit condition)/Sentient (Senses intact) - (Human being)
- Sensory perception (Physical)
- Extra sensory (Inference, Belief)
- Mental

ii. Object (subject/topic - any Padārtha) :
State :- Sleeping
- Dream - with only mind agent
- Diseased/Unconscious
- Intention

Distance of Agent from Object and auxiliaries for particular sense organs.

Group 3. Verbal Communications:

Speaking - Transitive; (Making Sound - Intransitive)
General Info. Exchange - Conversation; Expressing Feelings; Technical/literary - descriptive
1. Agent :
Speaker and Addressee - both Animate/Sentient
Different aspects - Praising, Blaming, Ordering, Requesting, Sharing info., Debating, Arguing, Asking ...

Speaker should be Expressive (not dumb; presupposes - knowing what and how to communicate)
2. Object :
Addressee should have hearing faculty intact
A second object (topic - not Animate) is possible
3. Instrument :
Mind - in all knowing frames
Sense organ is Mouth in speaking.

A preliminary ontology derived from the rules of Kāraka and Vibhakti sections of Aṣṭādhyāyī is given in Fig. 9.29-9.32. A preliminary semantic lexicon is found at Annexure - F.

7.6 Conflict resolution and disambiguation

Basically, a set of Sentences denoting a single notion or topic are said to form an extended sentence, set of connected sentences or discourse and it is supposed to convey a unique sense. In the light of such definition of a discourse, there may be occasions where individual sentence meanings contradict the supposed unique sense to be conveyed by the paragraph of which they are members. The sources for such a contradictory meanings arise from the processes of multiple denotative powers of words; even sentences.

Consider the three sentences : एषः स्थाणुः । एषः सेचनात् वर्धते । सिक्तोऽयं फलं प्रयच्छति । Here, from the Amarakośa, स्थाणू रुद्र उमापतिः , the word स्थाणुः can refer to Rudra or stem of a tree. The three sentences then accordingly pertain to stem or Rudra and respectively mean, by Secana, watering or ceremonial bathing (Abhiṣeka) and by Phala, fruit or desired results. This is an illustration of discourse level ambiguity, which needs Mīmāṃsā Nyāya-s to resolve.

The syntactic analysis invariably gives multiple identifications for almost all words due to the fact that identical word formations are possible through different base-suffix combinations denoting different semantic contents. These are to be reconciled to arrive at an acceptable sense for the contradicting words or sentences. There is a hierarchy of knowledge sources pertaining to this discussion at six levels as follows :

(1) श्रुतिः - Direct statement *i.e.* , which cannot be mistaken or a given thing; for example, a statement that we are going to discuss mathematics, where the topic is unanimously understood with a particular connotation. This level is context-invariant and is the strongest *i.e.* , will not be affected by other levels.

(2) लिङ्गम् - Expressive power of words at absolute and relative levels; as for example, when we say all expressions have an LHS and an RHS. Here, the word expression should denote a mathematical equation and not just any sentence, though expression can refer to both. The word all here also can only refer to the domain under consideration. This aspect of taking the restricted (relative) potency of words to denote objects at absolute level is one way of accounting for context.

(3) वाक्यम् - Sentence level denotation is further constraining word level potencies and includes effects of other words collocated in then same sentence by way of sentential expectancies, for example, $(X+2)^2 = 25$, fixes the value of X based on the conditions specified by the other components of the expression. With linguistic expressions, functor expectancies condition the meaning of a dubious

word. In the previous case, a potent word had known alternative meanings of which one of the other gets picked depending on whether the absolute or relative sense is permissible, because of the direct statement.

(4) प्रकरणम् - A subset of sentences under the total topic which may focus on a particular aspect which can guide resolution of ambiguous words or sentences comprising the set. Here we redefine the original topic into smaller connected sub-topics and this becomes a miniature discourse. For example, Consider the four sentences :

John reads.
Mary also.
Bill jumps.
Mary also.

Here, 'Mary also' occurring in a particular sequence after certain sentences conveys different meanings. Sentential expectancies are involved here.

(5) स्थानम् - The position of a particular statement in a discourse also lends help in disambiguation to some extent. This roughly corresponds to the chapters, sections, or paragraph headings of a book, *i.e.* , by virtue of some sentence being present under some heading, we inherit certain meaning for part or whole of the sentence.

(6) समाख्या - Intuitive links of certain nomenclatures with certain meanings also may be restored to for disambiguation when any other help is not forth coming. For example, A-Z of something is meant to denote everything by the similarity of the nomenclature abbreviating all alphabets of a language.

Here, we see a gradual decrease in certainty of meaning and increasing dependence on external knowledge for sense disambiguation at word, sentence, context, location or word level. Hence we say the latter levels are weaker due to the dependencies. This also establishes procedures for resolving conflicts between the possible permutations *i.e.* , 15 in this case, besides conflicts among sentences of the same type in terms of meanings.

Word-sense disambiguation

While the primary sense or literal meaning of words become in-admissible due to contextual constraints, a secondary or extended sense is preferred during analysis. This aspect has been deliberated upon in all the three Śāstra-s.

Among Grammarians, Bhartṛhari points out the following aspects as the basis for deciding such specific meanings: close connection, separation, association, contrariety, purpose, context/topic, significant expression, presence of other words, power/potency, propriety, particular region, time factor, notion of individual object and accents etc. (वाक्यात् प्रकरणात् अर्थात् औचित्यात् देश कालतः । शब्दार्थाः प्रविभज्यन्ते न शब्दादेव केवलात् ।। संसर्गो विप्रयोगश्च साहचर्य विरोधिता । अर्थः प्रकरणम् लिङ्गं शब्दस्यान्यस्य सन्निधिः ।। सामर्थ्यमौचिती देशः कालो व्यक्तिः स्वरादयः । शब्दार्थस्यानवच्छेदे विशेषस्मृति हेतवः ।। - Vakyapadiya 2.314-316).

Close connection : सवत्सो गौः - Cow with calf denotes milch cow.
Seperation : विवत्सो गौः 'Cow without calf denotes barren cow.

Association: रामलक्ष्मणौ - Rāma and Lakṣmaṇa means Daśaratha's son Rāma (for doubt about Rāma's identity).

Contrariety: रामार्जुनगतिस्तयोः Rāma and Arjuna denotes Paraśurāma and king Kārtavīrya Arjuna because of the relation between them not the popular Rāma and Arjuna.

Purpose: अञ्जलिना जुहोति - The word Añjali means folded palms related to wishing while it denotes opened palms related offering oblations in sacreficial fire.

Context/topic : सैन्धवमानय - Get Saindhava - denotes a horse of Sindhu origin in going to battle field while it denotes common salt while eating.

Significant expression: Anointed pebbles means anointed with ghee due to the presence of a sentence of a sentence commending ghee in the context.

Presence of other words: Rāma, the son of Jamadagni means Paraśurāma.

Power/potency: Find a good match means cutter as an instrument of cutting.

Particular region: The Lord reigns here means king, if the place is a capital city, else it could mean God.

Time factor: Citrabhānu glows - denotes Sun in day time and fire during night.

Notion of individual object: Mitra helps - denotes Sun or a friend depending on the gender of the word.

Accents: Accents define word meanings functionally, particularly in the case of compounc words. Where the same word- form could result from different compounding processes, their accents are necessarily different. Thus, the word 'स्थूल-पृषती' (fat-dotted cow) can denote a cow with large dots or a fat cow with dots depending on the accent of the word. Similarly, इन्द्र-शत्रु (Indra's enemy) can denote killer of Indra or one being killed by Indra, by accent variation.

In Nyāya-Sūtra (2.2.62) सहचरण-स्थान-तादर्थ्य-वृत्त-मान-धारण-सामीप्य-योग-साधनाधिपत्येभ्यः ब्राह्मण-मञ्च-कट-राज-सक्तु-चन्दन-गङ्गा-शाटकान्न-पुरुषेष्वतद्भावे अपि तदुपचारः

[In the case of (a) 'Brāhmaṇa', (b) 'Mañca' (platform) (c) 'Kaṭa' (mat) (d) 'Rājan' (king) (e) 'Saktu' ('flour) (f) 'Chandana' (sandal) (g) 'Gaṅgā', (h) 'Śaṭaka' (cloth) (i) 'Anna' (food) (j) 'Puruṣa' (man) there is secondary (indirect) application, due respectively to - (a) association (b) location (c) purpose (d) behaviour (e) measure (f) containing (g) proximity (h) connection (i) cause and (j)sovereignty], we find the mention of various factors to be considered, while dealing with figurative usage.

(1) Association: यष्टिकां भोजय - Feed the stick, means feed the Brahmin holding the stick.

(2) Location: मञ्चाः क्रोशन्ति - Platforms are shouting, means people on the platform shout.

(3) Purpose: कटं करोति - He makes a mat, means he aims at making the mat.

(4) Behaviour: यमः राजा - Yama (Chastiser) means a king by the similarity in punishing.

(5) Measure: आढकसक्तवः - Bushel of flour, means flour measured by a bushel.

(6) Containing: तुलाचन्दनम् - A balance-sandal, means sandalwood held in the balance.

(7) Proximity: गङ्गायां गावश्चरन्ति - Cows are grazing in the Ganges, means cows grazing in the adjoining lands of the Ganges.

(8) Connection: कृष्णाः शाटकः - A black (cloth), means a cloth marked with blackness.

(9) Cause: अन्नं प्राणाः - food is life, means that is the cause of life.

(10) Sovereignity: अयं पुरुषः कुलम्; अयं गोत्रम् - Mr. X is the clan, means X is the head of the clan.

The Mīmāṃsaka-s have elaborately described the factors governing sentential and discourse coherence in terms of the six major factors of commencement, conclusion, repetition, novelty or utility, illustration and propriety.

(उपक्रमोपसंहारौ अभ्यासोऽपूर्वता फलम् । अर्थवादोपपत्ती च लिङ्गम् तात्पर्य निर्णये ।।)

The topic of discussion/statement of intention or goal is meant by the factor commencement, which raises expectancy on this topic to be satisfied by the following sentences.

The conclusion signifies the lower limit of the discourse. In formal situations these two aspects are invariably present.

Mid-way through discourses maintaining continuity of the topic is accomplished by repeating or re-stating the initial goal. This may also be in the form of emphasis.

Novelty requires each sentence of a discourse to contribute some unknown aspect regarding the topic so that every sentence remains significant. This is in reply to the different expectancies accumulating. Utility deals with the purpose of each sentence meaning to be unique.

Illustration serves to amplify or explain the content of the discourse through instances, narration, supporting facts etc.

Propriety enjoins the logical compatibility of sentential meaning and inter-sentential relation.

Chapter 8 : Śābda-Bodha (Verbal import)

8.1 Concept of Śābdabodha

The ancient Indian approach to language analysis is based on the concept of viewing a sentence to be springing from the semantic message the speaker wishes to convey (in contrast to the phrase-structure model and the consequent binary parsing techniques for the English-like languages). **Śābda-bodha** is the description of the comprehension in a hearer from the utterance of a sentence. This very roughly corresponds to 'paraphrase' which contains the structural description of the sentence and makes the various elements in the meaning explicit, so that the original sentence is rendered unambiguous. This when represented, can capture the disambiguated meaning of the sentence and help inference.

Significant aspect in Sentential import

The मुख्य विशेष्य (emphasis or the significant factor) in sentential import is perceived differently in the three systems and hence, the format of the paraphrase is different. The Logicians treat the प्रथमान्तार्थ (Agent) as central in sentence meaning, while Grammarians view the धात्वर्थ (verbal meaning, *i.e,* activity) and मीमांसकs feel the भावना (efficient force or intended act) as prominent. However, there are specific situations where each of these are apt and hence, we take all these factors suitably.

How to represent meaning (or Knowledge, in general) in the computer? This is the crux of the whole exercise and we describe the ancient Indian method in this regard. This is necessarily empirical in nature. We develop a classification scheme for all substances in a structured form (पदार्थ विभागः - Ontology) to explain meanings and facilitate 'machine understanding'.

8.2 Fundamental Metaphysical categories

All knowable things are divided into **प्रमाण (** Means of valid knowledge) and **Prameya** (Object of knowledge). There are three प्रमाणs, *viz.* प्रत्यक्ष (**Perception**), अनुमान (**Inference**) and शब्द (**Verbal testimony**), and these include उपमान (comparison), अर्थापत्ति (presumption), अनुपलब्धि (non-apprehension) etc. suitably. Perception is knowledge generated by the direct contact of sense organs with objects. Inference is the knowledge of probandum based on probans established by logical concomitance. Verbal knowledge is the one arising from spoken words or verbal statements through the comprehension of the connected meaning. Perception and Verbal testimony can be stored as facts and inference as if-then rules in the machine.

Prameyas are classified under two broad headings as द्रव्य (**substance**) and अद्रव्य (**non-substance**). Substance is defined as 'अवस्थाश्रयम्' (having states or substratum for modifications). This is of two classes as जड (material) and अजड (non-material or spiritual). प्रकृति (Cosmic matter) and काल (time) are material substances. Spiritual substances could further be प्रत्यक् (self-revealed) or पराक् (revealed to others). Sentient beings are self-revealed.

Non-substance is defined as one which cannot be substratum to something else. It should be integrally (inseparably) related to a substance and not admit conjunction with another quality. There are ten basic metaphysical attributes, *viz.* sattva, rajas and tamas - representing whatever is fine or light, whatever is active and whatever is heavy - the three attributes of Cosmic matter; शब्द (sound), स्पर्श (touch), रूप (colour), रस (taste) and गन्ध (odour) - the five attributes of gross elements; संयोग (conjunction) and शक्ति (potency). These cover the innumerable other qualities belonging to substances in conformity with the प्रमाणas. Potency is the special property of a causal substance by virtue of which, it brings out the necessary effect. Conjunction is a form of relation which holds between two physical objects.

Substance and attribute are distinct but अपृथक् सिद्ध (integrally or inseparably) related. There are two types of entities in the Universe as those that are separable and those which are not. In the former case, the relation is conjunction, which is external, ceasing to exist the moment the objects are separated. In the latter case, we conceive a unique inherent or internal relation, which is the very nature of the relata lasting as long as the relata exist. Cause and effect are different states of one and the same substance. In terms of these, we define all things in the Universe. This classification is based on Tattva मुक्ता कलापa of वेदान्तa देशिकa, a Logician-philosopher of 13th century A.D. [19].

Heuristic rules are devised to resolve cases where no verb, more than one verb or non-functor words are present. These derive mainly from 'Laukika न्यायas', which are akin to formulation of common sense. Besides, मीमांसा also has clearly defined criteria for relative priority among various means of knowledge like, श्रुति (express mention), लिङ्ग (word meaning or explicit indication), वाक्य (syntactic connection, प्रकरण (context = उपक्रम - beginning, उपसंहार - end, अभ्यास - emphasis, अपूर्वता - novelty, फलम् - purpose and उपपत्ति - propriety), स्थान (relative position) and *समाख्या (name).*

Considering a passive sentence like 'मात्रा खाद्यते' ('being eaten by mother' or 'tablet is being eaten'), we get the first word to be nominative or instrumental case. When it is taken as instrument case, it denotes agent; while as nominative, it denotes object. Here, both the answers are output as valid as these words alone cannot disambiguate the sense any further. More examples are given in [23]. This shows the utility of the base-affix analysis of words.

8.3 case-relations used

It is seen that various relations are at play, due to context, when inflected words are used. To be able to pick the right relationship between the cases and the functors intended, the many-to-many K-V mappings are resorted to. The following list gives the illustrations of sentence construction for different case-relations being meant.

There are 7 different relations for Accusative Case, 11 for Instrumental, 11 for dative, 9 for Ablative, 17 for Genitive and 19 for Locative cases. Of these, 48 distinct shades of relations are seen (removing identical relations in different cases. Fig. 8.1 lists these. These, properly grouped and characterised, would help sense-determination determination and possibly, disambiguation also, when multiple options are valid. For this, we need a lexicon with semantic category info. stored (using the

ontology derived from these noun categories, relation-types and activities) alongwith lexical details for the vocabulary. We have appended a sample preliminary semantic lexicon used in **DEŚIKA** system.

The overlap between various shades of senses is difficult to avoid, as precise definitions of relations and exclusive contexts for them are impractical to comprehensively define. However, the grammatical and metaphysical categories help in many real-life situations to extract the possible meanings with the base/suffix meanings also considered in a systematic manner.

The 48 exclusive or distinct relations from the above are :

(1) निष्ठत्वम् (2) विषयत्वम् (3) विशेष्यत्वम् (4) प्रकारत्वम् (5) प्रतियोगित्वम् (6) निरूपितत्वम् (7) व्यापकत्वम् (8) कर्तृत्वम् (9) करणत्वम् (10) ज्ञानज्ञाप्यत्वम् (11) अभेदः (12) साहित्यम् (13) समवेतत्वम् (14) समानकालिकत्वम् (15) अवच्छिन्नत्वम् (16) उद्देश्यत्वम् (17) तृप्तिप्रयोजकत्वम् (18) निष्पत्तिप्रयोजकत्वम् (19) विकारित्वम् (20) वृद्धिप्रयोजकत्वम् (21) आश्रितत्वम् (22) स्वापहरणेच्छाप्रयोज्येच्छाविषयत्वम् (23) प्रयोजकत्वम् (24) इच्छाधीन इच्छाविषयत्वम् (25) अवधिमत्त्वम् (26) जन्यत्वम् (27) स्वकर्तृकोच्चारणाधीनत्वम् (28) आरम्भः (29) पर्यन्तः (30) तदपेक्षत्वम् (31) वृत्तिः (32) स्वामिता निरूपित स्वत्वम् (33) प्रतिपादकत्वम् (34) उच्चरितत्वम् (35) कर्मत्वम् (36) अवयवत्वम् (37) स्वसमभिव्याहृत पदार्थतावच्छेदक जाति शून्य षष्ठ्यन्तपदार्थ व्यावृत्तत्वम् (38) आधेयत्वम् (39) व्यापकत्वम् (40) अवच्छेद्यत्वम् (41) घटकत्वम् (42) प्रतिपाद्यत्वम् (43) सामानाधिकरण्यात्मक वैशिष्ट्यम् (44) पूर्वकालिकत्वम् (45) उत्तरकालिकत्वम् (46) स्वविषयकेच्छाधीनत्वम् (47) स्व समभित्याहृत पदार्थतावच्छेदक जातिशून्य सप्तम्यन्तपदार्थ व्यावृत्तत्वम् (48) कार्यकारण भावः.

8.4 Examples
Vaiyākaraṇa Śābda Bodha

Consider the sentence - 1. रामः अन्नम् पचति - Rāma cooks rice. It should first be analysed into the constituent lexical and grammatical elements (altogether six such elements here) as

- राम स् अन्न् अम् पच् (अ)ति -

It is the activity, which is presently taking place, which is tied to the substratum which is identical with one (single) Rāma and which is conducive to the softening located in the substratum which is identical with rice. This follows the old Vaiyākaraṇa-s.

The New Vaiyākaraṇa-s have a slightly different structural description. They would say :
The activity of cooking occuring in present time has an agent (qualified by Rāma as its agent) which is identical with Rāma and qualified by rice which is connected with it by way of being its object.

2. रामः काष्ठेन अन्नम् पचति - Rāma cooks rice with fire - wood.

रामः काष्ठेन पाकेन अन्नम् करोति - Rāma makes rice by cooking by firewood.
The final structural description of the knowledge would be - It is a making function, which is happening at present, which is done through the instrumentality of cooking (*i.e.*, qualified by cooking), which (cooking) has rice as its object - goal (Karma) and is done through the instrumentality of firewood, and it is the making function qualified by Rāma as its agent.

3. रामः महानसे काष्ठेन अन्नम् पचति - In the kitchen Rāma cooks rice with firewood.

"It is Rāma who is qualified by the effort that is conducive to cooking, which cooking has rice as its object - goal, (*i.e.*, qualified by the 'object - hood' in rice), which is qualified by instrumentality in firewood, and it is the same Rāma who is qualified by being located in the kitchen".

4. **पश्य, मृगो धावति** - "Look, the deer runs".

The Vaiyākaraṇa-s argue that since this is to be treated as a single sentence with one principle qualificand in the content of the verbal knowledge, their own analysis with the meaning of the verbal stem as the principal element provides a better structural description.

"It is the seeing (by you) which is at the present time and which has the running as its object, which running belongs to the deer as its agent".

Here since the object (Karman) is expressed by a verbal stem (dhāvana = run) and not by a nominal stem, one would not expect a second ending "am" (those endings are specifically meant for nominal stems).

कानिचन नैयायिक शाब्दबोध उदाहरणानि –

1. *णिजन्त प्रयोगः* । उदा - **चैत्रः मैत्रं गमयति** । - Caitra makes Maitra go.

पुंस्त्वैकत्व-विशिष्ट-मैत्रनिष्ठ-उत्तरप्रदेशसंयोगानुकूल-वर्तमानकालिक-कृत्येकत्वपुंस्त्वाश्रयः चैत्रः ।

Caitra is the substratum of activity (1) in present time favourable to an activity (2) which in turn is favourable to an unison of different place and has Maitra for its object. Śābdabodha identifies two activities: (1) Going to Maitra (2) The activity causing the going of Maitra for which Caitra is responsible.

2. **शिवः अर्जुनाय शस्त्रं अयच्छत्** । - Śiva gave the weapon to Arjuna.

शस्त्रकर्मक-अर्जुनोद्देश्यक-स्वस्वत्व-निवृत्तिपूर्वक-परस्वत्व-आपादनानुकूल-भूतकालिक-व्यापाराश्रयः शिवः ।

Śiva is the substratum of an activity favourable to transfer of (Tvā of hood; possessivehood) ownership pertaning to past time which (activity) has Arjuna as the beneficent and Śastra (weapon) as object.

3. **रामः अरण्ये बाणेन वालिनं जघान** । - Rāma killed Vālī in the forest with an arrow.

अरण्याधिकरणक-बाणकरणक-वालिकर्मक-हननानुकूल-भूतकालिक-व्यापाराश्रयः रामः ।

Rāma is the substratum of activity favourable to killing pertaining to past time, which has Vālī as object and location of the activity is the Araṇya (forest) and Bāṇa (arrow) is instrumental in the activity.

4. **चैत्रः रामस्य गृहं गच्छति** । - Caitra goes to Rāma's house.

रामसम्बन्धि-गृहाभिन्नोत्तरदेश-संयोगानुकूल-वर्तमानकालिक-व्यापाराश्रयः चैत्रः ।

Caitra is the substratum of an activity pertaining to present time, favourable to unison of a place which is not different froom the house possessed by Rāma.

5. **यदा शबरी रामाय फलं अयच्छत् तदा सः तत् अखादत्** । - When Śabarī gave Rāma a fruit he ate it.

शबरीकर्तृक-फलकर्मक-रामोद्देश्यक-स्वस्वत्वनिवृत्तिपूर्वक-परस्वत्वापादनानुकूल-उत्तरकालीन-गलबिलाधः-संयोगानुकूल-भूतकालिक-व्यापाराश्रयः रामः।

Rāma is the substratum of an activity (1) pertaining to past time, favourable to eating, which takes place after an activity (2) favourable to transfer of ownership which (activity (2)) has Śabarī as the agent, Phala as the object and Rāma as the beneficent.

6. **त्वं कुत्र गच्छसि** ? - Where are you going ?

कीदृशस्थलोद्देश्यक-उत्तरदेशसंयोगानुकूल-वर्तमानकालिक-व्यापाराश्रयः त्वम् ?
You are the substratum of an activity, pertaining to present time, favourable to unison of which place (?)

7. **अहं विश्वसिमि यत् कृष्णः भगवान्** । - I believe that Kṛṣṇa is God.

'देवाभिन्नः कृष्णः' इत्याकारक-विश्वासानुकूल-वर्तमानकालिक-व्यापाराश्रयः अहम् ।
I am the substratum of an activity pertaining to present time favourable to the belief that Kṛṣṇa is none other than God.

8. **कृष्णः अर्जुनं समाश्वासयत्** । - Kṛṣṇa comforted Arjuna.

अर्जुनोद्देश्यक-समाश्वासनानुकूल-भूतकालिक-व्यापाराश्रयः कृष्णः ।
Kṛṣṇa is the substratum of an activity pertaining to past time which resulted in comfortment of Arjuna.

9. **चैत्रः सौहार्दात् देवदत्ताय ओदनं घटे अग्निना पचति** । - Out of friendship Caitra cooks rice for Devadatta in a pot with fire.

अग्निकरणक-घटाधिकरणक-सौहार्दकरणक-देवदत्तोद्देश्यक-ओदनकर्मक-विक्लित्त्यनुकूल-वर्तमानकालिक-व्यापाराश्रयः चैत्रः ।
Caitra is the substratum of an activity pertaining to present time favourable to softening of rice which has Devadatta as the beneficent, which (activity) takes place in the pot with fire as instrument and which (activity) has friendship as the cause.

10. **देवदत्तहस्तेन पुत्रस्य गोः ब्राह्मणाय मोहात् दीयते** । - Son's cow is given by Devadatta's hands to Brāhmaṇa out of affection.

पुत्रसम्बन्धि-गोकर्मक-हस्तकरणक-ब्राह्मणोद्देश्यक-मोहकरणक-स्वस्वत्वनिवृत्तिपूर्वक-परस्वत्वापादनानुकूल-वर्तमान-कालिक-व्यापाराश्रयः देवदत्तः ।
Devadata is the substratum of an activity pertaining to present time, favourable to transfer of ownership which has Brāhmaṇa as the beneficent, Devadatta's hand as the instrument, love as the cause and the cow possessed by his son as the object.

Points of difference in perception among Vyākaraṇa, Nyāya and Mīmāṃsā views:

A. **Failure of nominative qualificand**

1. पश्य, मृगो धावति - See the deer runs.

1a. मृगो धावति, तम् पश्य - Deer runs, see him.

1b. धावनानुकूल-कृतिमान् मृगः - The deer possesses the volitional effort conducive to running.

and मृगकर्मक - दर्शनाश्रयः त्वम् - You resort to the act of seeing in which deer is the object.

1c. धावनानुकूल - कृतिमन् - मृगनिष्ठ - कर्मता - निरूपक - दर्शनाश्रयः त्वम्

You resort to the act of seeing, determined by the objective status of deer, possessing the volitional effort, conducive to running.

1d. धावन्तं मृगं पश्य - See the running deer.

1e. मृगाभिन्नैकाश्रय-वृत्ति-उत्तरदेशसंयोगानुकुल-वेगजन्य-विलक्षण-व्यापारकर्मक-ईक्षणानुकूलव्यापारः त्वदभिन्नैकाश्रय वृत्तिः

The action in favour of seeing, having the object relating to an (extraordinary) action borne out of speed conducive to contact with the next region, resorting to deer only, is inherent in you.

1f. धावनकर्तृ - मृगकर्मक - दर्शनाश्रयः त्वम् - You resort to the action seeing, the object of which is the deer, the agent of running.

2. शृणु, गर्जति मेघः - Hear the cloud roars.

3. नटो गायति, शृणु - Hear, the dancer sings.

4. घटोऽनश्यत्, पश्य - See, the pot destroyed.

2a. गर्जनाश्रय - मेघ - कर्मक - श्रवणाश्रयः त्वम्
You resort to the action 'hearing', to which the cloud is related as an object resorting to roaring.

3a. गानाश्रय - नटकर्मक - श्रवणाश्रयः त्वम्
You resort to the action 'hearing', to which the dancer is related as an object resorting to singing.

4a. अतीतोत्पत्तिक - नाशप्रतियोगिक - घट - कर्मक - दर्शनाश्रयः त्वम्
You resort to the action 'seeing', to which the pot is related as an object which is the counter positive of destruction produced in past.

2b. मेघाभिन्नैकाश्रयवृत्ति - गर्जन - कर्मक - श्रवणं त्वदभिन्नैकाश्रयकम्
The action hearing, to which roaring is related as an object belonging to cloud only, resorts to you.

3b. नटाभिन्नैकाश्रयनृत्ति - गान - कर्मक - श्रवणं त्वदभिन्नैकश्रयकम्
The action hearing, to which singing is related as an object belonging to dancer only, resorts to you.

4b. घटाभिन्नैक - प्रतियोगिताक - अतीतोत्पत्तिक - नाश - विषयकदर्शनम् त्वदभिन्नैकाश्रयकम्
The action seeing, in which the content is destruction produced in past, the counter positive of which is pot, resorts to you.
[2a, 3a, 4a - Nai. Śā. ; 2b, 3b, 4b - Vai. Śā.]

5. शंखाः श्रूयन्ते - The conches are heard.

5a. शंखाभिन्न - बहुत्वावच्छिन्न - कर्मवृत्ति - श्रवणानुकूलो व्यापारः (अस्ति, भवति)
The action, in favourable of hearing, belongs to object delimited by manyness resorting to conch.

6. त्रयः कालाः (सन्ति) - Three times (exist).

6a. त्रित्व - विशिष्टाभिन्नाः कालाः (प्रतीयन्ते, ज्ञायन्ते) - Times qualified by the number three.

B. **Failure of action-predominance** (Taddhita suffixes)

7. वैयाकरणः - Grammarian, i.e, person who knows or reads grammar (4.2.59).

8a. रथ्यः - horse, who bears or carries the chariot (4.4.76).

8b. युग्यः - bullock, who bears or carries the yoke (4.4.76).

8c. प्रासङ्ग्यः - a trainee, who bears or carries the yoke-like weed for cattle (4.4.76).

C. **Particle as the main qualificand**

9. भूतले न घटः - There is no pot on the ground.

9a. भूतलानुयोगिक-घटप्रतियोगिकः अभावः - The absence in which pot is counter positive and ground is locus.

10. शिवाय नमः - Obeisance to Lord Śiva.

11. अलंकरोति - Embellishes.

12. आविष्करोति - Discovers.

13. नमस्करोति - Hails / Salutes / Greets.

14. पुरस्करोति - Makes forward.

15. साक्षात्करोति - Meets.

16. तिरस्करोति - Discards.

17. हरये नमः - Obeisance to Hari.

18. स्वस्ति भवते - Blessings to you.

17a. स्व - प्रयोक्तृ - पुरुष - कर्तृक - नमः - प्रयुक्त - त्यागः - Offering by a person who uses the very word.

18a. भवदीय - हितविषयिणी - मदिया - इच्छा - My wish (embedded in the meaning of the word svasti) in which your welfare is the content.

D. **Problem of meaning of a word**

19. पुत्रेण सह आगतः - (He) has come with son.

19a. पुत्रवृत्ति - कर्तृताक - आगति - कालीन - आगति - कर्तृतावान्

He is possessed of agency of the act 'coming', contemporary with the act of agency of which lies in son.

20. पुत्रेण सह मित्रं प्राप्तः चैत्रः - i. Caitra with (his) son obtained a friend. ii. Caitra obtained a friend with (his) son.

20a. पुत्रकर्तृक - मित्र - प्राप्ति - कालीन - मित्र - कर्मक - प्राप्ति - कर्तृत्ववान् चैत्रः

Caitra is possessed of the money of the act of obtaining a friend, contemporary with the act of obtaining a friend, undertaken by his son.

ii. चैत्रस्य पुत्रकर्मक - प्राप्तिकर्तृत्वम्

The state of Caitra's being the agent of the act of obtaining a friend, contemporary with the act of obtaining (Caitra's) son is indicated.

21. धनेन सह अयम् ग्रामं प्राप्तः - This man obtained village with wealth.

22. घटत्वेन सह घटम् अवगाहते बुद्धिः - The knowledge receives jar with jarness.

23. पटः पटत्वेन सह भासते - The cloth appears (simultaneously) with clothness.

24. पत्न्या सह अग्निम् आदध्यात् - One should consecrate the sacrificial fire with his wife.

Points of difference in perception among Vyākaraṇa, Nyāya and Mīmāṃsā views:

A. **Failure of nominative qualificand**

1. पश्य, मृगो धावति - See the deer runs.

1a. मृगो धावति, तम् पश्य - Deer runs, see him.

1b. धावनानुकूल-कृतिमान् मृगः - The deer possesses the volitional effort conducive to running.

and मृगकर्मक - दर्शनाश्रयः त्वम् - You resort to the act of seeing in which deer is the object.

1c. धावनानुकूल - कृतिमन् - मृगनिष्ठ - कर्मता - निरूपक - दर्शनाश्रयः त्वम्

You resort to the act of seeing, determined by the objective status of deer, possessing the volitional effort, conducive to running.

1d. धावन्तं मृगं पश्य - See the running deer.

1e. मृगाभिन्नैकाश्रय-वृत्ति-उत्तरदेशसंयोगानुकुल-वेगजन्य-विलक्षण-व्यापारकर्मक-ईक्षणानुकूलव्यापारः त्वदभिन्नैकाश्रय वृत्तिः

The action in favour of seeing, having the object relating to an (extraordinary) action borne out of speed conducive to contact with the next region, resorting to deer only, is inherent in you.

1f. धावनकर्तृ - मृगकर्मक - दर्शनाश्रयः त्वम् - You resort to the action seeing, the object of which is the deer, the agent of running.

2. शृणु, गर्जति मेघः - Hear the cloud roars.

3. नटो गायति, शृणु - Hear, the dancer sings.

4. घटोऽनश्यत्, पश्य - See, the pot destroyed.

2a. गर्जनाश्रय - मेघ - कर्मक - श्रवणाश्रयः त्वम्

You resort to the action 'hearing', to which the cloud is related as an object resorting to roaring.

3a. गानाश्रय - नटकर्मक - श्रवणाश्रयः त्वम्

You resort to the action 'hearing', to which the dancer is related as an object resorting to singing.

4a. अतीतोत्पत्तिक - नाशप्रतियोगिक - घट - कर्मक - दर्शनाश्रयः त्वम्

You resort to the action 'seeing', to which the pot is related as an object which is the counter positive of destruction produced in past.

2b. मेघाभिन्नैकाश्रयवृत्ति - गर्जन - कर्मक - श्रवणं त्वदभिन्नैकाश्रयकम्

The action hearing, to which roaring is related as an object belonging to cloud only, resorts to you.

3b. नटाभिन्नैकाश्रयनृत्ति - गान - कर्मक - श्रवणं त्वदभिन्नैकश्रयकम्

The action hearing, to which singing is related as an object belonging to dancer only, resorts to you.

घटाभिन्नैक - प्रतियोगिताक - अतीतोत्पत्तिक - नाश - विषयकदर्शनम् त्वदभिन्नैकाश्रयकम्

The action seeing, in which the content is destruction produced in past, the counter positive of which is pot, resorts to you.

5. शंखाः श्रूयन्ते - The conches are heard.

5a. शंखाभिन्न - बहुत्वावच्छिन्न - कर्मवृत्ति - श्रवणानुकूलो व्यापारः (अस्ति, भवति)
The action, in favourable of hearing, belongs to object delimited by manyness resorting to conch.

6. त्रयः कालाः (सन्ति) - Three times (exist).

6a. त्रित्व - विशिष्टाभिन्नाः कालाः (प्रतीयन्ते, ज्ञायन्ते) - Times qualified by the number three.

B. **Failure of action-predominance** (Taddhita suffixes)

7. वैयाकरणः - Grammarian, i.e, person who knows or reads grammar (4.2.59).

8a. रथ्यः - horse, who bears or carries the chariot (4.4.76).

8b. युग्यः - bullock, who bears or carries the yoke (4.4.76).

8c. प्रासङ्ग्यः - a trainee, who bears or carries the yoke-like weed for cattle (4.4.76).

C. **Particle as the main qualificand**

9. भूतले न घटः - There is no pot on the ground.

9a. भूतलानुयोगिक-घटप्रतियोगिकः अभावः - The absence in which pot is counter positive and ground is locus.

10. शिवाय नमः - Obeisance to Lord Śiva.

11. अलंकरोति - Embellishes.

12. आविष्करोति - Discovers.

13. नमस्करोति - Hails / Salutes / Greets.

14. पुरस्करोति - Makes forward.

15. साक्षात्करोति - Meets.

16. तिरस्करोति - Discards.

17. हरये नमः - Obeisance to Hari.

18. स्वस्ति भवते - Blessings to you.

17a. स्व - प्रयोक्तृ - पुरुष - कर्तृक - नमः - प्रयुक्त - त्यागः - Offering by a person who uses the very word.

18a. भवदीय - हितविषयिणी - मदिया - इच्छा - My wish (embedded in the meaning of the word svasti) in which your welfare is the content.

D. **Problem of meaning of a word**

19. पुत्रेण सह आगतः - (He) has come with son.

19a. पुत्रवृत्ति - कर्तृताक - आगति - कालीन - आगति - कर्तृतावान्

He is possessed of agency of the act 'coming', contemporary with the act of agency of which lies in son.

20. पुत्रेण सह मित्रं प्राप्तः चैत्रः - i. Caitra with (his) son obtained a friend. ii. Caitra obtained a friend with (his) son.

20a. पुत्रकर्तृक - मित्र - प्राप्ति - कालीन - मित्र - कर्मक - प्राप्ति - कर्तृत्ववान् चैत्रः

Caitra is possessed of the money of the act of obtaining a friend, contemporary with the act of obtaining a friend, undertaken by his son.

ii. चैत्रस्य पुत्रकर्मक - प्राप्तिकर्तृत्वम्

The state of Caitra's being the agent of the act of obtaining a friend, contemporary with the act of obtaining (Caitra's) son is indicated.

21. धनेन सह अयम् ग्रामं प्राप्तः - This man obtained village with wealth.

22. घटत्वेन सह घटम् अवगाहते बुद्धिः - The knowledge receives jar with jarness.

23. पटः पटत्वेन सह भासते - The cloth appears (simultaneously) with clothness.

24. पत्न्या सह अग्निम् आदध्यात् - One should consecrate the sacrificial fire with his wife.

Types of sentences for illustrating Śābdabodha

1. Active voice without object. {da,dv} ;
2. Active voice with object. {da,o,dv}
3. Active voice with two objects. {da,o,o,dv} ;
4. Active voice with two agents. {da,da,o,dv}
5. Passive voice. {ia,o,iv} ;
6. Passive voice with two agents. {ia,ia,o,iv}
7. Passive voice with two objects. {ia,o,o,iv} ;
8. Passive voice with agent as object. {o,iv}
9. Impersonal Passive voice. {ia,iv}

key:- da = direct agent ; dv = direct verb ; o = object ; ia = indirect agent ; iv = indirect verb.

१. *अकर्मकर्तरि प्रयोगः* । उदा - **रामो विराजते** । 'Rāma shines'. शाब्दबोधश्च -

नैयायिक-मते - विशिष्टदीप्त्यनुकूल - वर्तमानकालिक - कृतिमान् रामः ।
वैयाकरण-मते - रामाभिन्नैककर्तृनिष्ठ - वर्तमानकाल - विशिष्टदीप्त्यनुकूलः व्यापारः ।
मीमांसक-मते - रामनिष्ठकर्तृत्वसमानाधिकरणा विशिष्टदीप्त्यनुकूला वर्तमानकालिकी भावना ।

२. *सकर्मकर्तरि प्रयोगः* । उदा - **रामः भूमण्डलं अपालयत्** । 'Rāma ruled the Earth'. शाब्दबोधश्च -

नै. - भूमण्डलकर्मक - रक्षणानुकूल - अनद्यतनभूतकालिक - कृतिमान् रामः ।
वै. - एकत्वावच्छिन्न-रामाभिन्न-कर्तृनिष्ठः अनद्यतनभूतकालिकः भूमण्डलाभिन्न-कर्मनिष्ठ-रक्षणानुकूलः व्यापारः ।
मी. - भूमण्डलनिष्ठरक्षणानुकूला रामनिष्ठा अनद्यतनकालिकी भावना ।

३. *कर्मणि प्रयोगः* । उदा - **रामेण भुवि अन्यादृशी कीर्तिः उपार्जिता** । 'Unique fame was earned on Earth by Rāma'. शाब्दबोधश्च -

नै. - रामनिष्ठभावनाविष्ठयक - सर्वेतरविलक्षण-भू-अधिकरणक - भूतकालिक - संपादनजन्य - फलशाली कीर्तिः ।
वै. - रामाभिन्नकर्तृनिष्ठः भूतकालिकः भू-अधिकरणकः एकत्वावच्छिन्न-सर्वेतरविलक्षण-यशोभिन्न-कर्मनिष्ठः संपादनानुकूलः व्यापारः ।
मी. - सर्वेतरविलक्षण-यशोभिन्नाश्रित-कर्मत्व-समानाधिकरणा भू-अधिकरणका संपादनप्रयोजिका भूतकालिकी रामनिष्ठ - कर्तृत्वसमानाधिकरणा भावना ।

४. *द्विकर्मक-कर्तरि प्रयोगः* । उदा - **रामचन्द्रः दशग्रीवं परलोकं अनीनयत्** । 'Rāma took (sent) Rāvaṇa to the other world'. शाब्दबोधश्च -

नै. - दशग्रीव-कर्मक - परलोक-अकथितकर्मक - नयनानुकूल - भूतकालिककृतिमान् रामचन्द्रः ।
वै. - दशग्रीवाभिन्नकर्मनिष्ठः परलोकाभिन्न-अकथितकर्मनिष्ठः रामचन्द्राभिन्नकर्तृनिष्ठः भूतकालिकः नयनानुकूलः व्यापारः ।
मी. - दशग्रीवनिष्ठकर्मत्व - परलोकनिष्ठाकथितकर्मत्व - समानाधिकरणा रामचन्द्रनिष्ठा भूतकालिकी नयनानुकूला भावना ।

५. *द्विकर्तृक-कर्तरि प्रयोगः* । उदा - **सीतापतिः स्वयं कपिभिः सेतुं बन्धयामास** । 'Rāma (got) built a bridge by monkeys'. शाब्दबोधश्च

नै. - सेतुनिष्ठकर्मक - बहुकपिकरणक - स्वकर्तृक - परोक्षभूतकालिक - बन्धनानुकूलकृतिमान् सीतापतिः ।
वै. - एकत्वावच्छिन्न-सीतापत्यभिन्न-तत्कर्तृनिष्ठः बहुत्वावच्छिन्न-कप्यभिन्नकरणनिष्ठः सेत्वभिन्नकर्मनिष्ठः परोक्षभूतकालिकः बन्धनानुकूलः व्यापारः ।
मी. - सेतुनिष्ठ - कर्मका साक्षात्त्वविशिष्ट-सीतापतिनिष्ठ - कर्तृका बहुत्वविशिष्ट-कपिनिष्ठ - करणका परोक्षभूतकालिकी बन्धनानुकूला भावना ।

६. *द्विकर्तृक-कर्मणि प्रयोगः* । उदा - **वायुपुत्रेण रामेण लङ्का भस्मसात् अकारि** । 'Laṅkā was rendered to ashes by Hanūmān and Rāma'. शाब्दबोधश्च

नै. - वायुपुत्रनिष्ठ-रामनिष्ठ-समुच्चितभावनाविष्ठयक - भूतकालिक - दहनजन्यफलशालिनी लङ्का ।
वै. - वायुपुत्राभिन्न-रामाभिन्न-समुच्चयनिष्ठ - द्विकर्तृकः भूतकालिकः लङ्काभिन्नकर्मनिष्ठः दहनानुकूलः व्यापारः ।
मी. - वायुपुत्रनिष्ठ-रामनिष्ठ - कर्तृत्वद्वयसमानाधिकरणा लङ्काभिन्नाश्रित - कर्मत्वविशिष्ट - दहनक्रियाफलप्रयोजिका भूतकालिकी भावना ।

७. *द्विकर्मक-कर्मणि प्रयोगः* । उदा - **दशास्यः जानकीशेन त्रिविष्टपं अनायि** । ' Rāvaṇa was taken (sent) to the other world by Rāma'. शाब्दबोधश्च -
नै. - जानकीशनिष्ठ - भावनाविषयक - भूतकालिक - त्रिविष्टपाधिकरणक - नयनजन्यफलशाली दशास्यः ।
वै. - जानकीशाभिन्नकर्तृक - भूतकालिक - एकत्वावच्छिन्न-दशास्याभिन्नकर्मनिष्ठ - त्रिविष्टपाधिकरणक - नयनानुकूलः व्यापारः ।
मी. - जानकीशनिष्ठ-कर्तृत्वसमानाधिकरणा दशास्याभिन्नाश्रित-कर्मत्व - त्रिविष्टपाश्रय-अधिकरणत्व- समानाधिकतणा भूतकालिकी नयनानुकूला भावना । शाब्दबोधश्च -

८. *कर्मकर्तरि प्रयोगः* । उदा - **रामः स्वयं पूज्यते** । 'Rāma is being worshipped'. शाब्दबोधश्च -

नै. - वर्तमानकालिक - पूजाजन्य - फलशाली स्वाभिन्नः रामः ।
वै. - स्वाभिन्नविशिष्टैकत्वावच्छिन्न-रामाभिन्नकर्तृनिष्ठः वर्तमानकालिकः पूजानुकूलः व्यापारः ।
मी. - स्वाभिन्नत्वविशिष्ट-रामाभिन्नाश्रित-कर्मत्व-समानाधिकरणा वर्तमानकालिकी पूजाजन्य फलप्रयोजिका भावना ।

९. *भावे प्रयोगः* । उदा - **रामचन्द्रेण भूयते** । '(It is) being existed by Rāma'. शाब्दबोधश्च -

नै. - वर्तमानकालिक - सत्ताप्रयोजक - कृत्याश्रयः रामचन्द्रः ।
वै. - रामचन्द्राभिन्नाश्रयकः वर्तमानकालिकः सत्तानुकूलः व्यापारः ।
मी. - रामचन्द्रनिष्ठ-कर्तृत्व-समानाधिकरणा वर्तमानकालिकी सत्तानुकूला भावना ।

Relations used in Śābdabodha - Case-relations

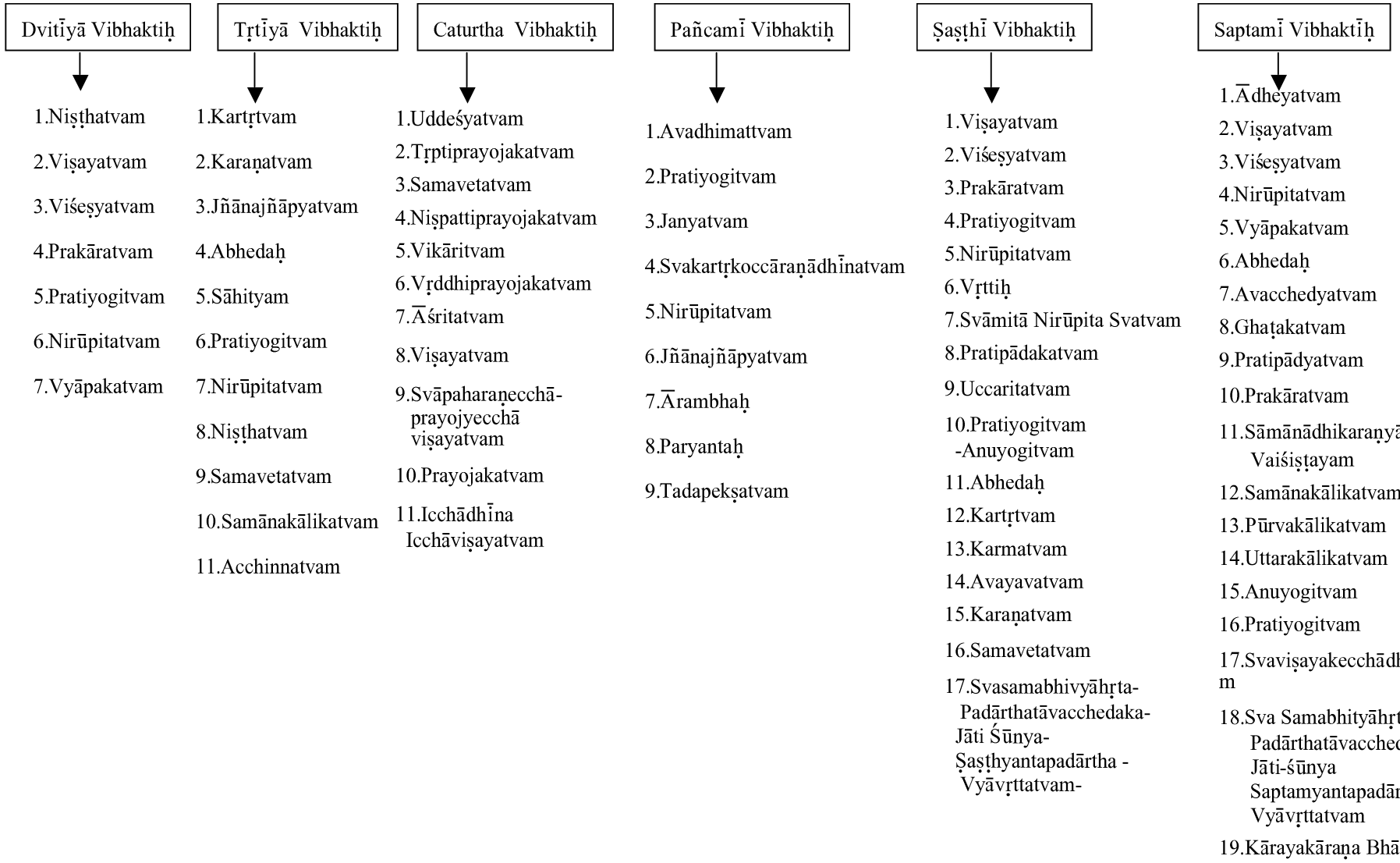


Figure 8.1

PART IV

Chapter 9 : Design and expected outputs

9.1 System details of General Purpose Sanskrit Parser

The proposed GP Sanskrit Parser is to take a Sanskrit word or sentence (plain or accented) as input. Fig. 9.1 is the flow chart of the parser. If it is a sentence, it is to be split into words. Then for each word, morphological analysis is to be performed to recognise the possible word categories. Accented inputs need to be processed for accent information also. Nouns, verbs or indeclinables will be the categories of output with standard details. When multiple solutions exist, all are considered.

9.2 Algorithm for morphological analysis

Accented input

The Flow chart enclosed Fig. 9.2 outlines morphological analysis for accented inputs and Fig. 9.3 gives Vedic syllable syntax, the steps involved in converting from Saṃhitā-pāṭha to Pada-pāṭha and vice versa, as also generating the various combinatorial patterns called Veda-Vikṛti-s. Essentially, we have an additional dimension of accent-markers to be handled which are always associated with vowels or (special) Vedic characters. Itiga-s need to be split and 'iti' should be removed. The attendant accent-change and rewriting of lṅgya etc. are to be carried out. The examples below from Saṃhitā to Pada Pāṭha have these steps illustrated. Special Vedic characters are also used and their results after the processing are included. Additional Information like Vedic fonts key maps, sample accented texts of different Veda-s etc. are presented in the Annexure - B.

Unaccented input

After taking care of accent info. if input is accented, and directly otherwise, we can adopt the following general strategy :

1. **Take** a Sanskrit word input.
2. **Check** whether the word is an avyaya (indeclinable). If so, store the identification.
3. **Check** whether the word is any special form. If so, store the identification.
4. **Generate** all possible splits of a word (along with second level character substitution) giving rise to different sub-string combinations.
5. **For** all sub-string combinations

do

6. **Check** the last sub-string in the Nominal suffix and Verbal suffix databases.
7. **if the last sub-string is a Nominal suffix**

then

8. **While** there are more sub-strings in the combination

do

9. **Take** the next sub-string.
10. **Check** the sub-string in the following databases in turn.

i. Kṛt suffix
ii. Taddhita suffix
iii. Feminine suffix
iv. Nominal base
v. Verbal root
vi. Upasarga

if a match is found in any of the databases, carry out a compatibility type **check** between the corresponding sub-strings, **and if** OK, store the identification along with the relevant details of the sub-string in the combination.

11. **if** the sub-string does not match in any of the databases, **break** out of the **while** loop;

12. **If** the total identification of all the sub-strings of the combination is such that it is one of the valid word structures **then** store the total identification of the word.

13. **Else if the last sub-string is a Verbal suffix**

then

14.**While** there are more sub-strings in the combination

do

15. **Take** the next sub-string.

16. **Check** the sub-string in the following databases in turn.

i. Modal suffix

ii. Conjugational suffix

iii. Various augments

iv. Verbal root

v. Upasarga

if a match found in any of the databases, carry out a compatibility type **check** between the corresponding sub-strings, **and if** OK store the identification along with the relevant details of the sub-string in the combination.

17. **if** the sub-string does not match in any of the databases, **break** out of the **while** loop;

18. **If** the total identification of all the sub-strings of the combination is such that it is one of the valid word structures, **then** store the total identification of the word.

19. **Else**

20. **Take** the last sub-string.

21. **While** there are more sub-strings in the combination

do

22. **Check** the sub-string in the avyaya list, **if** a match is found, store the identification, **take** the next sub-string and **continue.**

23. **Check** the sub-string in the following databases in turn.

i. Kṛt suffix

ii. Taddhita suffix

iii. Feminine suffix

iv. Nominal base

v. Verbal root

vi. Upasarga

if a match is found in any of the databases carry out a compatibility type **check** between the corresponding sub-strings, **and if** OK, store the identification along with the relevant details of the sub-string in the combination.

24. **if** the sub-string does not match in any of the databases, **break** out of the **while** loop;

25. **If** the total identification of all the sub-strings of the combination is such that it is one of the valid word structures, **then** store the total identification of the word.

26. **If** there is no identifcation for the word (for any of the sub-string combination), **then print** the word is not a valid Sanskrit word; **else print** all the grammatical identifications of the word.

27. **END.**

Discussion

Here, the database sizes and content would decide the data structures and techniques used for searching, storage etc. The nominal bases used are about 3000 words, the suffixes of various types about 1000, list of avyaya-s about 700; while the roots are 2000, suffixes and preverbs with their combinations about 500. The databases are of the order of 0.4 MB in size as far as verbal roots, suffixes etc. are concerned. The lexicon has a provision for 'update' as it can be infinite. But, certain decision processes are involved as verbal suffixes are dependent on about 7 parameters as Gaṇa, Padī, It, Iṭ, Karmitva, accent, meaning, besides tense/mood, person, number. Hence, the arrangement, order and storage of intermediate results at each iteration of the word-split process would have to be considered. The process is not linearly decomposable. Preverbs (even upto three of them) alter some of these attributes of verbal roots.

Hence, storage as hashed files or reverse suffix table, tries etc. are candidates. Paradigm types are frozen for nominal bases and suffixes, verbs are used with partial finality and indeclinables and other suffixes are fixed. Hence, depending on the level and stage of the split, the LHS and RHS strings can be matched in the database. String-length based sorting and alphabetical indexing can also be employed.

From the above algorithm, we see that it mainly involves generating of all possible combinations and then checking in different databases. If the check is successful and the structure of word is one of the valid forms, then the given input is a valid word, whose grammatical identification can be built up from the identifications of part sub-strings. The detailed description of the algorithm for generating all possible combinations is given below alongwith its importance for NLU in Sanskrit and other Indian Languages. It also brings out the theoretical problem in detail, highlighting the importance of the procedure for NLU.

The common algorithm is explained with various detailed stages of processing in Fig. 9.4 and morphological validation details are found in Fig. 9.5.

9.3 Śābdabodha Flow-chart explanation

The process of Śābda Bodha analysis begins with taking the verbal identifications from the syntactic analysis stage and the functor identification from the semantic analysis stage. For the verbs, the roots are collected and then their category specifications are consulted.

Given these information, we check all other words of the sentence other than the verb to find out if there are any candidates for mandatory Kāraka-s. On finding candidates, the type check is carried out. If it passes, that word with that functor, forms part of a possible meaning. This is continued for all mandatory kāraka-s. Whenever there are no candidates for even one mandatory kāraka, then the sentence is declared incomplete for that verb in that sense. When more than one candidate exists for the same mandatory kāraka, then extra rule base will be required to disambiguate (or the sentence can have more meaning).

The similar check for presence and compatibility of optional kāraka-s is carried out when there is atleast one candidate for all mandatory kāraka-s. After this, the sentence should be checked for any non-kāraka word like adjectives, adverbs, indeclinables etc. and their meanings to be determined. This is again done by suitable rulebase.

After all the words are accounted for in the sentence, the output is given in one of the three traditional formats due to the three schools of ancient Indian sciences. *i.e*, Logic, Grammar or Epistemology. These formats differ only in the emphasis laid in the description of the meaning. The logicians treat the agent as central to semantic extraction, while the grammarians treat activity content as central and Mīmāṃsaka-s treat the intention of the agent as significant in the sentence meaning.

For syntactic analysis, the Śāstra-ic approach is illustrated by Fig. 9.6, the expectancy-based syntactic analyser. Fig. 9.7 gives the procedure Ākāṅkṣā. Here, we begin with the verb, usually (if present). By means of putting various questions as to various functional roles with respect to the verbal activity, the verbal import is arrived at.

Next, we get to decide the meanings and inter-relation of words. We begin with the verb and link nouns through syntactico-semantic mappings. These are devised as shown in Figs. 9.8 to 9.14 for case-ending (Vibhakti-s) sub-types based on functional roles (Kāraka) and collocations (Upapada) as described in Pāṇini's rules 2.3.1 - 2.3.73. Total classification is in 9.8 and individual syntactic conditions for identification are given Figs. 9.9 to 9.14. Figs. 9.15 to 9.21 are for case-relation (Kāraka-s) sub-types based on specific contexts as described in Pāṇini's rules 1.4.23 - 1.4.55. Total classification is in 9.15 and individual conditions for identification are given 9.16- 9.21.

Starting with the activity category, we look at activity specification in terms of expected noun types. We have grouped all activities (Ref. Annexure - C) into eleven suitable groups in Annexure - E. The characterisation and grouping into major functor types of activities are shown in Figs. 9.22 to 9.28. An ontology derived out of these details is presented in Fig. 9.29 to 9.32.

The conditions for mapping any particular sub-type of a case ending to a functor role sub-type is achieved through the link of semantic conditions shown alongside the types. Disambiguation takes special features into account as a weighted choice between alternatives by constraints and cost assignment. The output is given in the form of paraphrase (Śābdabodha) in the traditional format.

The classification of Śābdabodha is in Fig. 9.33. Verbal import of Vedic inputs is shown in Fig. 9.34. The three formats in traditional Śāstra-s are shown diagrammatically for the sentence : 'Caitraḥ Hariṃ Bhajati' - Caitra serves Hari - in Figs. 9.35 for Vyākaraṇa, 9.36 for Nyāya and 9.37 for Mīmāṃsā.

9.4 Expected outputs

A set of expected outputs for an incrementally built up sentence, step by step, and including moods, tenses etc. are given in Figs. 9.38 to 9.54.

The sentence taken is: The potter makes (Karoti Kulālaḥ). We then add object, instrument etc. by adding the words, for a pot, by a stick, for money, from mud, at home, in his home, and further add adjectives for all cases like, smart potter, big pot, long stick, lot of money, fertile mud, old house and familial house etc. (Caturaḥ Kulālaḥ Mahāntaṃ Ghaṭaṃ Dīrgheṇa Daṇḍena Prabhūtāya Dhanāya Praśastāt Mṛdaḥ Sakuṭumbasya Svasya Purāṇe Gṛhe Āśu Karoti).

Past tense of imperfect (Akarot - Parasmaipadī, Dhanaṃ Alabhata Kulālaḥ - Ātmanēpadī) and aorist (Akārṣīt) types, future (Kariṣyati), optative/benedictive moods (Karotu), indeclinable particle of assertion (Devadatta Eva Pacati), activity sequence using infinitive/gerunds and participials (Paṭhitvā Likhāmi Aham, Guruṃ Avanamya Likhāmi Aham, Pītaḥ Kṛṣṇaḥ, Gatavān Rāmaḥ) are included.

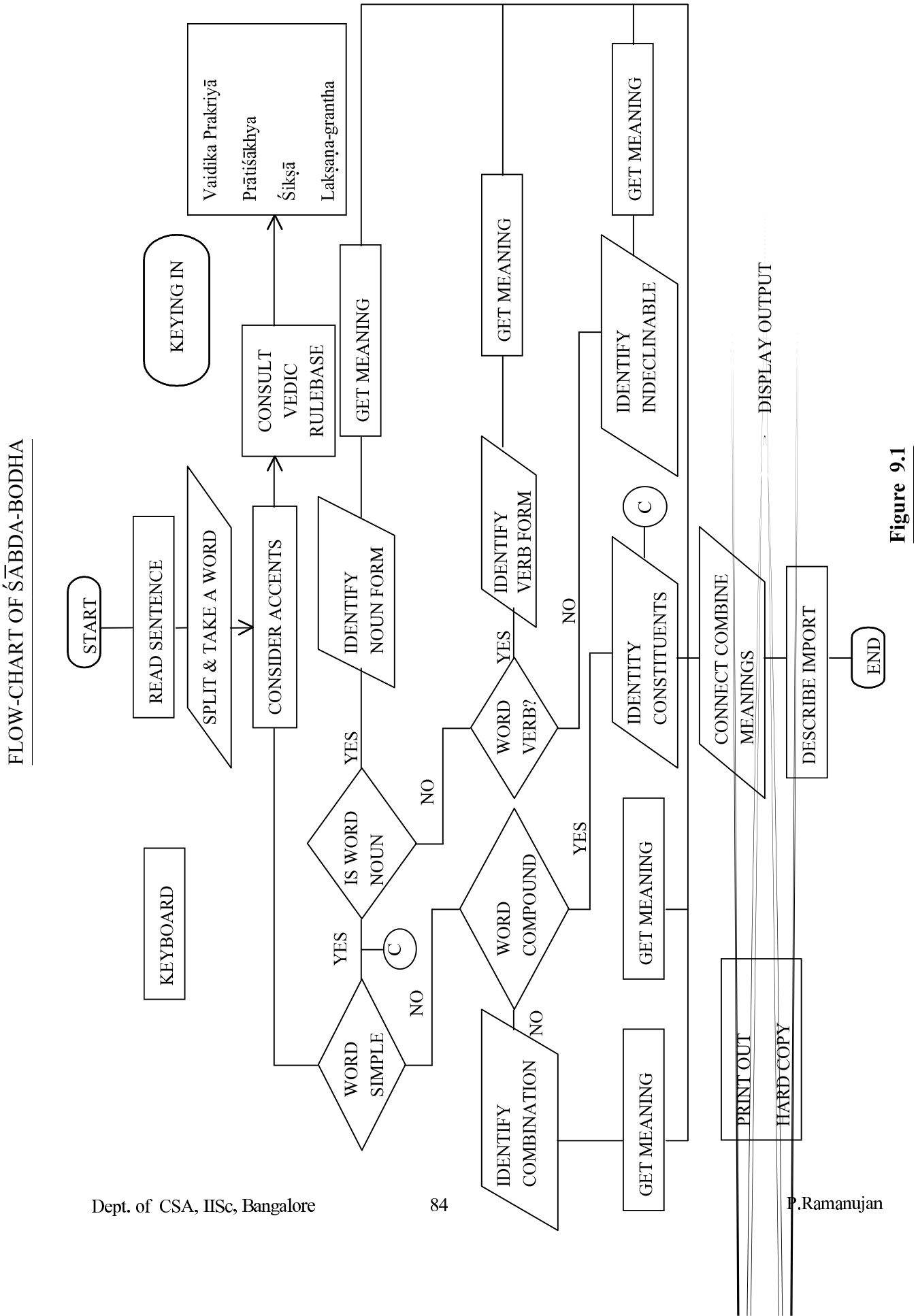

FLOW-CHART OF ŚĀBDA-BODHA
START
READ SENTENCE
SPLIT & TAKE A WORD
CONSIDER ACCENTS
CONSULT VEDIC RULEBASE
Vaidika Prakriyā
Prātiśākhya
Śikṣā
Lakṣaṇa-grantha
KEYING IN
KEYBOARD
IS WORD NOUN
YES
NO
C
IDENTIFY NOUN FORM
GET MEANING
WORD VERB?
YES
NO
IDENTIFY VERB FORM
GET MEANING
WORD COMPOUND
YES
NO
IDENTIFY COMBINATION
GET MEANING
WORD SIMPLE
YES
NO
C
IDENTITY CONSTITUENTS
IDENTIFY INDECLINABLE
GET MEANING
CONNECT COMBINE MEANINGS
DISPLAY OUTPUT
DESCRIBE IMPORT
END
PRINT OUT
HARD COPY

Figure 9.1

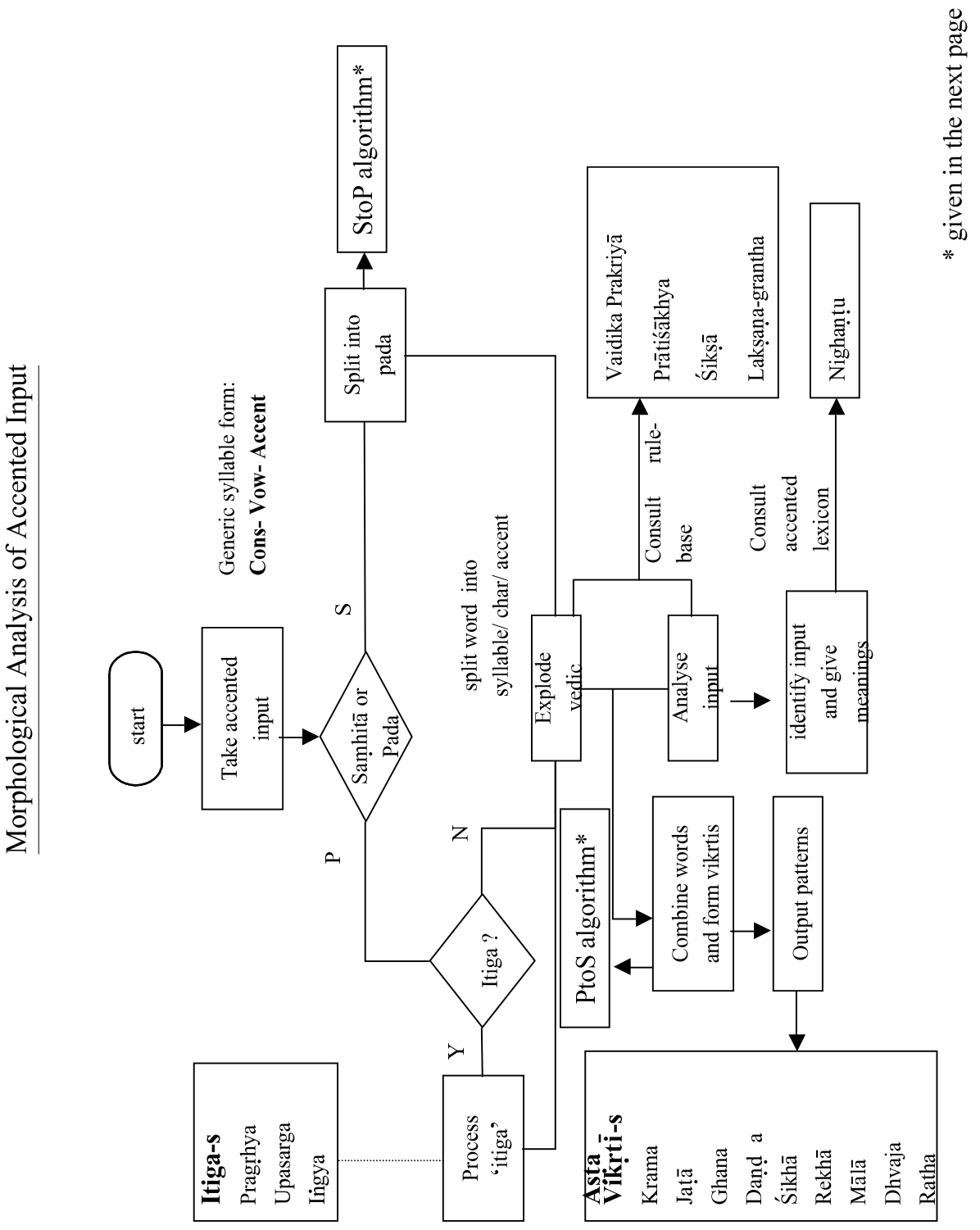

Morphological Analysis of Accented Input
start
Take accented input
Generic syllable form:
Cons- Vow- Accent
Samhitā or Pada
S
P
Split into pada
StoP algorithm*
N
Itiga ?
Y
Process 'itiga'
Itiga-s
Pragṛhya
Upasarga
Iṅgya
split word into syllable/ char/ accent
Explode vedic
Analyse input
Consult rule-base
Vaidika Prakriyā
Prātiśākhya
Śikṣā
Lakṣaṇa-grantha
Consult accented lexicon
identify input and give meanings
Nighaṇṭu
PtoS algorithm*
Combine words and form vikrtis
Output patterns
Aṣṭa Vikṛti-s
Krama
Jaṭā
Ghana
Daṇḍ a
Śikhā
Rekhā
Mālā
Dhvaja
Ratha
* given in the next page

Figure 9.2

Vedic Syllable Syntax - ISCII code syntax extension [S = Svara R = Non-Svara **]**

Vedic-Syllable::= Vedic-Cons-Vowel-Syllable | Vedic-Cons-Vowel-Syllable |Full-Vedic-Syllable

Vedic-Cons-Vowel-Syllable ::= Cons-Vowel-Syllable S

Vedic-Vowel-Syllable::=Vowel-Syllable S

Full-Vedic-Syllable:: R S

A) Samhita to Padapatha - StoP Algorithm

1) convert svarita & pracaya-s to anudāttā

2) Expand Iṅgya-s, add iti and rewrite seperately

3) With udāttā-s as pivot, accent each syllable

4) If there are any jātya (nitya) svaritā-s, write it as it is.

5) Undo any sandhis in saṃhitā (prose) form & write pausa form cerebralization / elongation non-sandhi etc.

6) Other specialities of pada-pāṭha to be added. Itiga-s

B) Pada Pāṭha to Samhita PtoS Algorithm

1) dissolve all special features of pada-pāṭha

2) make anudāttā-s all svarita-s and pracaya

3) combine words make sandhi euphonically as per rule base.

4) making udāttā-s fixed, complete accent marking

5)hrasva / dīrgha kampa-s are to be placed appropriately

6) carry out elongation cerebralization etc.

e.g, śrúdhi in saṃhitā; śrudhi in padapāṭha

Figure 9.3

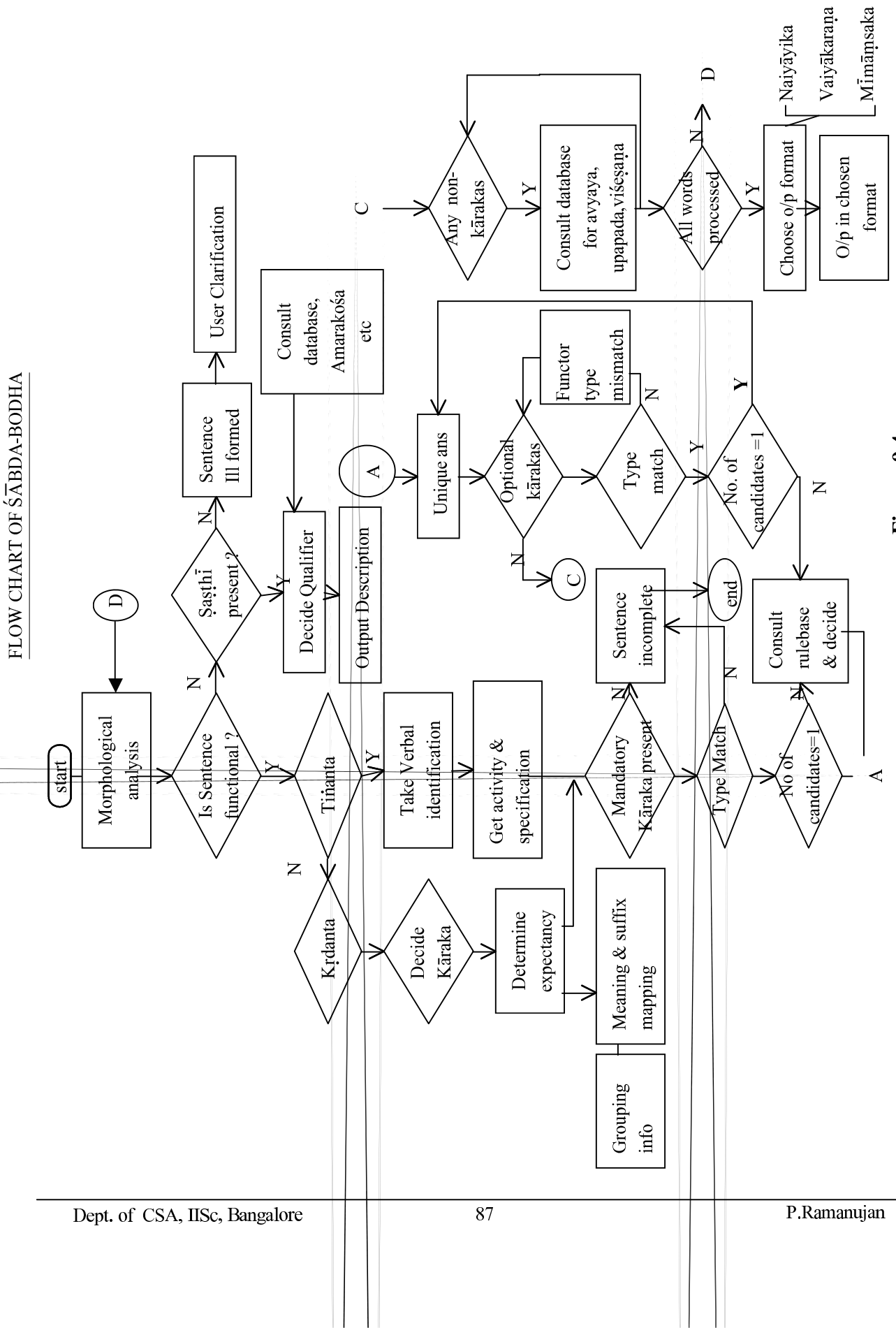

FLOW CHART OF ŚĀBDA-BODHA
start
Morphological analysis
Is Sentence functional ?
Y
N
Tiṅanta
Y
N
Kṛdanta
Take Verbal identification
Decide Kāraka
Get activity & specification
Determine expectancy
Grouping info
Meaning & suffix mapping
Mandatory Kāraka present
N
Y
Sentence incomplete
Type Match
N
Y
No of candidates=1
N
Y
Consult rulebase & decide
A
Ṣaṣṭhī present ?
N
Y
Sentence Ill formed
User Clarification
Decide Qualifier
Consult database, Amarakośa etc
Output Description
A
Unique ans
Optional kārakas
N
Y
C
Functor type mismatch
Type match
N
Y
No. of candidates =1
N
Y
end
C
Any non-kārakas
Y
N
Consult database for avyaya, upapada,viśeṣaṇa
All words processed
N
Y
D
Choose o/p format
Naiyāyika
Vaiyākaraṇa
Mīmāṃsaka
O/p in chosen format
D

Figure 9.4

FLOW CHART FOR MORPHOLOGICAL VALIDATION

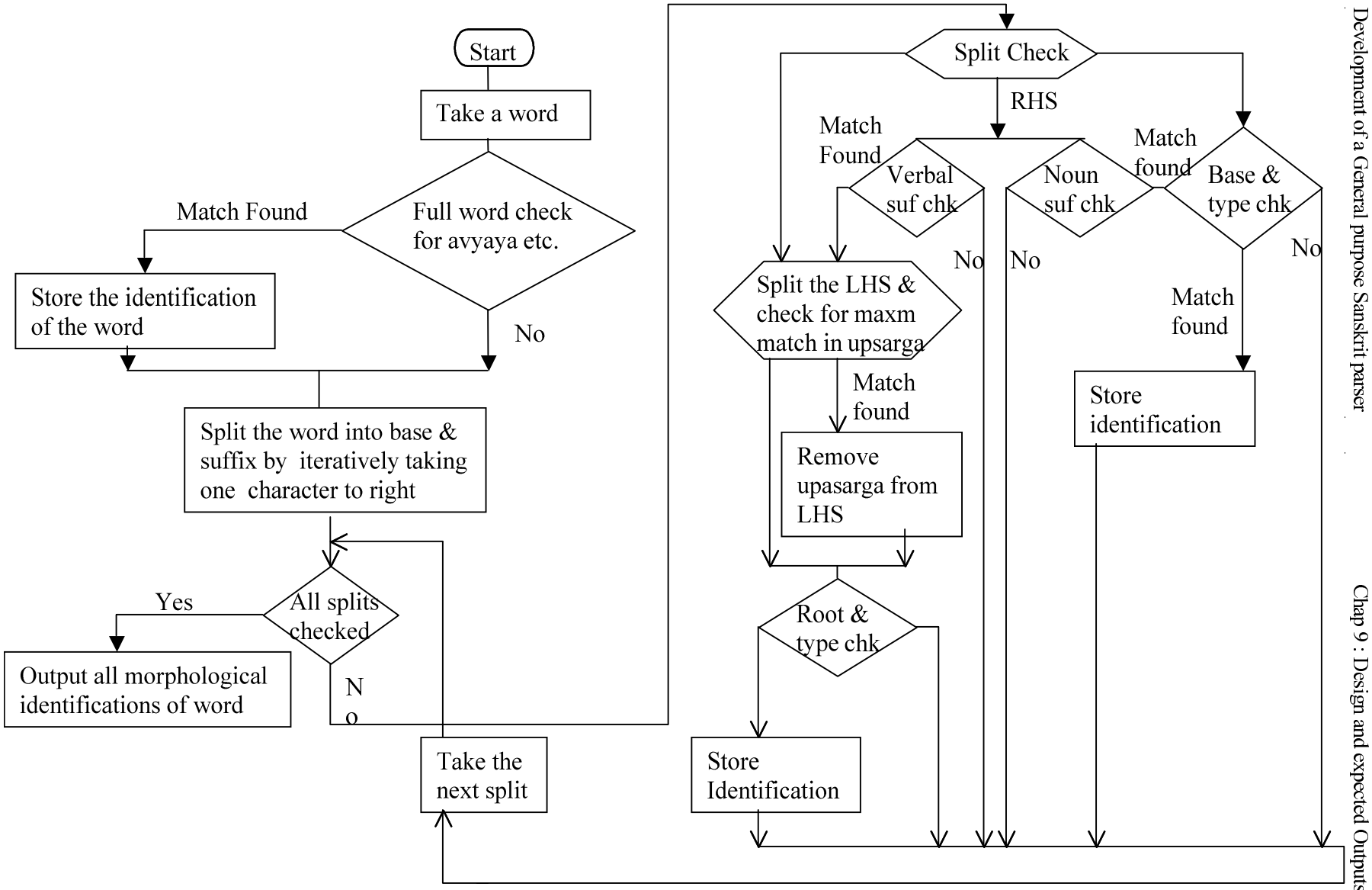
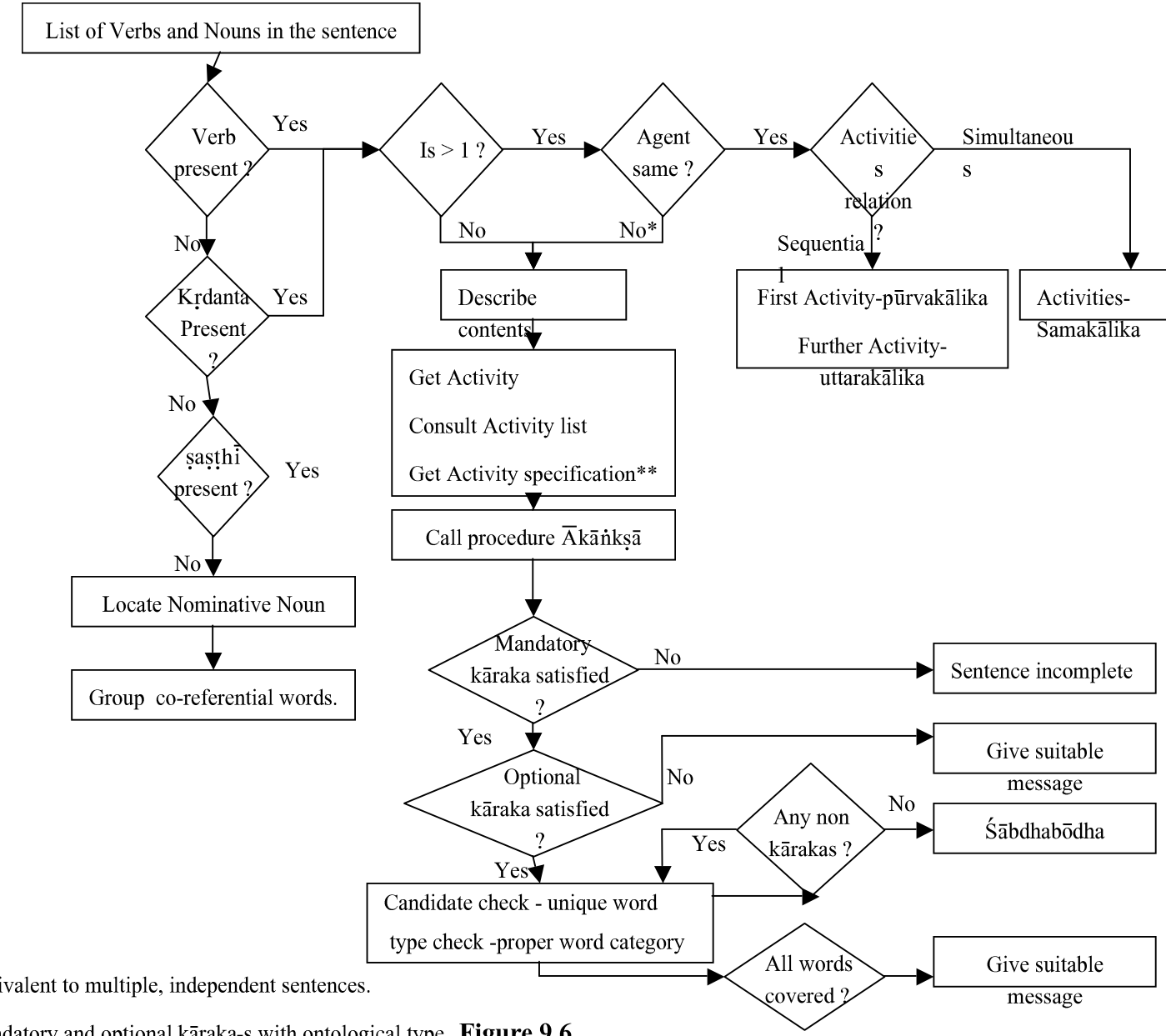


Figure 9.5

Expectancy based Syntactic Analyser

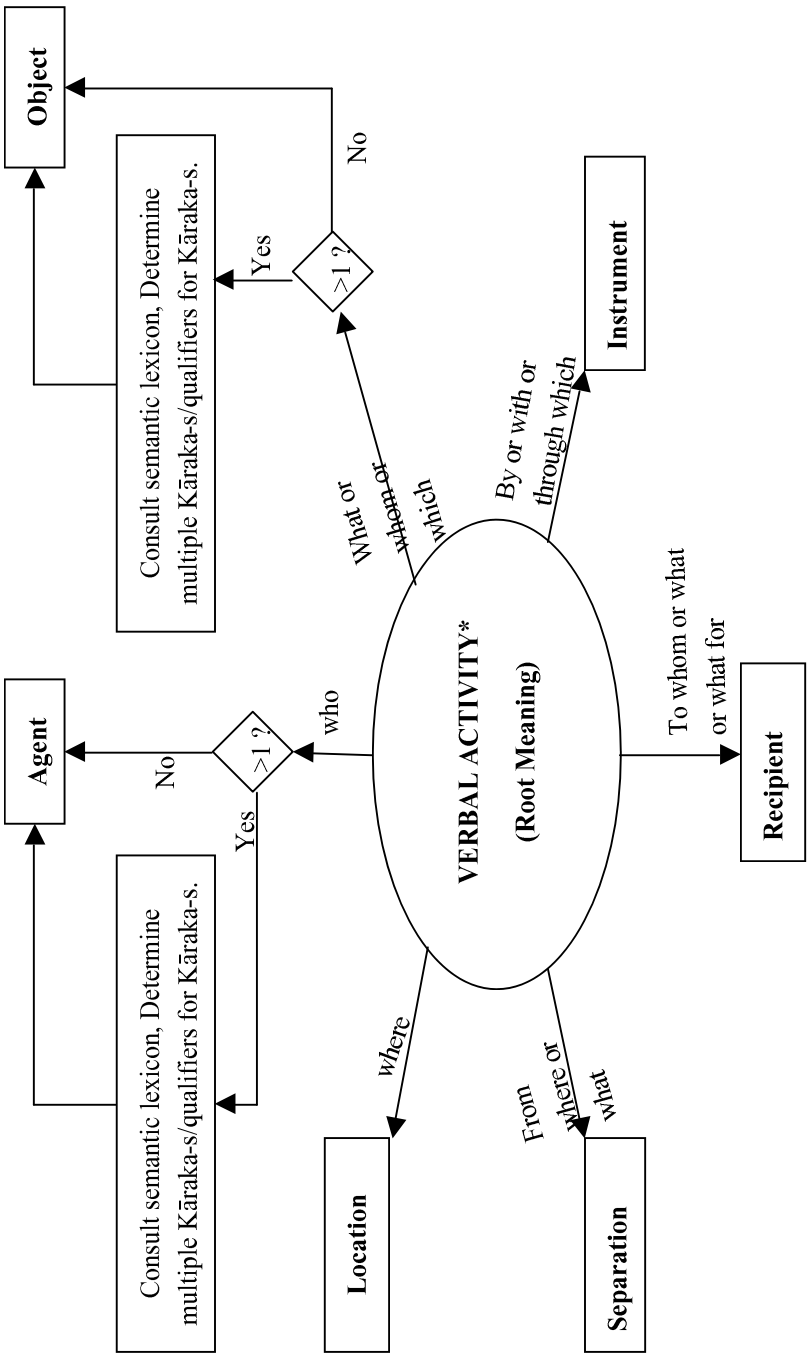


Note :

* -Equivalent to multiple, independent sentences.

** -Mandatory and optional kāraka-s with ontological type **Figure 9.6**

Procedure Ākāṅkṣā



Note : * - If no activity marker, I.e., Verb or Kṛdanta or multiple activities - refer to syntactic analyser.

Figure 9.7

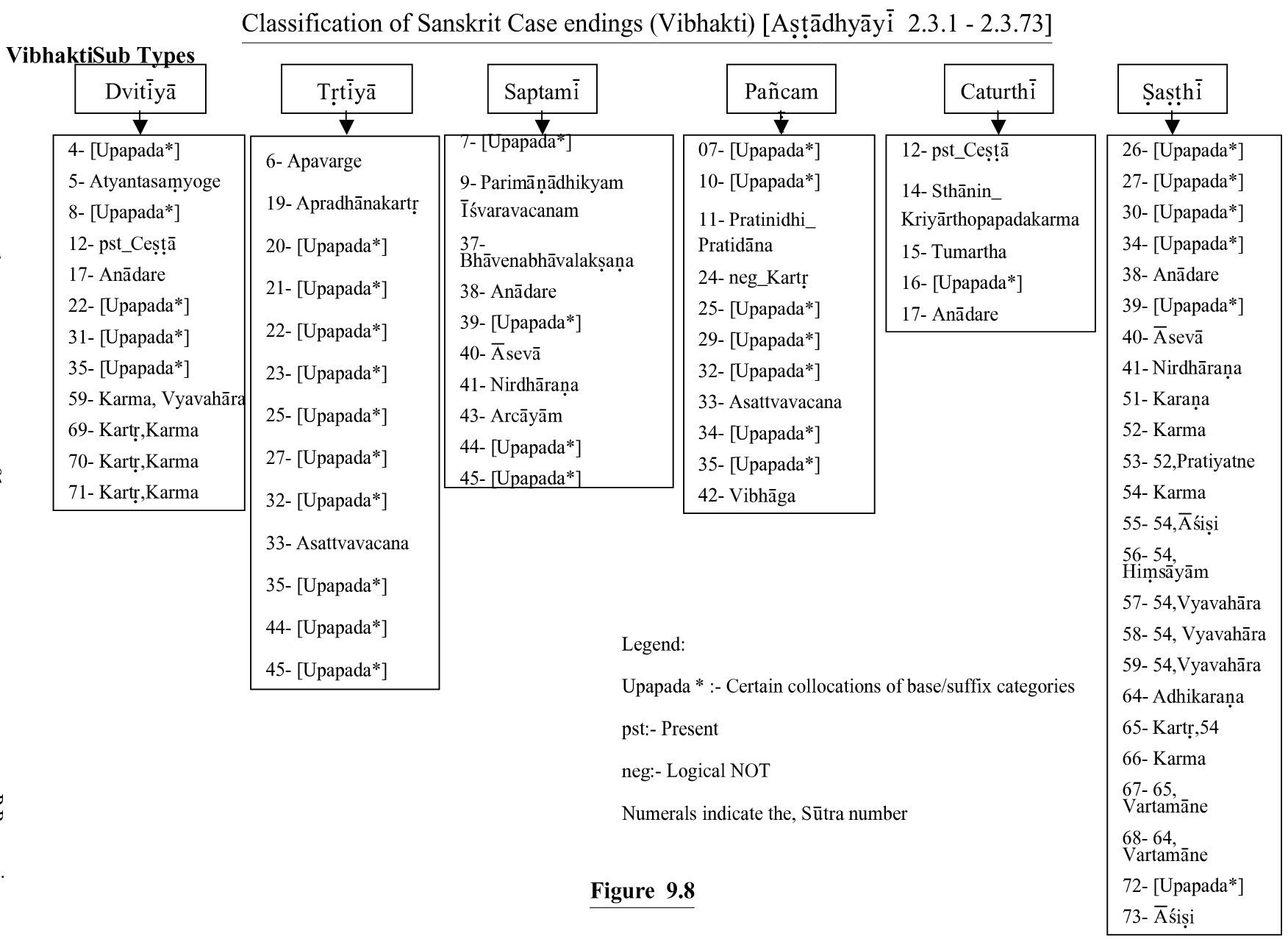

Classification of Sanskrit Case endings (Vibhakti) [Aṣṭādhyāyī 2.3.1 - 2.3.73]
VibhaktiSub Types
Dvitīyā
4- [Upapada*]
5- Atyantasaṃyoge
8- [Upapada*]
12- pst_Ceṣṭā
17- Anādare
22- [Upapada*]
31- [Upapada*]
35- [Upapada*]
59- Karma, Vyavahāra
69- Kartṛ,Karma
70- Kartṛ,Karma
71- Kartṛ,Karma
Tṛtīyā
6- Apavarge
19- Apradhānakartṛ
20- [Upapada*]
21- [Upapada*]
22- [Upapada*]
23- [Upapada*]
25- [Upapada*]
27- [Upapada*]
32- [Upapada*]
33- Asattvavacana
35- [Upapada*]
44- [Upapada*]
45- [Upapada*]
Saptamī
7- [Upapada*]
9- Parimāṇādhikyam Īśvaravacanam
37- Bhāvenabhāvalakṣaṇa
38- Anādare
39- [Upapada*]
40- Āsevā
41- Nirdhāraṇa
43- Arcāyām
44- [Upapada*]
45- [Upapada*]
Pañcam
07- [Upapada*]
10- [Upapada*]
11- Pratinidhi_ Pratidāna
24- neg_Kartṛ
25- [Upapada*]
29- [Upapada*]
32- [Upapada*]
33- Asattvavacana
34- [Upapada*]
35- [Upapada*]
42- Vibhāga
Caturthī
12- pst_Ceṣṭā
14- Sthānin_ Kriyārthopapadakarma
15- Tumartha
16- [Upapada*]
17- Anādare
Ṣaṣṭhī
26- [Upapada*]
27- [Upapada*]
30- [Upapada*]
34- [Upapada*]
38- Anādare
39- [Upapada*]
40- Āsevā
41- Nirdhāraṇa
51- Karaṇa
52- Karma
53- 52,Pratiyatne
54- Karma
55- 54,Āśiṣi
56- 54, Hiṃsāyām
57- 54,Vyavahāra
58- 54, Vyavahāra
59- 54,Vyavahāra
64- Adhikaraṇa
65- Kartṛ,54
66- Karma
67- 65, Vartamāne
68- 64, Vartamāne
72- [Upapada*]
73- Āśiṣi
Legend:
Upapada * :- Certain collocations of base/suffix categories
pst:- Present
neg:- Logical NOT
Numerals indicate the, Sūtra number

Figure 9.8

Syntactic conditions to identify Second case ending Sub Type

Figure 9.9

Syntactic conditions to identify Third case ending Sub Type

Figure 9.10

Syntactic conditions to identify Seventh case ending Sub Type

Figure 9.11

Syntactic conditions to identify Fifth case ending Sub Type

Figure 9.12

Syntactic conditions to identify Fourth case ending Sub Type

Figure 9.13

Syntactic conditions to identify Sixth case ending Sub Type

Pst- Present (Upapada)
wc- Word category
upr- Upasarga and root
root- Dhātu
avysuf- Taddhita vibhakti suffix
c- Condition
base- Prātipadika
krd-Kṛdanta
suf- Suffix
anyr- Upasarga
andd- logical AND
neg- logical NOT
any- any Upapada

Figure 9.14

Chart of Kāraka subtypes

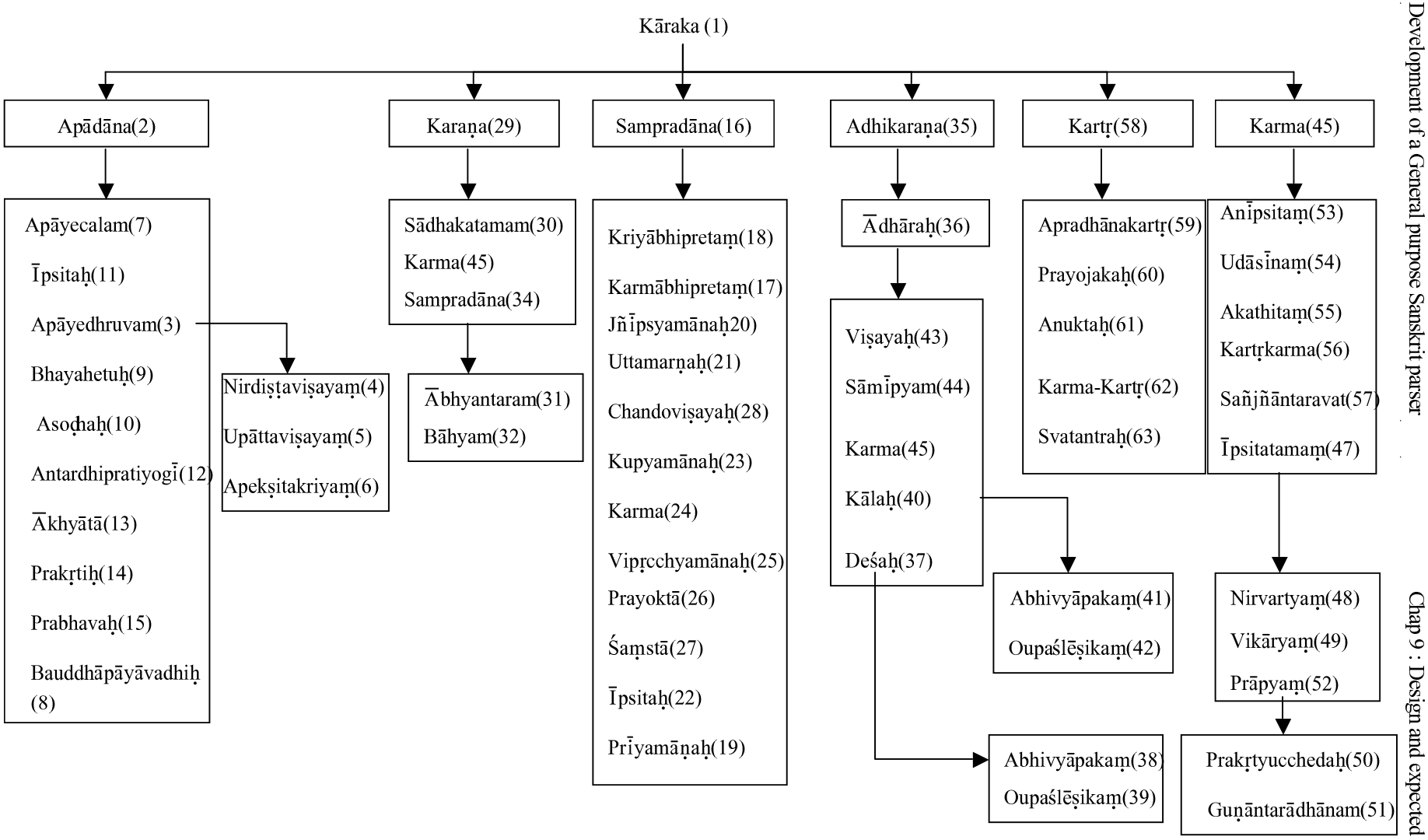


Figure 9.15

Conditions for Apādāna Kāraka subtypes

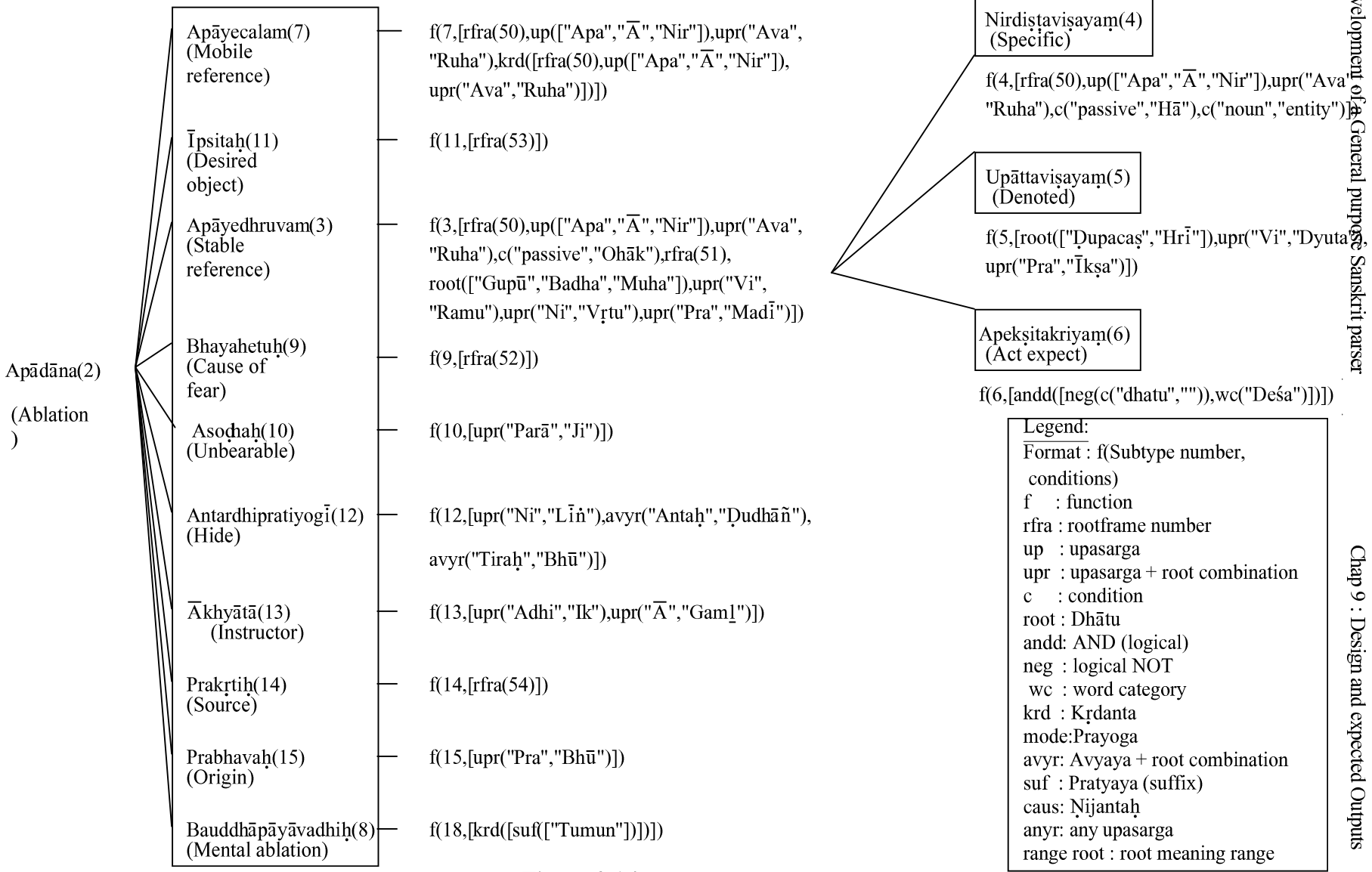


Figure 9.16

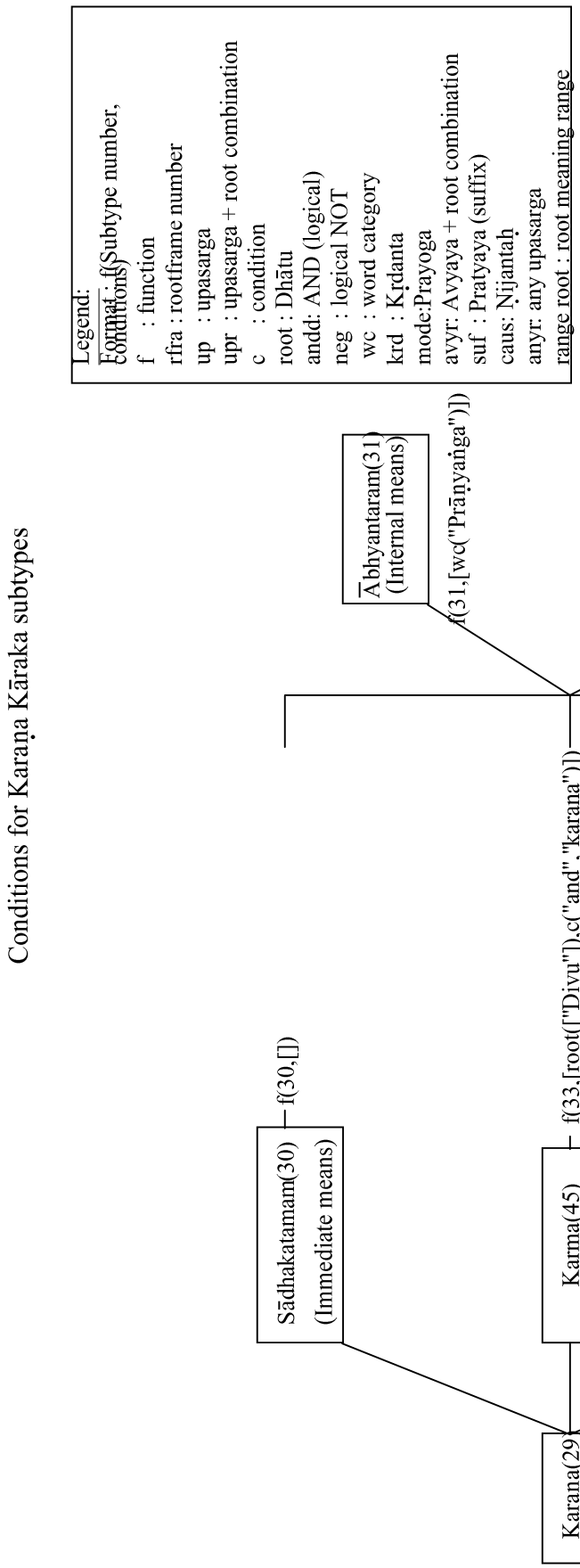

Conditions for Karaṇa Kāraka subtypes
Legend:
Format : f(Subtype number, conditions)
f : function
rfra : rootframe number
up : upasarga
upr : upasarga + root combination
c : condition
root : Dhātu
andd: AND (logical)
neg : logical NOT
wc : word category
krd : Kṛdanta
mode:Prayoga
avyr: Avyaya + root combination
suf : Pratyaya (suffix)
caus: Ṇijantaḥ
anyr: any upasarga
range root : root meaning range
Karaṇa(29)
(Means)
Sādhakatamam(30)
(Immediate means)
f(30,[])
Karma(45)
(Object)
f(33,[root(["Divu"]),c("and","karana")])
Sampradāna(34)
(Donee)
f(34,[upr("Pari","Ḍukrīñ"),c("or","karana")])
Ābhyantaram(31)
(Internal means)
f(31,[wc("Prāṇyaṅga")])
Bāhyam(32)
(External means)
f(32,[neg(wc("Prāṇyaṅga"))])

Figure 9.17

Conditions for Sampradāna Kāraka subtypes

Figure 9.18

Conditions for Adhikaraṇa Kāraka subtypes

Figure 9.19

Conditions for Karma Kāraka subtypes

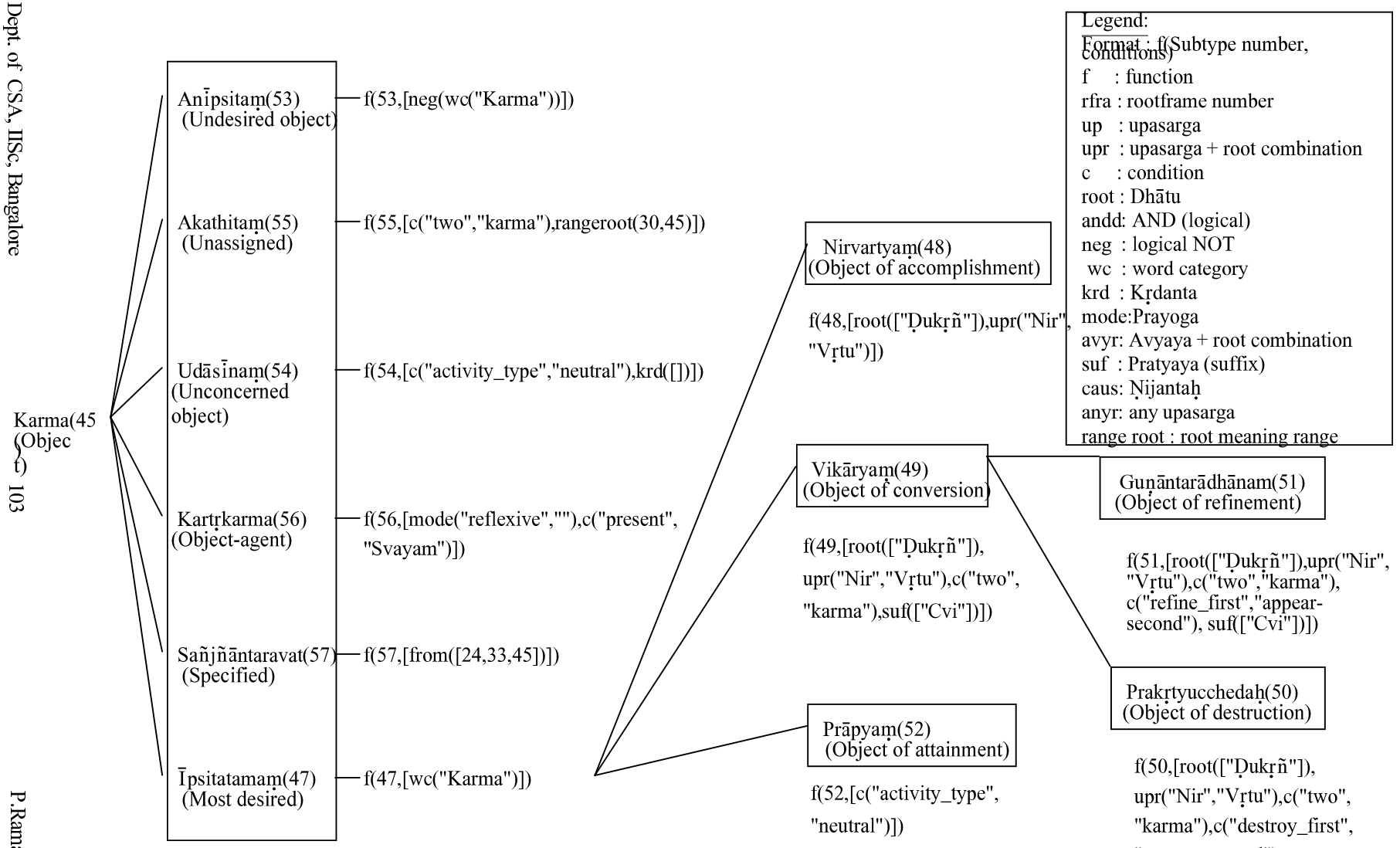


Figure 9.20

Conditions for Kartṛ Kāraka subtypes

Figure 9.21

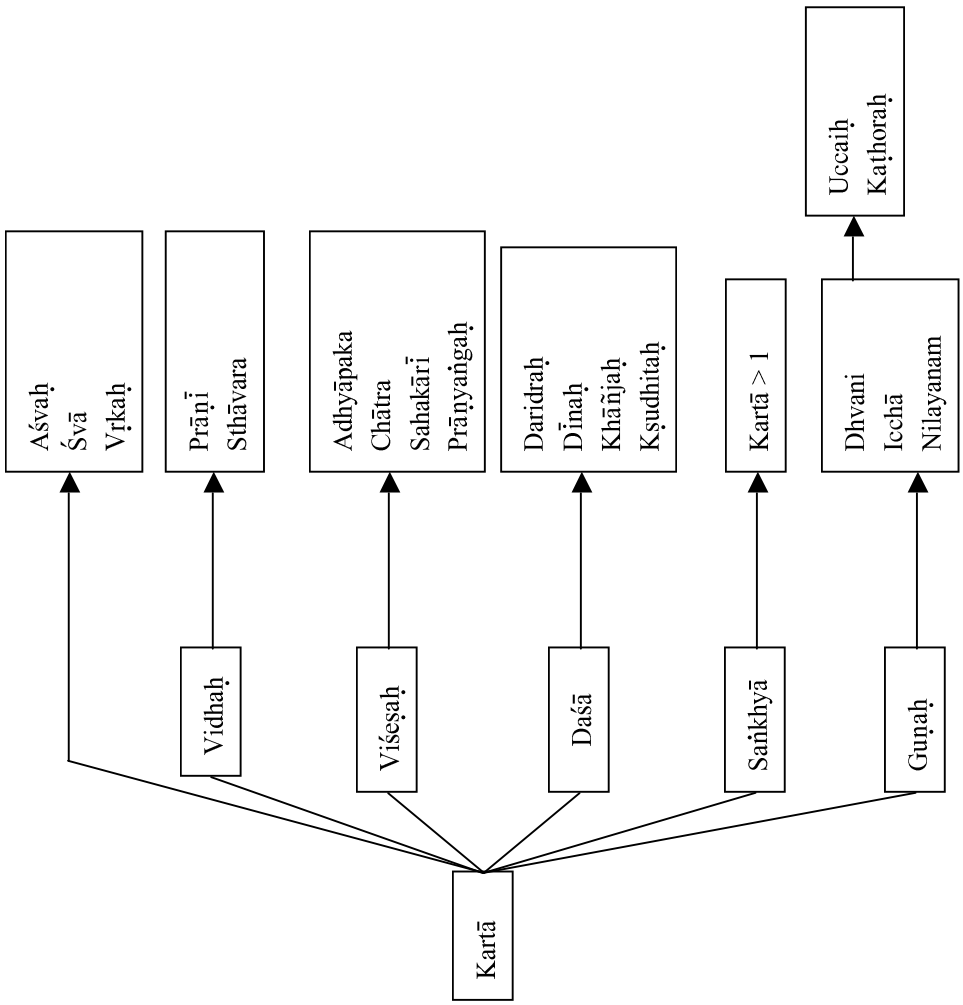

Kartā
Vidhaḥ
Viśeṣaḥ
Daśā
Saṅkhyā
Guṇaḥ
Aśvaḥ
Śvā
Vṛkaḥ
Prāṇī
Sthāvara
Adhyāpaka
Chātra
Sahakārī
Prāṇyaṅgaḥ
Daridraḥ
Dīnaḥ
Khāṇjaḥ
Kṣudhitaḥ
Kartā > 1
Dhvani
Icchā
Nilayanam
Uccaiḥ
Kaṭhoraḥ

Figure 9.22

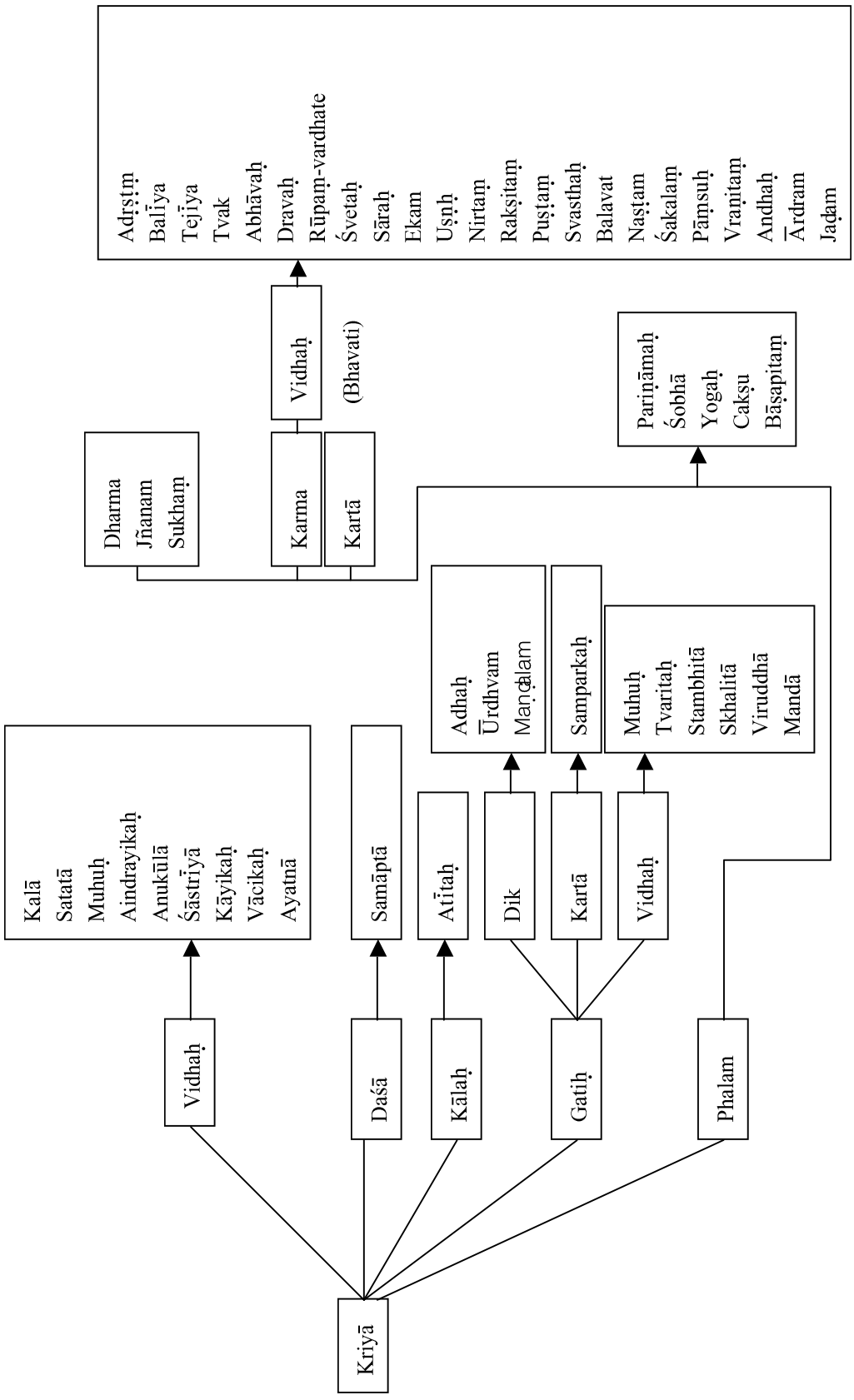

Kriyā
Vidhaḥ
Daśā
Kālaḥ
Gatiḥ
Phalam
Kalā
Satatā
Muhuḥ
Aindrayikaḥ
Anukūlā
Śāstrīyā
Kāyikaḥ
Vācikaḥ
Ayatnā
Samāptā
Atītaḥ
Dik
Kartā
Vidhaḥ
Adhaḥ
Ūrdhvam
Maṇḍalam
Samparkaḥ
Muhuḥ
Tvaritaḥ
Stambhitā
Skhalitā
Viruddhā
Mandā
Dharma
Jñanam
Sukhaṃ
Karma
Kartā
Vidhaḥ
(Bhavati)
Adṛṣṭṃ
Balīya
Tejīya
Tvak
Abhāvaḥ
Dravaḥ
Rūpaṃ-vardhate
Śvetaḥ
Sāraḥ
Ekam
Uṣṇḥ
Nirtaṃ
Rakṣitaṃ
Puṣṭaṃ
Svasthaḥ
Balavat
Naṣṭam
Śakalaṃ
Pāṃsuḥ
Vraṇitaṃ
Andhaḥ
Ārdram
Jaḍam
Pariṇāmaḥ
Śobhā
Yogaḥ
Cakṣu
Bāṣapitaṃ

Figure 9.23

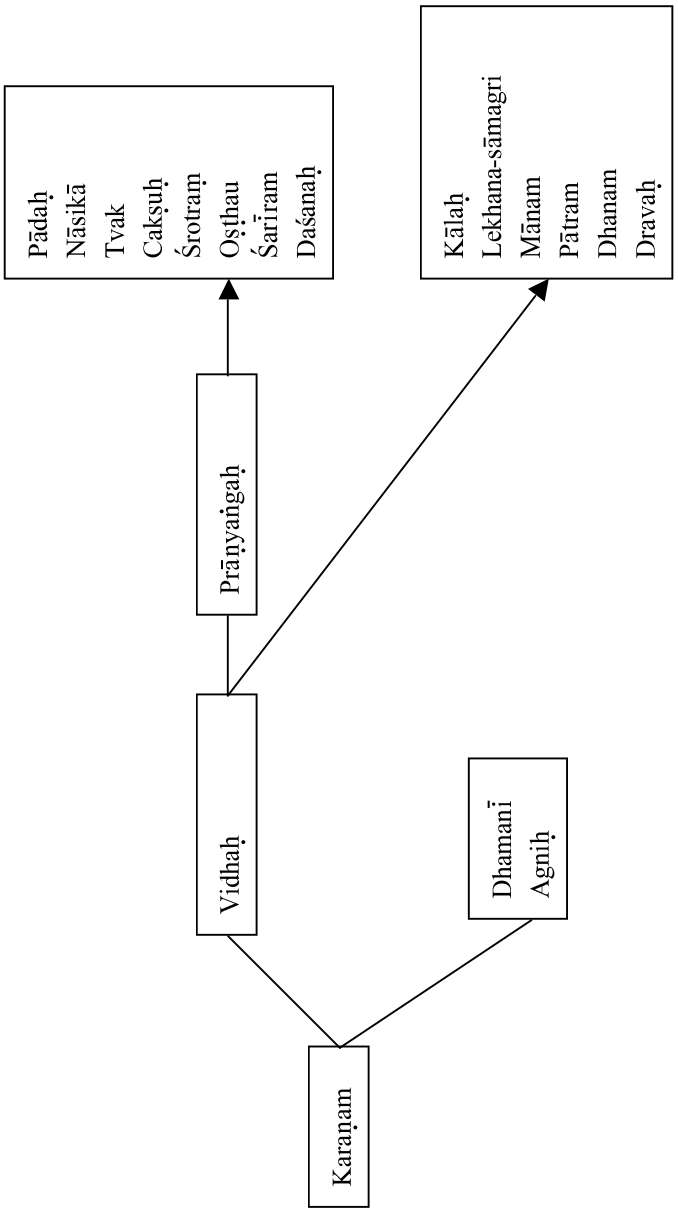

Pādaḥ
Nāsikā
Tvak
Cakṣuḥ
Śrotraṃ
Oṣṭhau
Śarīram
Daśanaḥ
Kālaḥ
Lekhana-sāmagri
Mānam
Pātram
Dhanam
Dravaḥ
Prāṇyaṅgaḥ
Vidhaḥ
Dhamanī
Agniḥ
Karaṇam
Utkaṭaḥ
Vidhaḥ
Prayatnaḥ

Figure 9.24

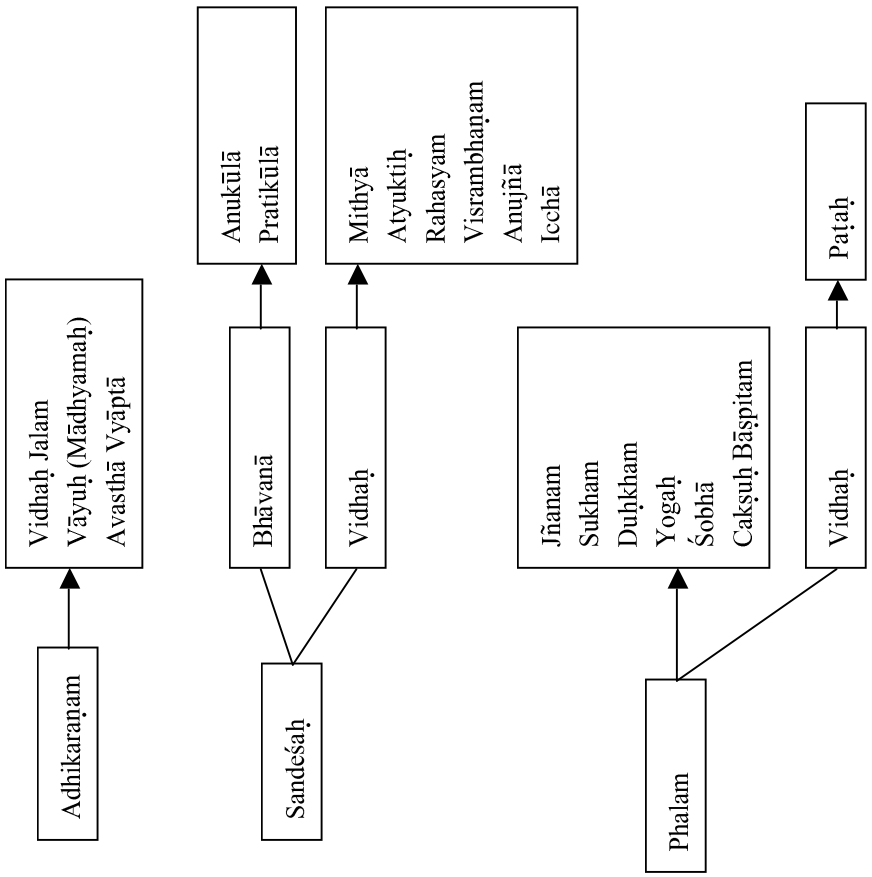

Adhikaraṇam
Vidhaḥ Jalam
Vāyuḥ (Mādhyamaḥ)
Avasthā Vyāptā
Sandeśaḥ
Bhāvanā
Anukūlā
Pratikūlā
Vidhaḥ
Mithyā
Atyuktiḥ
Rahasyam
Visrambhaṇam
Anujñā
Icchā
Phalam
Jñanam
Sukham
Duḥkham
Yogaḥ
Śobhā
Cakṣuḥ Bāṣpitam
Vidhaḥ
Paṭaḥ

Figure 9.25

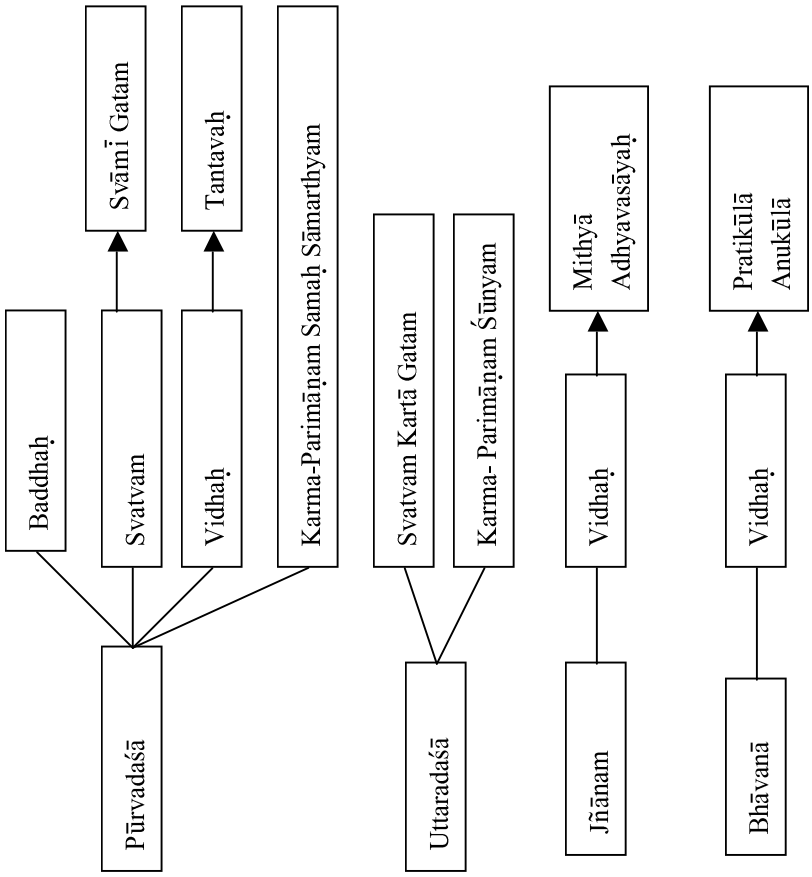
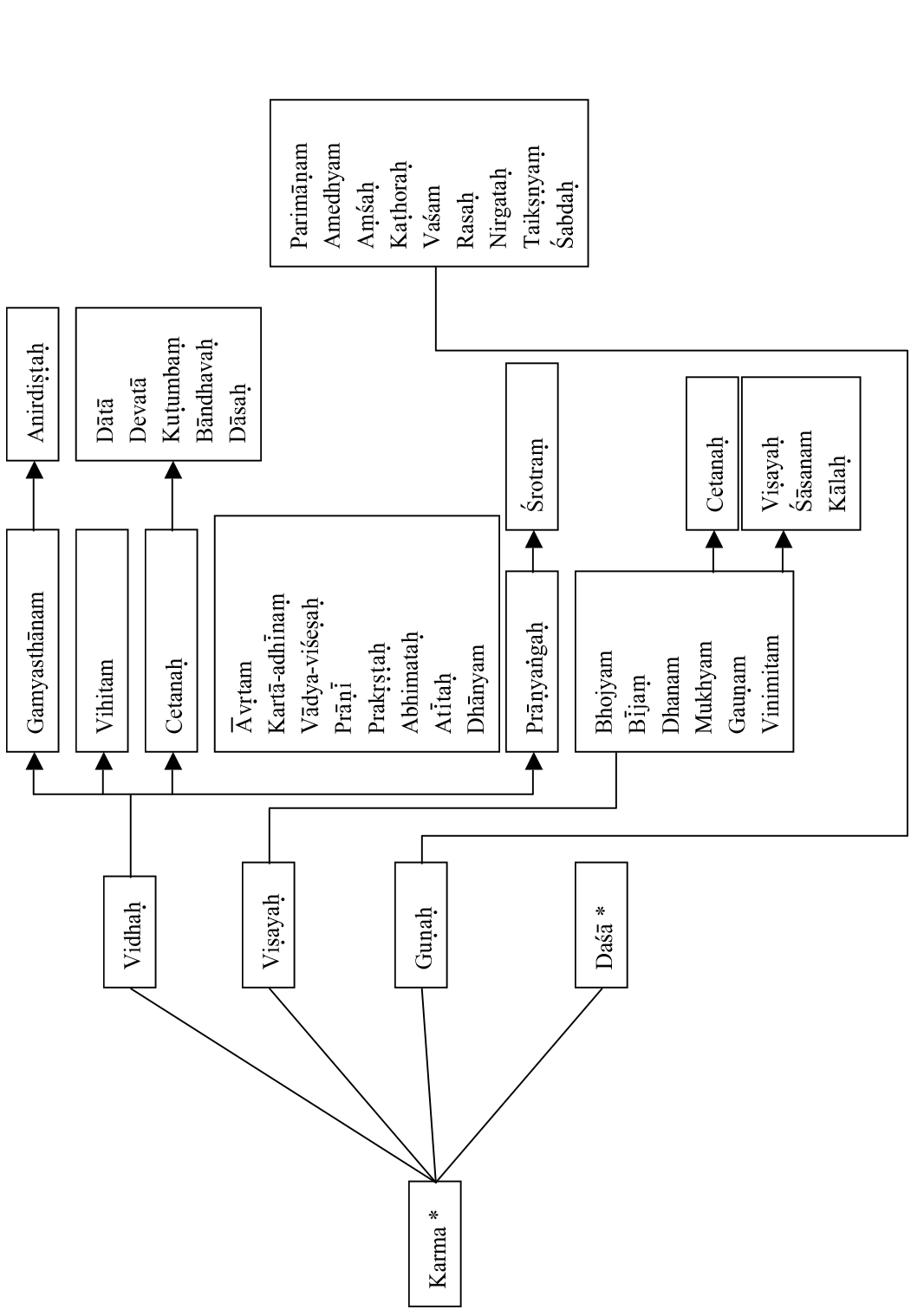
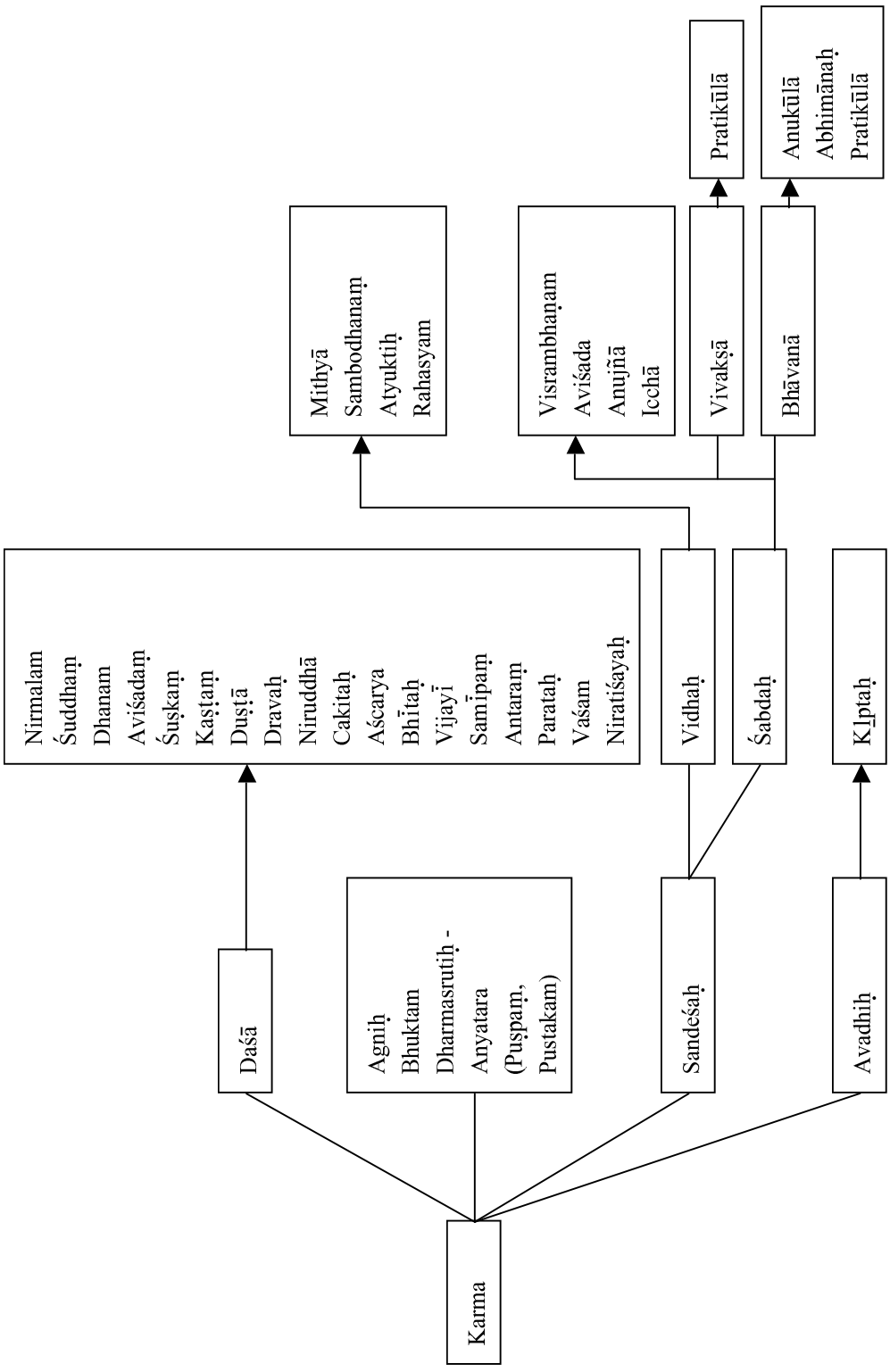

Baddhaḥ
Svatvam
Vidhaḥ
Karma-Parimāṇam Samaḥ Sāmarthyam
Svāmī Gatam
Tantavaḥ
Pūrvadaśā
Svatvam Kartā Gatam
Karma- Parimāṇam Śūnyam
Uttaradaśā
Vidhaḥ
Mithyā
Adhyavasāyaḥ
Jñānam
Vidhaḥ
Pratikūlā
Anukūlā
Bhāvanā

Figure 9.26



* marked item is continued in the next page

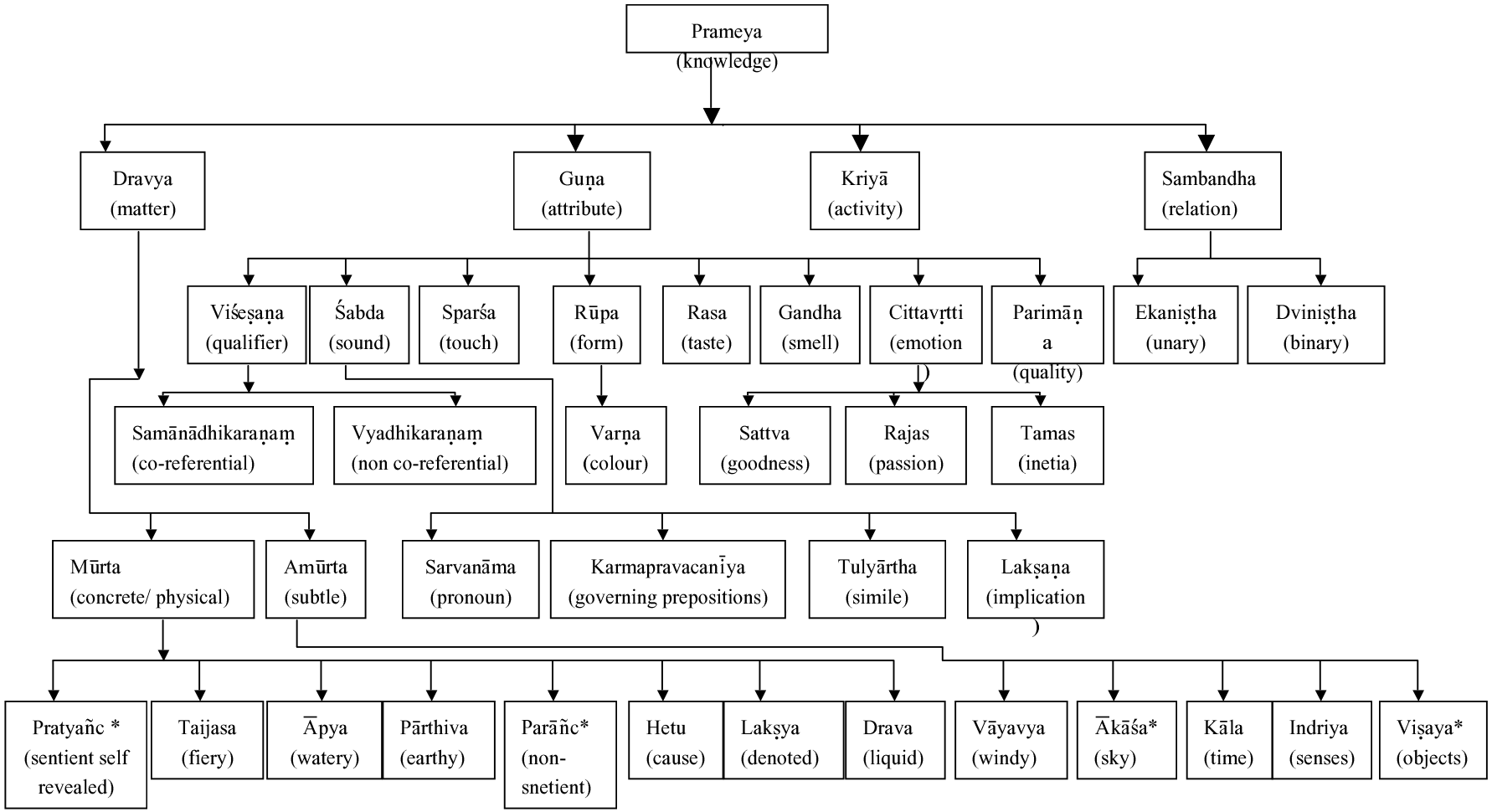
Figure 9.27



Continued from the previous page.

Figure 9.28

Padārtha Vibhāgaḥ - An Ontology based on Grammer rules of Kāraka and Vibhakti



* marked items contd.... in next page

Figure 9.29

Padārtha Vibhāgaḥ - An Ontology based on Grammer rules of Kāraka and Vibhakti

Contd.. From previous page

Figure 9.30

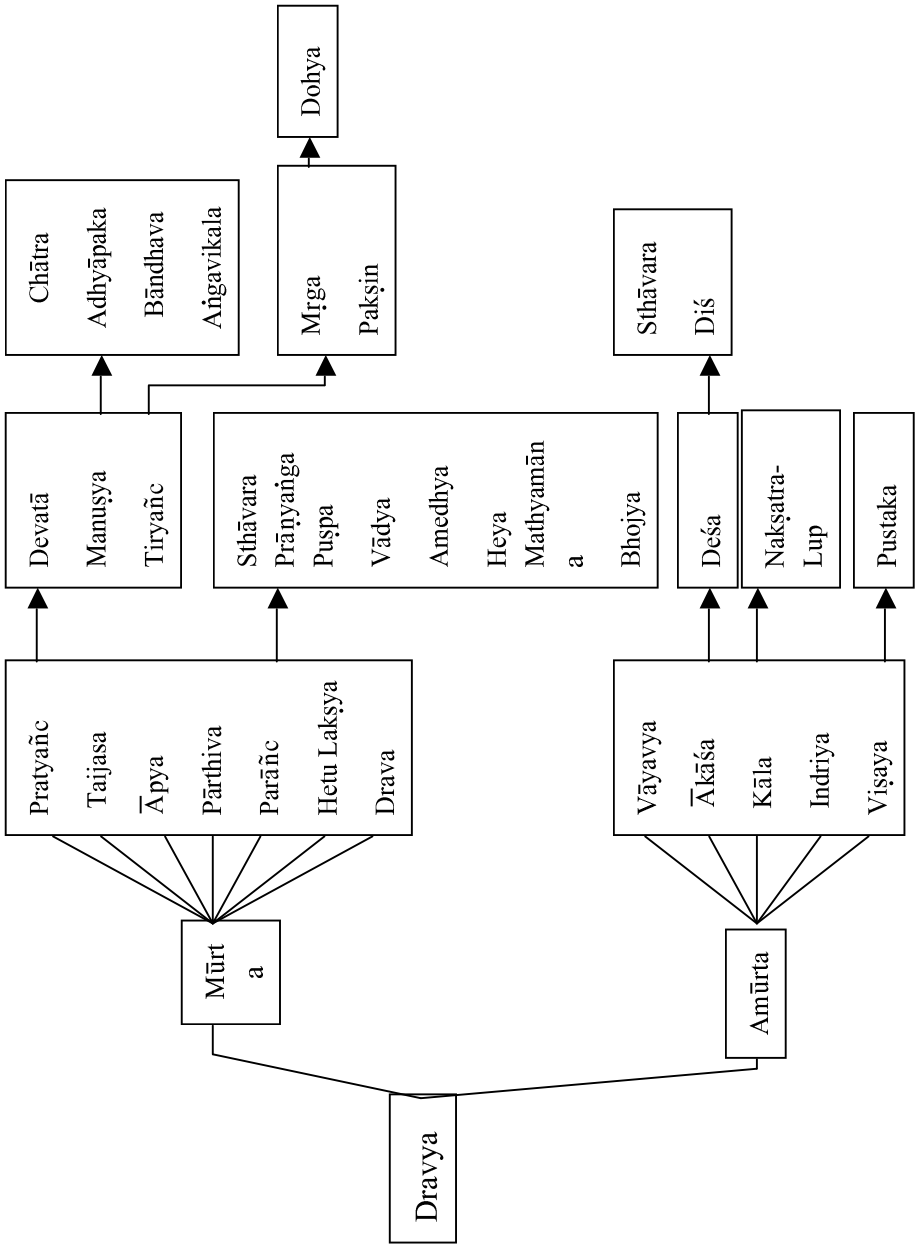

Dravya
Mūrta
Amūrta
Pratyañc
Taijasa
Āpya
Pārthiva
Parāñc
Hetu Lakṣya
Drava
Vāyavya
Ākāśa
Kāla
Indriya
Viṣaya
Devatā
Manuṣya
Tiryañc
Sthāvara
Prāṇyaṅga
Puṣpa
Vādya
Amedhya
Heya
Mathyamān
a
Bhojya
Deśa
Nakṣatra-
Lup
Pustaka
Chātra
Adhyāpaka
Bāndhava
Aṅgavikala
Mṛga
Pakṣin
Dohya
Sthāvara
Diś

Figure 9.31

Padārtha Vibhāgaḥ - An Ontology based on Grammer rules of Kāraka and Vibhakti

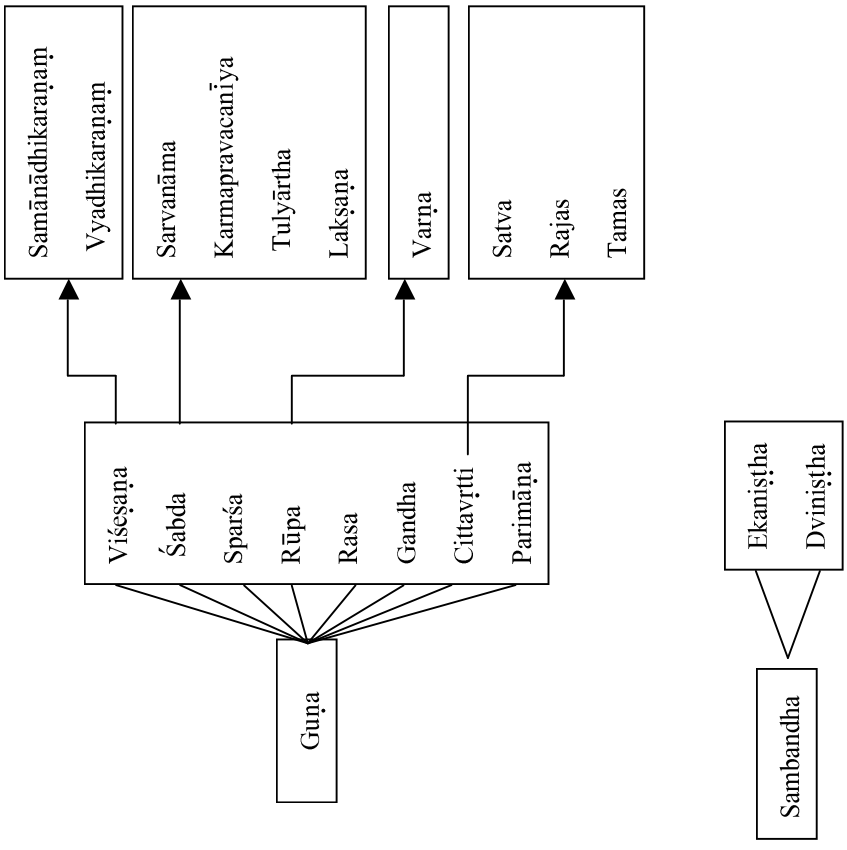


Figure 9.32

Śābda-bodha classification

Diagram -1

Diagram -2

Figure 9.33

Vedic Śābda-bodha

Input sentence : hariḥ om - All udātta at the commencement of Vedic recitation /chanting

Vedopakramanirūpaka-praṇavaḥ - Haririty atra Vartamānavisargo'rdhamātraḥ . Viramastvekamātrākālikaḥ . Dvayoḥ Padayoḥ Sandhyabhāvarūpā Vivṛtiḥ Vatsānusṛtisañjñikā .Praṇavaścaturmātraḥ . Haripadamudāttam . Atra Svaraviśiṣṭo Varṇakramātmako Bodhaḥ - Udāttadharmaviśiṣṭa Okāraḥ . Haririty atra Hakārodāttākāra-rephodāttekārodāttavisarjanīyāḥ Iti .

h —— Hakāra dhvani janita kaṇṭha sthāna tadadhobhāga karaṇābhyantara sañjñika vivṛtaprayatnabāhya sañjñika nādaghoṣa mahāprāṇa prayatnārdhamātrākāla sūryadevatāka śūdrajātīya parasvarāṅga hakāra

a —— Nāda dhvani janita nātyupasaṃhṛtahanusthāna tādṛśoṣṭhakaraṇābhyantarasañjñika saṃvṛtaprayatna bāhyasañjñika saṃvāra nādālpaprāṇa prayatnaikamātrika vāyudevatāka brāhmaṇa jāti sātvikaguṇasahita tarjanyaṅguli madhyarekhāvinyāsa yogyādhyetṛdehadairghya dhvani kāṭhinyakaṇṭhākāśakṛśatvakara niṣādagāndhārasvaradvayahetubhūta mūrdhasthānodbhavodāttadharma viśiṣṭa vāyudevatāka brāmhaṇajātīyākāra

r —— Nāda dhvanijanitadantamūloparibhāgasthānajihvāgramadhyakaraṇābhyantasañjñikeṣa tspṛṣṭaprayatnabāhyasañjñika saṃvāra nādāghoṣālpa prāṇaprayatnāgnidevatākavaiśyajātīyaparasvarāṅga repha

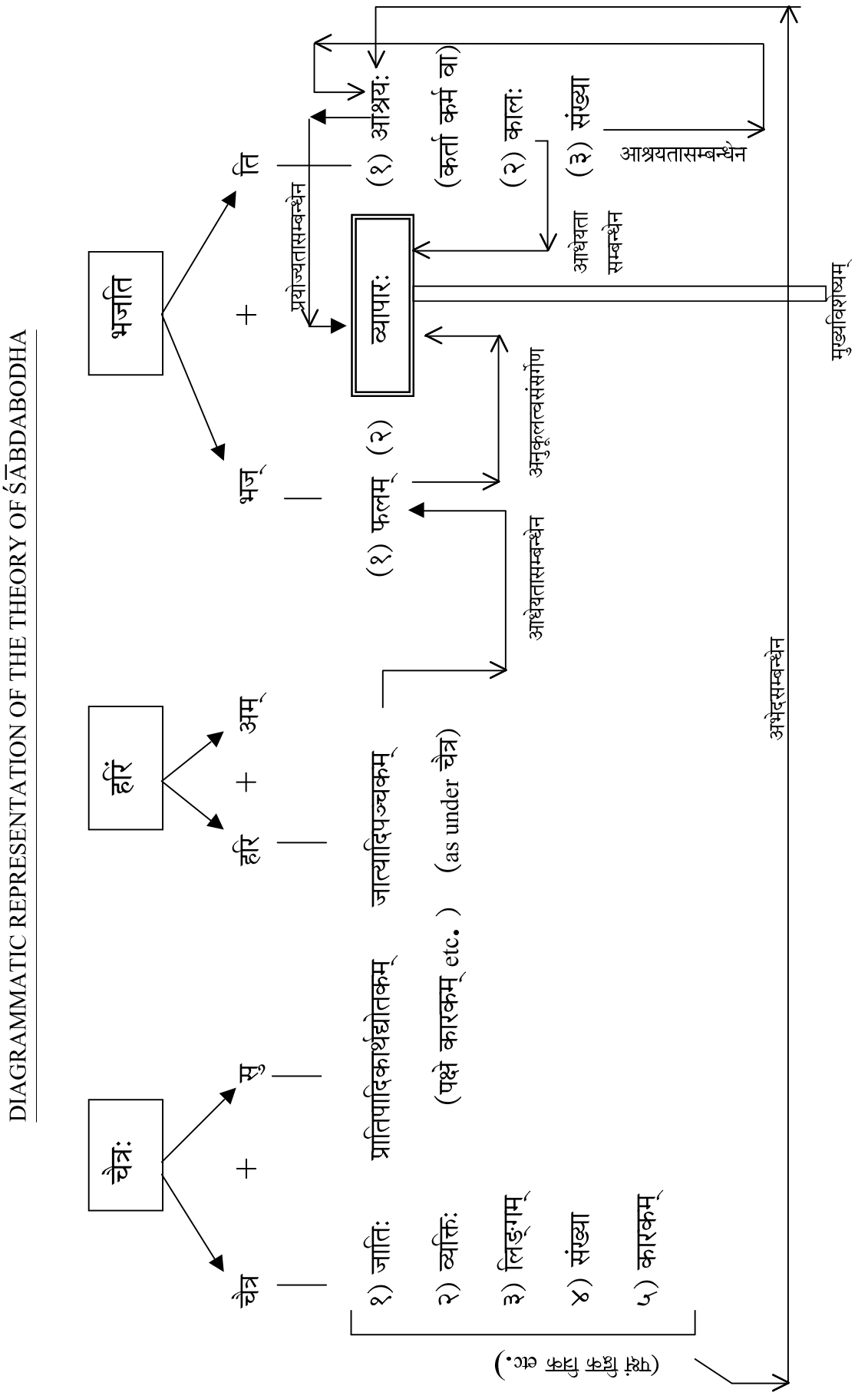
i —— Nāda dhvatijanita tālusthāna jihvāmadhyakaraṇavivṛtaprayatnaikamātrika vāyudevatāka brāhmaṇa jāti sātvikaguṇasahita tarjanyaṅgulimadhyarekhāvinyāsayogyādhyetṛdehadairghya dhvanikāṭhinyakaṇṭhākāśakṛśatvakara niṣādagāndhārasvaradvaya hetubhūta mūrdhasthānodbhavodāttadharmaviśiṣṭekāra

ḥ —— śvāsa dhvanijanita hanūmūlasthānajihvāmūlakaraṇavivṛtaprayatnārdhamātrākāla vāyudevatāka śūdra jātivisarjanīya

o —— Mātrikavatsānusṛti sañjñika vivṛttivirāma nāda dhvani janita hanūttaroṣṭobhayasthāna nātyupasaṃhṛtauṣṭhādharoṣṭho bhayakaraṇa vivṛtaprayatna traimātrika bhūmidevatāka brāhmaṇajāti sātvikaguṇasahita tarjanyaṅgulimadhyarekhāvinyāsayogyādhyetṛdehadairghya dhvani kāṭhinyakaṇṭhākāśakṛtvakara niṣādagāndhārasvaradvayahetubhūta mūrdhasthānodāttadharmaviśiṣṭa okāra

m —— Nāda dhvani janitottaroṣṭhasthānādharoṣṭhakaraṇābhyantara sañjñika spṛṣṭaprayatna bāhyasañajñikasaṃvāraghoṣālpaprāṇa prayatnaikamātrikārkadevatāka vaiśyajāti pūrvasvarāṅgamakārāḥ

Figure 9.34



I The Vaiyākaraṇa Theory of Śābdabodha (वैयाकरण शाब्दबोधः)

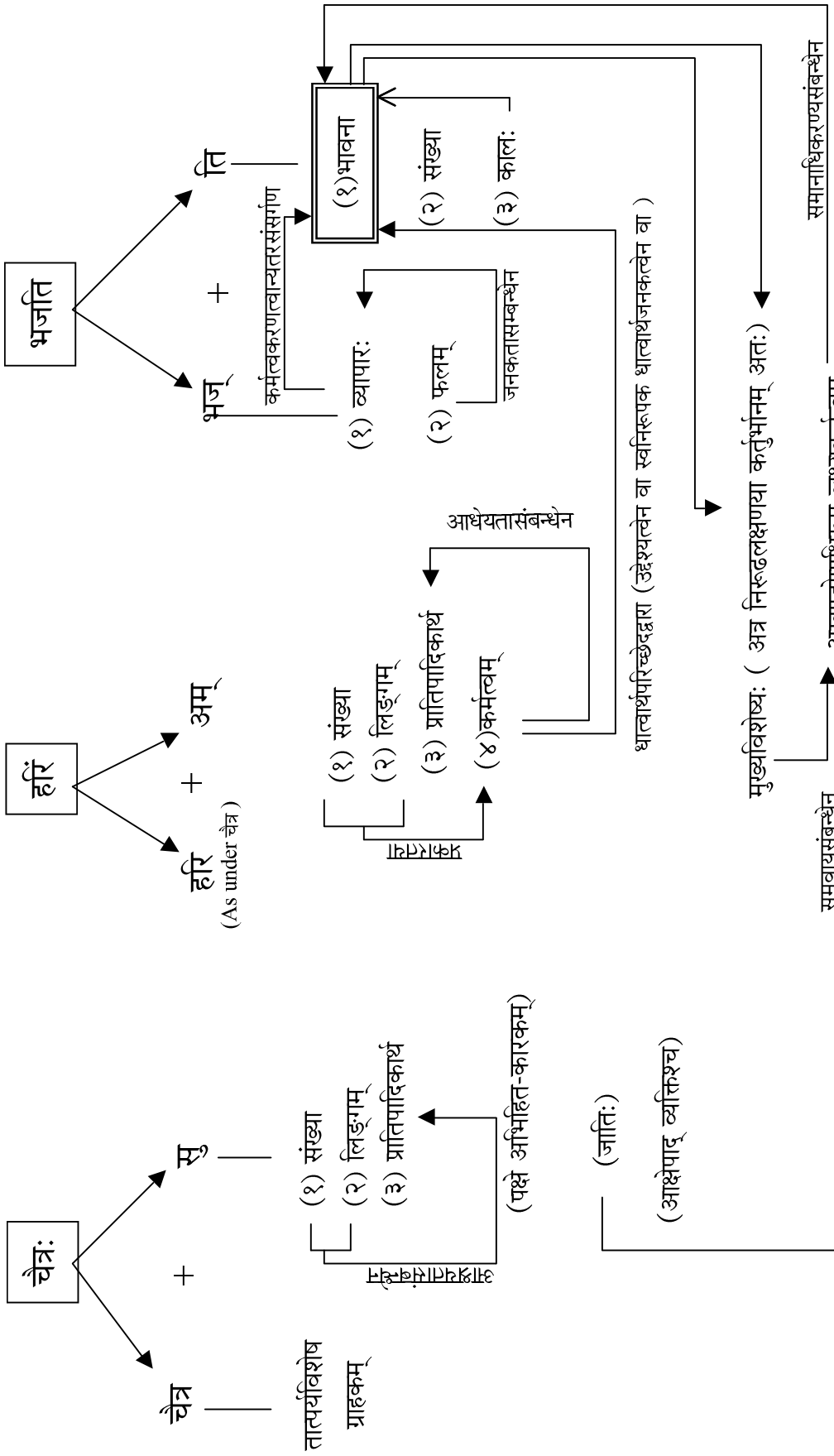
Figure 9.35

DIAGRAMMATIC REPRESENTATION OF THE THEORY OF ŚĀBDABODHA

II The logical Theory of Śābdabodha (नैयायिक शाब्दबोधः)

Figure 9.36

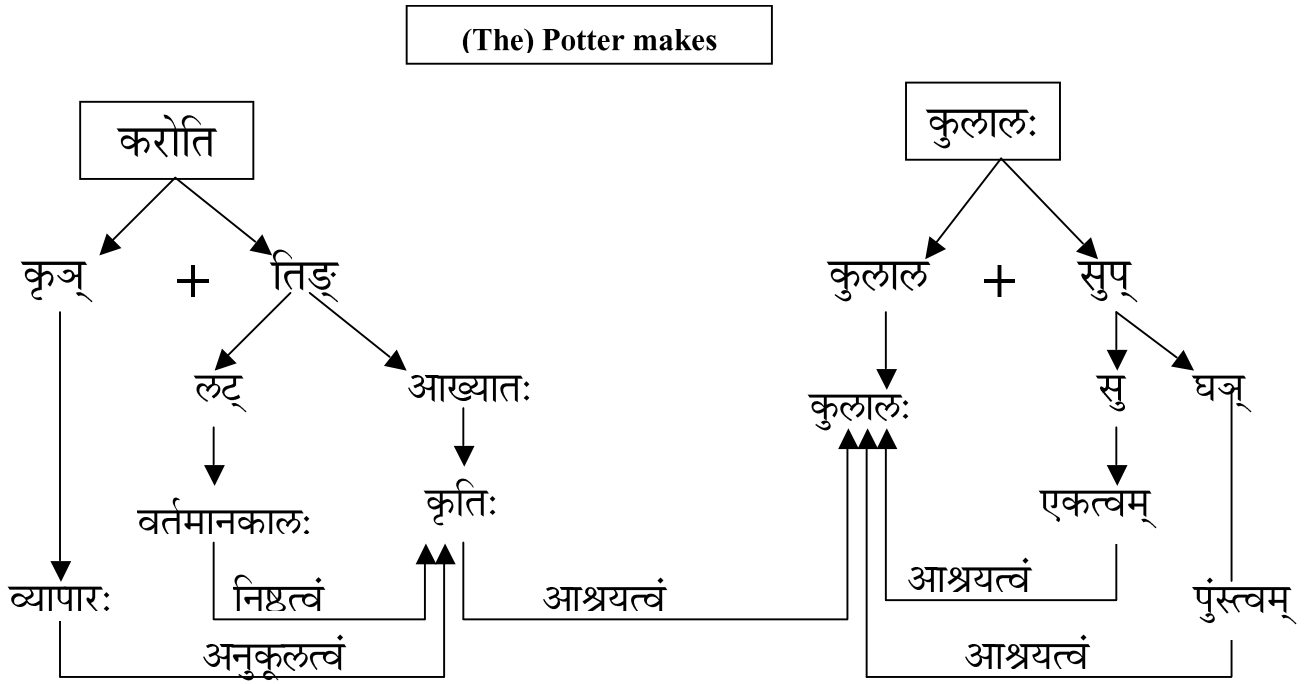
DIAGRAMMATIC REPRESENTATION OF THE THEORY OF ŚĀBDABODHA



III Mīmāṃsā Theory of Śabdabodha
(मीमांसकशाब्दबोधः)

Figure 9.37

नैयायिक शाब्दबोध प्रक्रिया



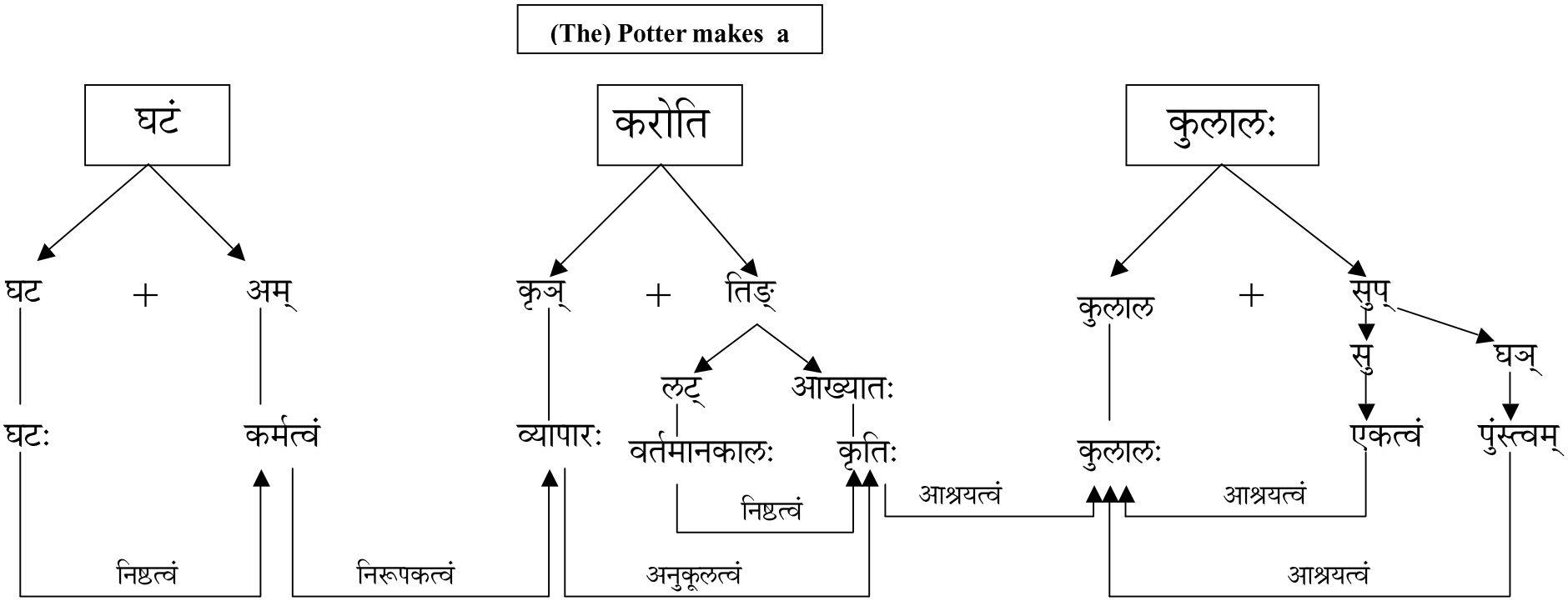
परिवृत्तिसहम् - प्रकृत्यर्थद्वयम् - 2 Variables; परिवृत्त्यसहम् - प्रत्ययार्थत्रयम् , सम्बन्धचतुष्टयम् - 3 + 4 Constants

सम्बन्धः - आश्रयता - सम्बन्धोऽयं सर्वत्र आख्यातार्थ-प्रकृत्यर्थयोः एव । अनुकूलता - सम्बन्धोऽयं सर्वत्र धात्वर्थाख्यातार्थयोः एव ।

शाब्दबोधः - व्यापारानुकूल वर्तमानकालवृत्ति कृत्याश्रयैकत्वाश्रय पुंस्त्वाश्रयः कुलालः ।

Naiyāyika-Śābda-bodhaḥ - Prathama-staraḥ - Prathamānta-Subanta-kriyā-Pada-yutaḥ - Kartṛ-kārakam

Figure 9.38



परिवृत्तिसहम् - प्रकृत्यर्थत्रयम् 3 Variables ; परिवृत्त्यसहम् - प्रत्ययार्थपञ्चकम् + -सम्बन्धसप्तकम् 5+7 = 12 Constants

नियमाः -

१. सुबर्थ-प्रकृत्यर्थयोः घञर्थ-प्रकृत्यर्थयोः सर्वत्र आश्रयत्वम् । एवं आख्यातार्थ-प्रकृत्यर्थयोः च ।

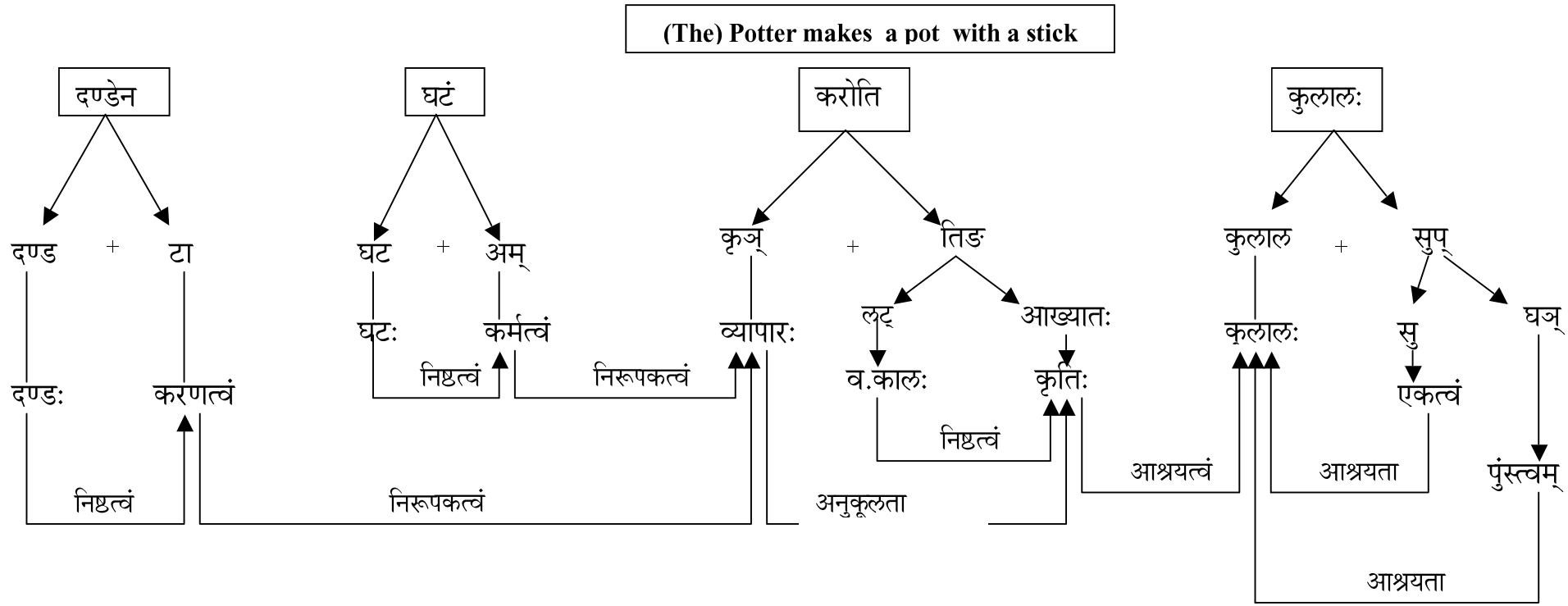
२. प्रकृत्यर्थ-प्रत्ययार्थयोः लडर्थाख्यातार्थयोः निष्ठत्वम् ।

३. प्रत्ययार्थ-धात्वर्थयोः सर्वत्र निरूपकत्वमर्थः । अपवादः - षष्ठीविभक्तिप्रत्ययार्थस्य तु सप्तम्याः प्रकृतौ अन्वयः ।

शाब्दबोधः - घटनिष्ठ-कर्मता-निरूपक-व्यापारानुकूल-वर्तमानकाल-वृत्ति-कृत्याश्रयैकत्वाश्रयपुंस्त्वाश्रयः कुलालः।

Naiyāyika-Śābda-bodhaḥ - Dvitīya-staraḥ - Kartṛ-karma-kriyā-Pada-yutaḥ - Karma-kārakam

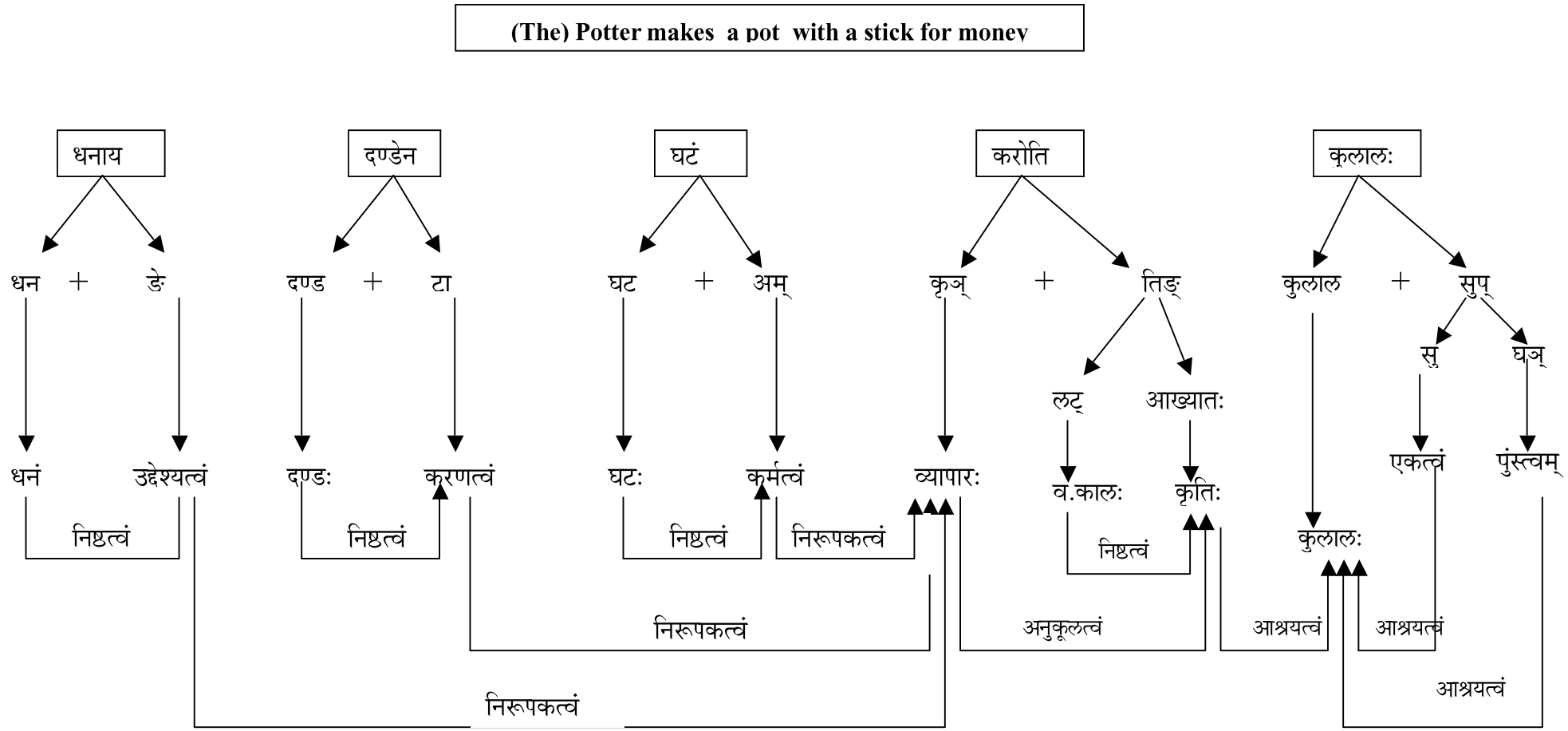
Figure 9.39



शाब्दबोधः - दण्डनिष्ठ-करणतानिरूपक-घटनिष्ठ-कर्मता-निरूपक-व्यापारानुकूल-वर्तमानकाल-वृत्ति-कृत्याश्रयैकत्वाश्रयपुंस्त्वाश्रयः कुलालः।

Naiyāyika-Śābda-bodhaḥ - Tṛtīya-staraḥ - Kartṛ-karma-Karaṇa-kriyā-Pada-yutaḥ - Karaṇa-kārakam

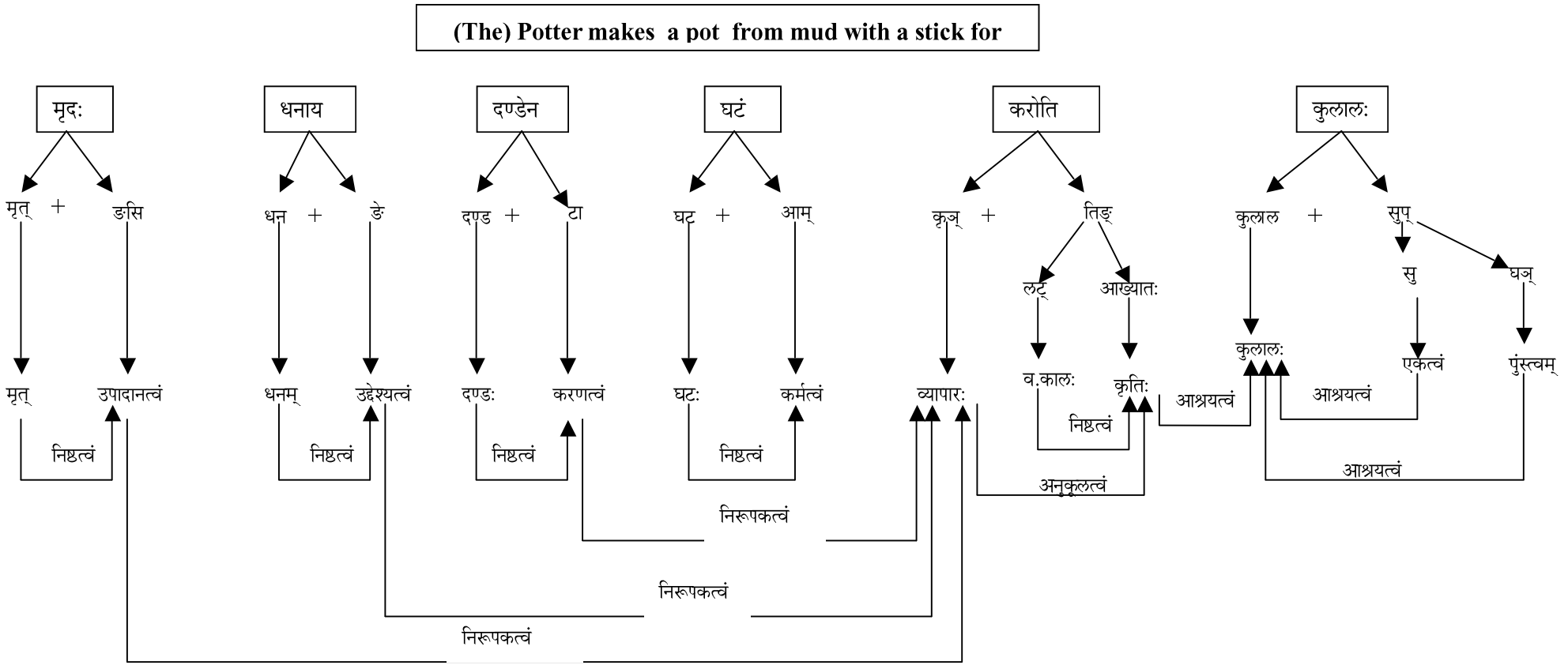
Figure 9.40



शाब्दबोधः - धननिष्ठोद्देश्यतानिरूपक-दण्डनिष्ठकरणतानिरूपक-घटनिष्ठकर्मतानिरूपक-व्यापारानुकूल-वर्तमानकालनिष्ठ-कृत्याश्रयैकत्वाश्रयपुंस्त्वाश्रयः कुलालः [illegible]

Naiyāyika-Śābda-bodhaḥ - Caturtha-staraḥ - Kartṛ-karma-Karaṇa-sampradāna-kriyā-Pada-yutaḥ - Sampradāna-kārakam

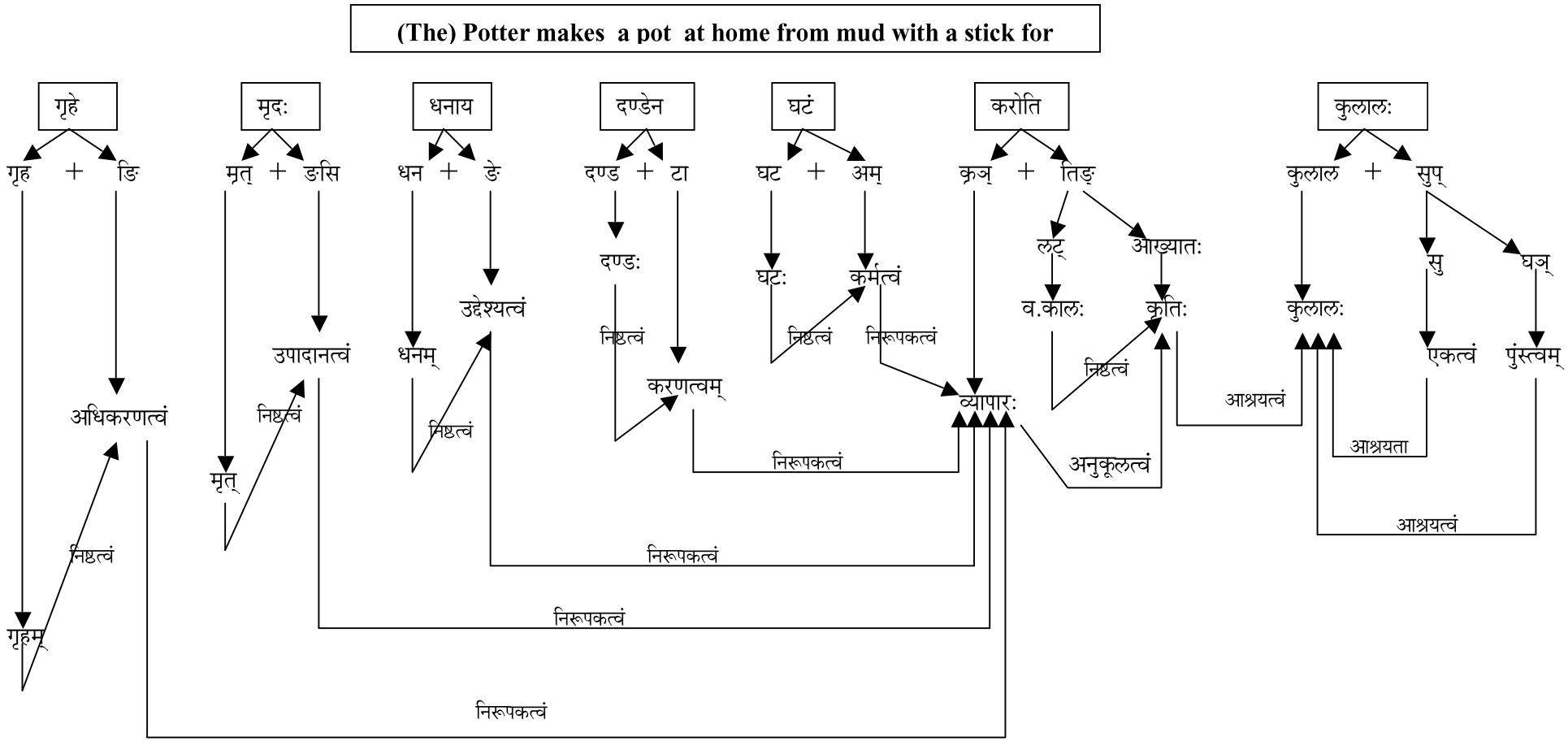
Figure 9.41



शाब्दबोधः :- मृन्निष्ठोपादानतानिरूपक-धननिष्ठोद्देश्यतानिरूपक-दण्डनिष्ठकरणतानिरूपक-घटनिष्ठकर्मतानिरूपक-व्यापारानुकूल-वर्तमानकालवृत्ति-कृत्याश्रयैकत्वाश्रयःपुंस्त्वाश्रयः कुलालः

Naiyāyika-Śābda-bodhaḥ - Pañcama-staraḥ - Kartṛ-karma-Karaṇa-sampradāna-Apādāna-kriyā-Pada-yutaḥ - Apādāna-kārakam

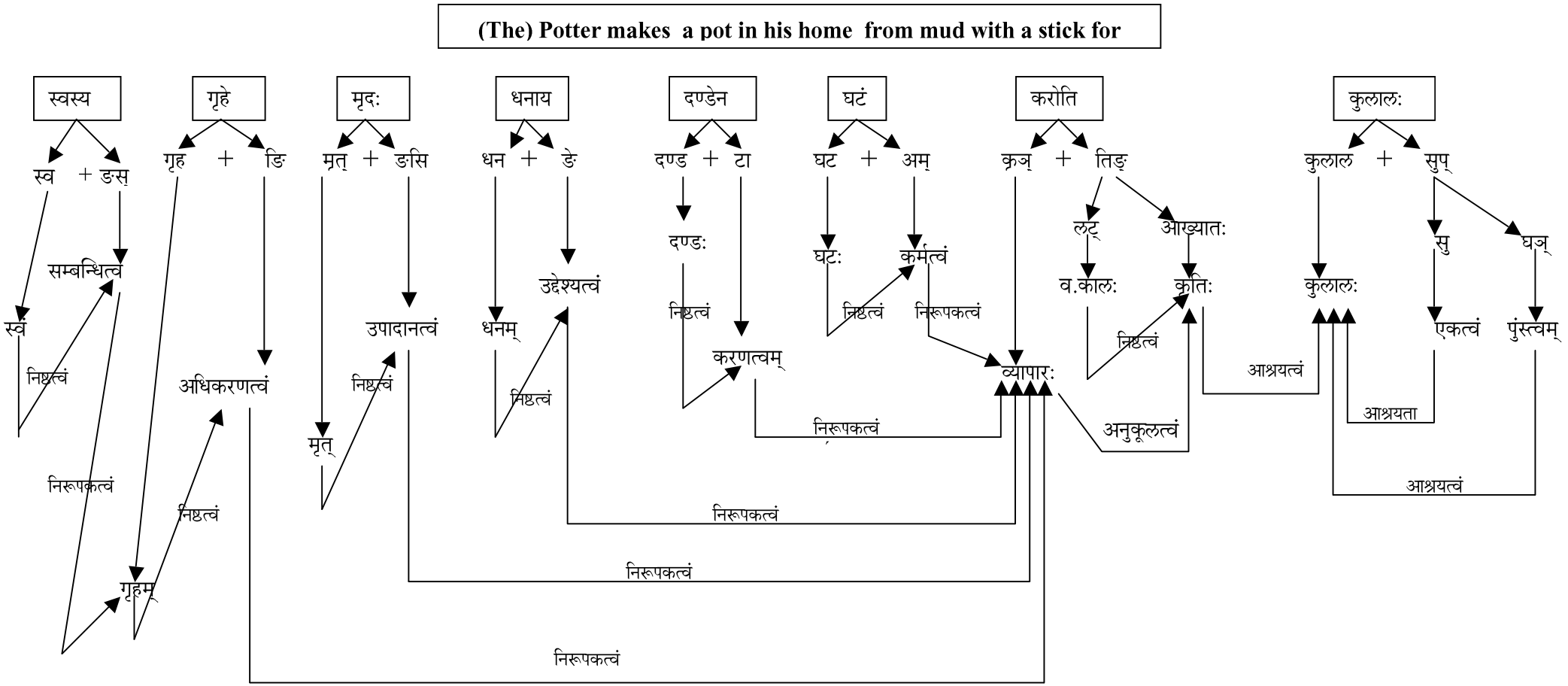
Figure 9.42



शाब्दबोधः - गृहनिष्ठाधिकरणतानिरूपक-मृन्निष्ठोपादानतानिरूपक-धननिष्ठोद्देश्यतानिरूपक-दण्डनिष्ठकरणतानिरूपक-घटनिष्ठकर्मतानिरूपक-व्यापारानुकूल-वर्तमानकालवृत्ति-कृत्याश्रयैकत्वाश्रयः पुंस्त्वाश्रयः कुलालः

Naiyāyika-Śābda-bodhaḥ - Ṣaṣṭha-staraḥ - Kartṛ-karma-Karaṇa-sampradāna-Apādāna-Adhikaraṇa-kriyā-Pada-yutaḥ - Adhikaraṇa-kārakam

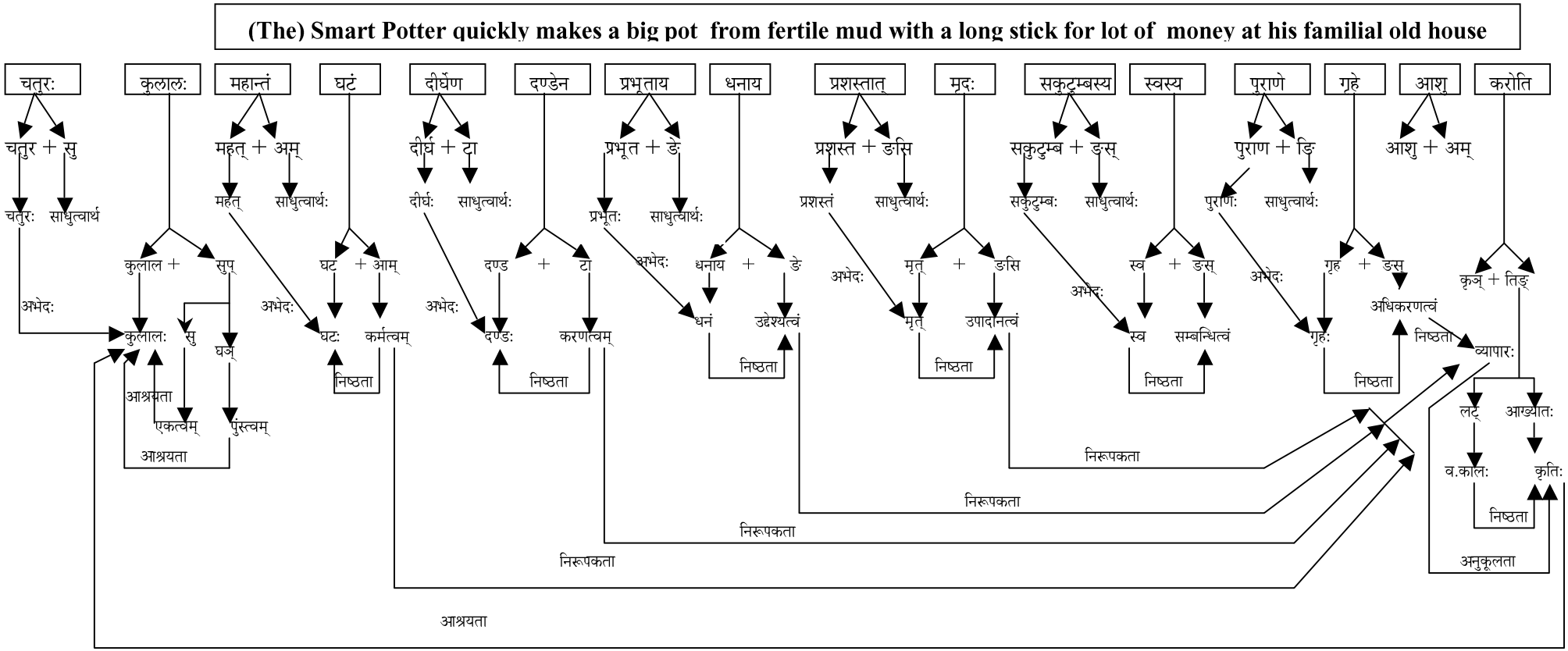
Figure 9.43



शाब्दबोधः - -स्वनिष्ठसम्बन्धितानिरूपक-गृहनिष्ठाधिकरणतानिरूपक-मृन्निष्ठोपादानतानिरूपक-धननिष्ठोद्देश्यतानिरूपक-दण्डनिष्ठकरणतानिरूपक-घटनिष्ठकर्मतानिरूपक-व्यापारानुकूल-वर्तमानकालवृत्ति-कृत्याश्रयैकत्वाश्रय-पुंस्त्वाश्रयः कुलालः

Naiyāyika-Śābda-bodhaḥ-Saptama-staraḥ-Kartṛ-karma-Karaṇa-sampradāna-Apādāna-Adhikaraṇa-Sambandha-kriyā-Pada-yutaḥ-Sambandhaḥ

Figure 9.44

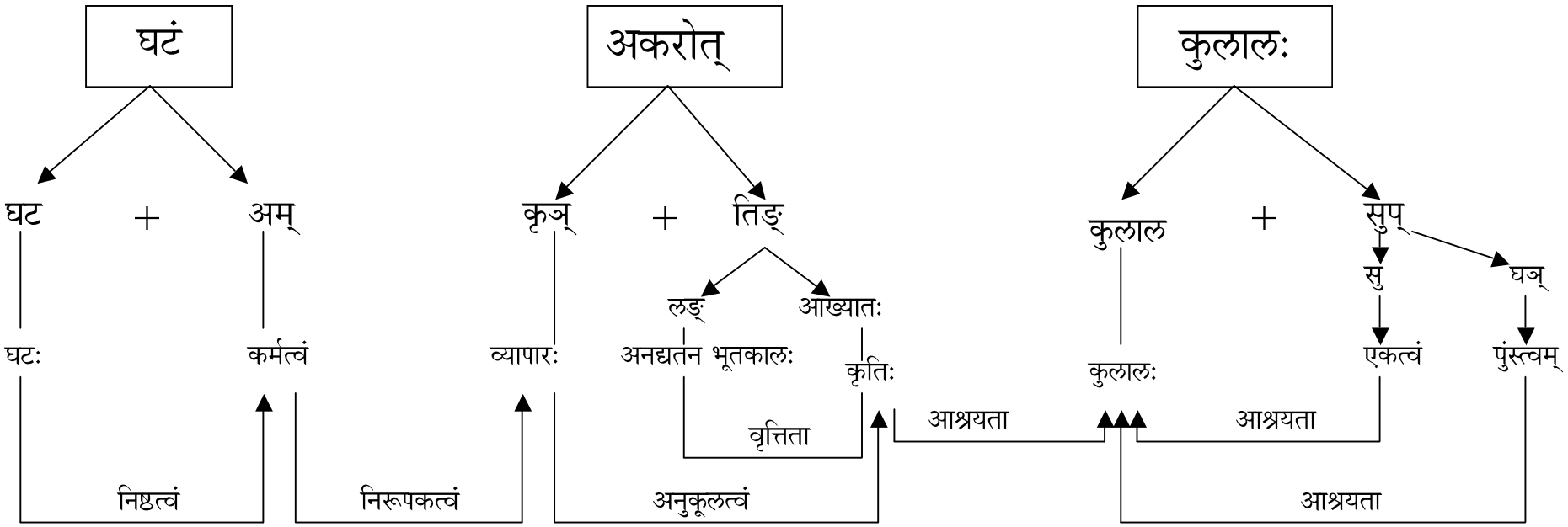


शाब्दबोधः - महदभिन्न-घटनिष्ठकर्मतानिरूपक-दीर्घाभिन्नदण्डनिष्ठकरणतानिरूपक-प्रभूताभिन्नधननिष्ठोद्देश्यतानिरूपक-प्रशस्ताभिन्न-मृन्निष्ठोपादानतानिरूपक-सकुटुम्बाभिन्न-स्वनिष्ठसम्बन्धितानिरूपक-पुराणाभिन्नगृहनिष्ठाधिकरणतानिरूपक-आश्वभिन्न-व्यापारानुकूल-वर्तमानकालनिष्ठ-कृत्याश्रयः एकत्वाश्रयः पुंस्त्वाश्रयः चतुराभिन्न-कुलालः ।

Naiyāyika-Śābda-bodhaḥ - Aṣṭama-staraḥ - Saviśeṣaṇa-Sarvakāraka-Sasambandha-kriyā-Pada-yutaḥ - Viśiṣṭa-Bodhaḥ

Figure 9.45

(The) Potter made a pot (imperfect tense)

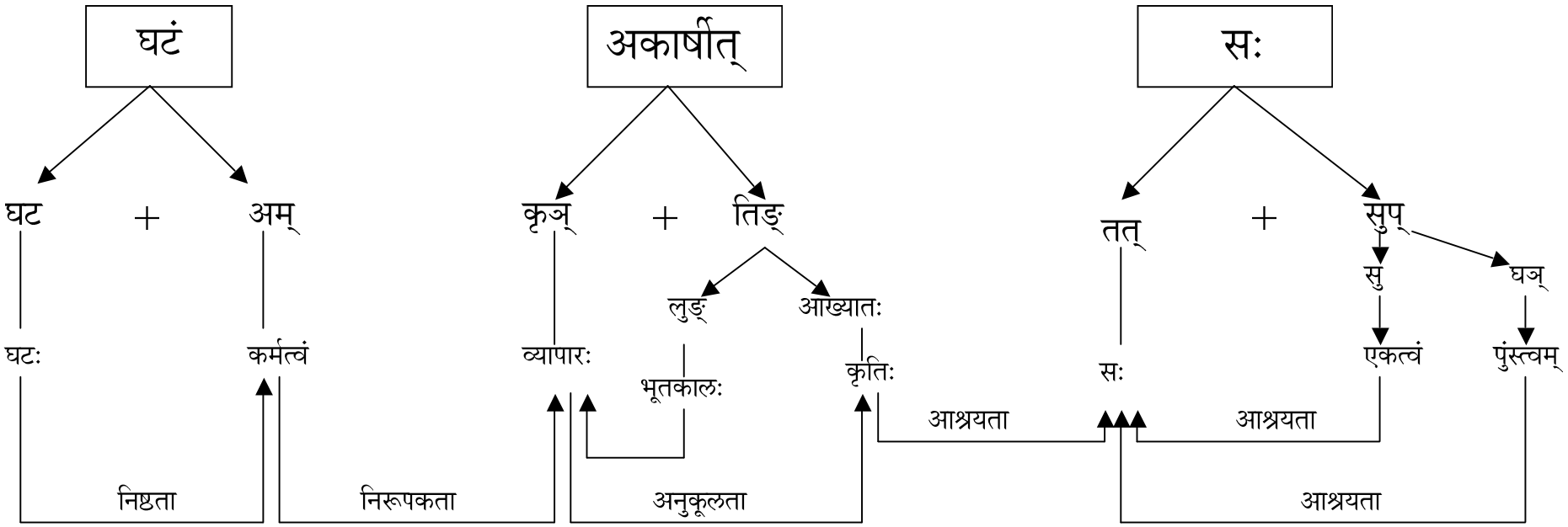


शाब्दबोधः - घटनिष्ठ-कर्मता-निरूपक-अनद्यतन-भूतकाल-वृत्ति-कृत्याश्रयैकत्वाश्रय-पुंस्त्वाश्रयः कुलालः

Naiyāyika-Śābda-bodhaḥ - Dvitīya-staraḥ - Kartṛ-Karma-kriyā-Pada-yutaḥ - Karma-kārakam (Bhūta-kālikaḥ)

Figure 9.46

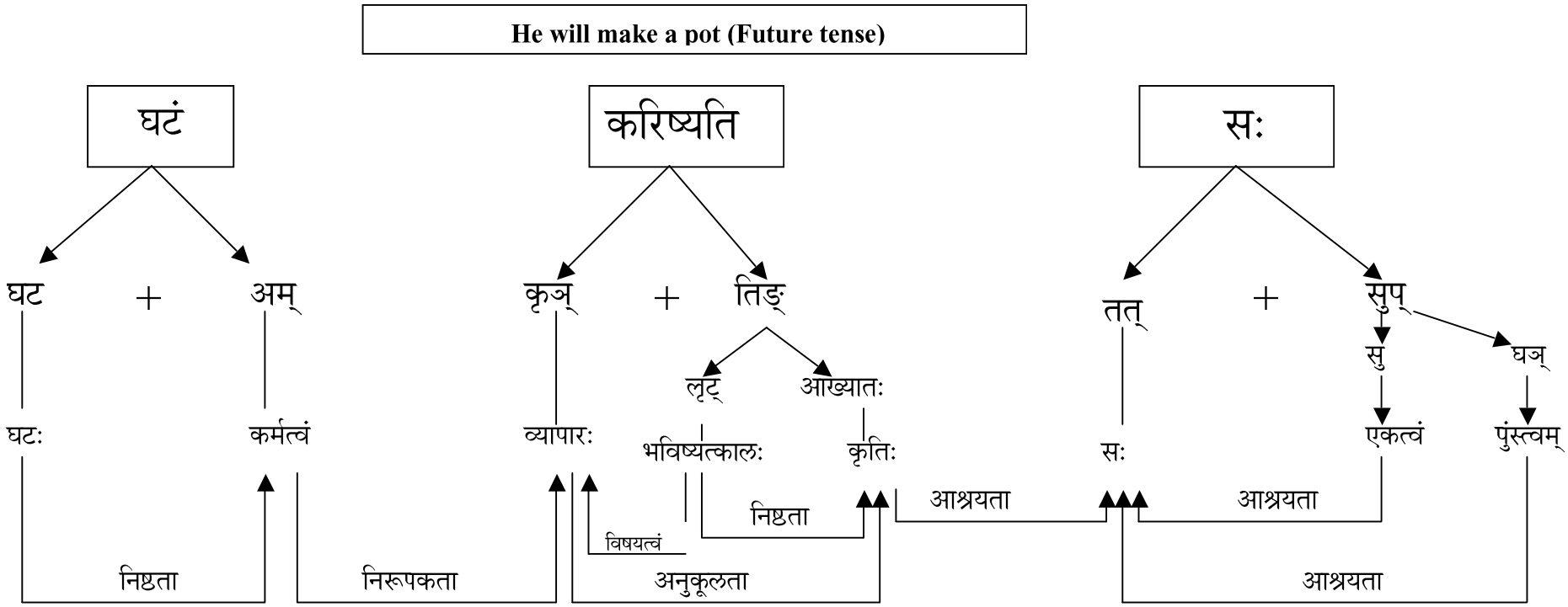
He made a pot (Aorist tense)



शाब्दबोधः - घटनिष्ठकर्मतानिरूपक-व्यापारानूकूल-भूतकालवृत्ति-कृत्याश्रयः एकत्वाश्रयः पुंस्त्वाश्रयः सः (कुलालः) ।

Naiyāyika-Śābda-bodhaḥ - Dvitīya-staraḥ - Kartṛ-Karma-kriyā-Pada-yutaḥ - Karma-kārakam (Anadyatana-Bhūta-kālikaḥ)

Figure 9.47

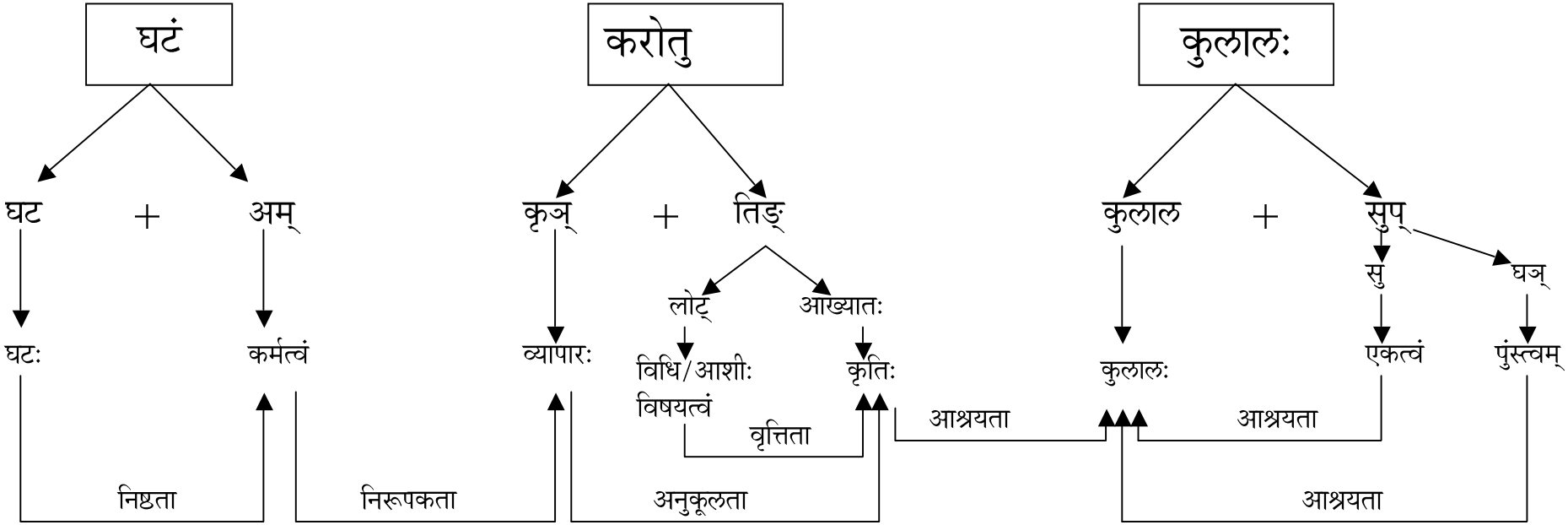


शाब्दबोधः - घट-निष्ठकर्मतानिरूपक-व्यापारानूकूल-भविष्यकाल-विषयः व्यापारः तदनुकूल-कृत्याश्रयः एकत्वाश्रयः पुंस्त्वाश्रयः सः (कुलालः) ।

Naiyāyika-Śābda-bodhaḥ - Dvitīya-staraḥ - Kartṛ-Karma-kriyā-Pada-yutaḥ - Karma-kārakam (Bhaviṣyat-kālikaḥ)

Figure 9.48

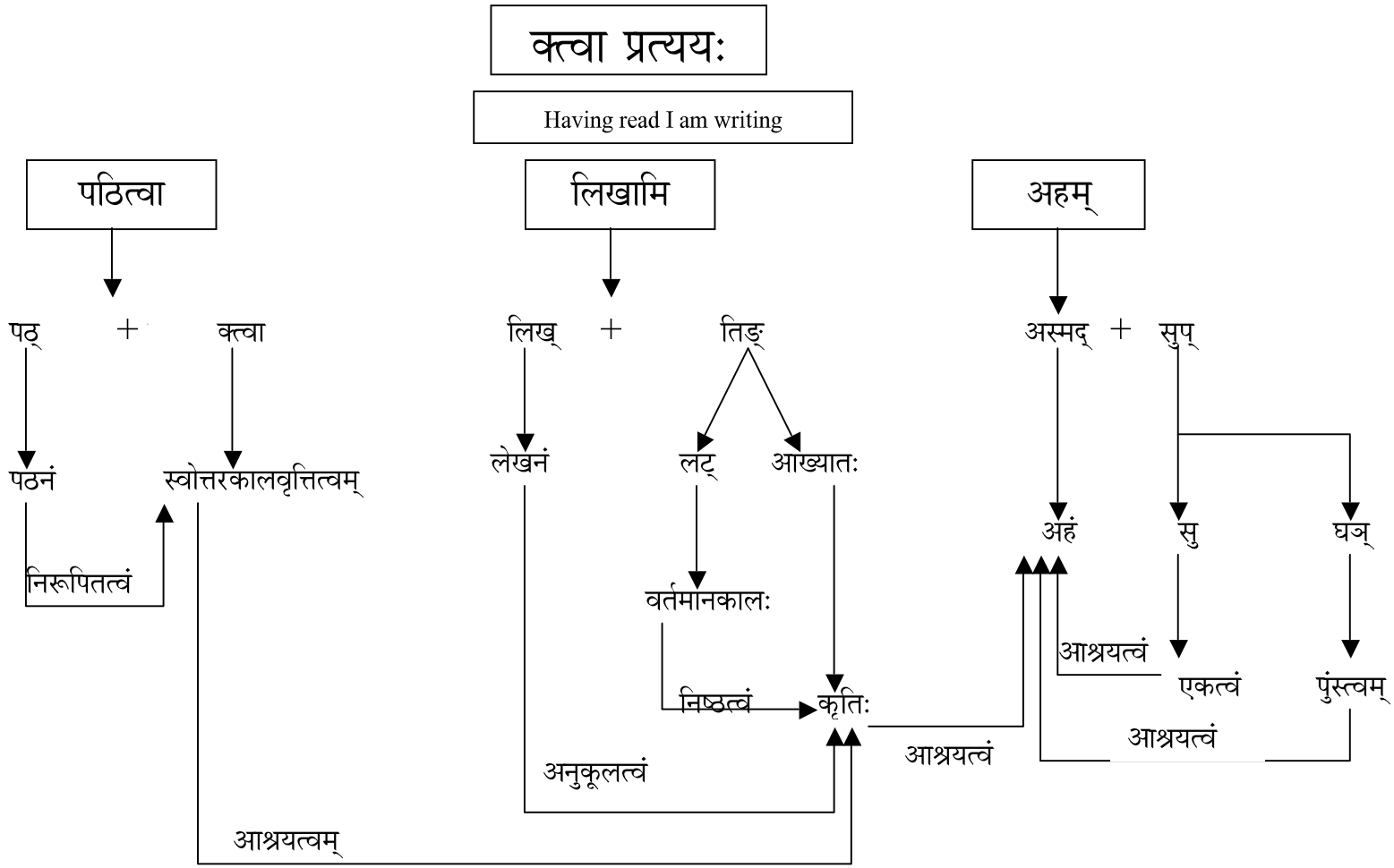
Let The Potter make a pot (optative/potential mode)



शाब्दबोधः - घटनिष्ठ-कर्मता-निरूपक-व्यापारानुकूल-विध्याशीरन्यतर-विषय-वृत्ति-कृत्याश्रयैकत्वाश्रय-पुंस्त्वाश्रयः कुलालः

Naiyāyika-Śābda-bodhaḥ - Dvitīya-staraḥ - Kartṛ-Karma-kriyā-Pada-yutaḥ - Karma-kārakam (Vidhi/Āśīḥ Anyatarārthakaḥ)

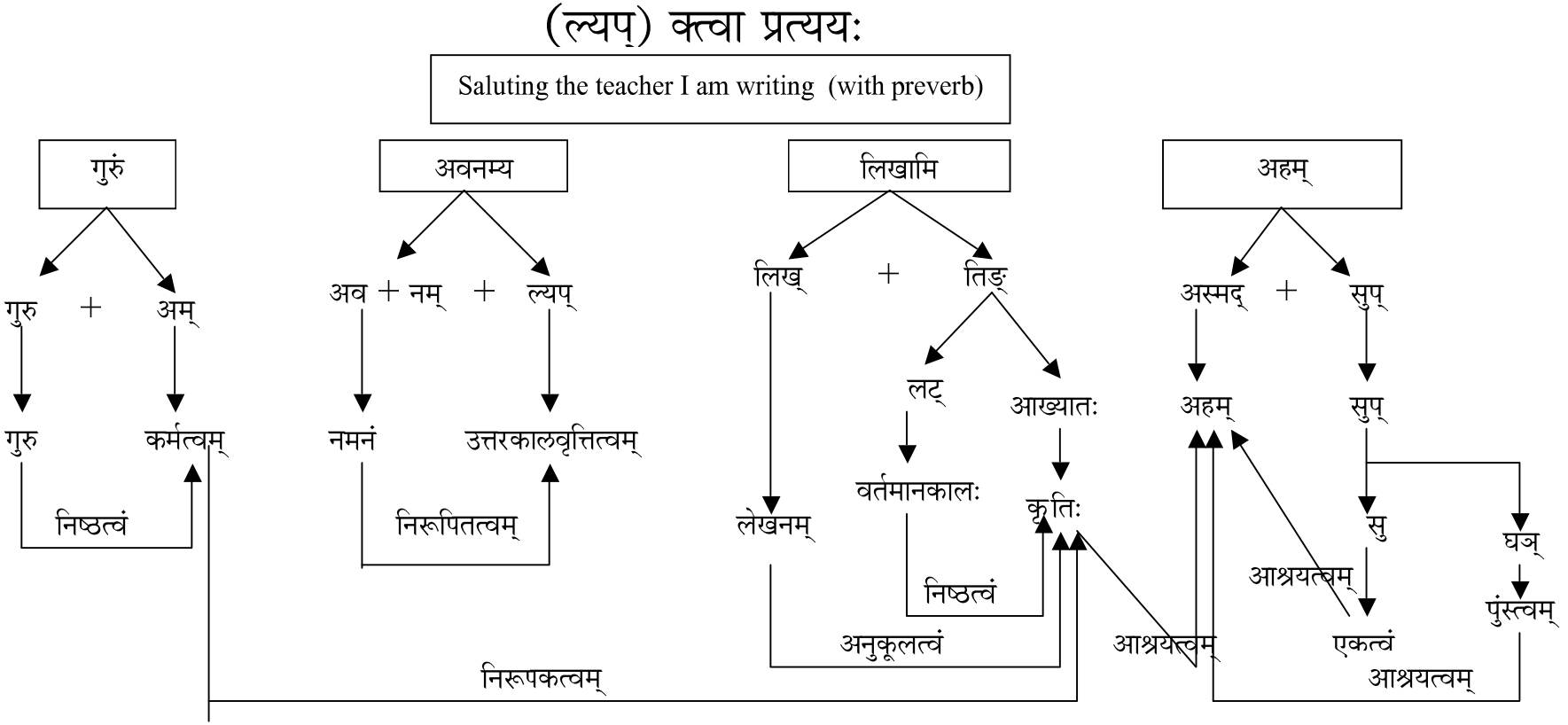
Figure 9.49



शाब्दबोधः - पठननिरूपितस्वोत्तरकालवृत्तित्वाश्रयः लेखनानुकूलवर्तमानकाल निष्ठकृत्याश्रयः एकत्वाश्रयः पुंस्त्वाश्रयः अहम् । or
पठननिरूपितस्वोत्तरकालवृत्तित्ववल्लेखनानुकुलवर्तमानकालनिष्ठकृत्याश्रयैकत्वाश्रयपुंस्त्ववानहम्

Naiyāyika-Śābda-bodhaḥ - Prathama-staraḥ- Kartṛ-kriyā-Pada-yutaḥ - Kramika-kriyā-dvayam (Pūrvakāla-kriyārthakaḥ)

Figure 9.50

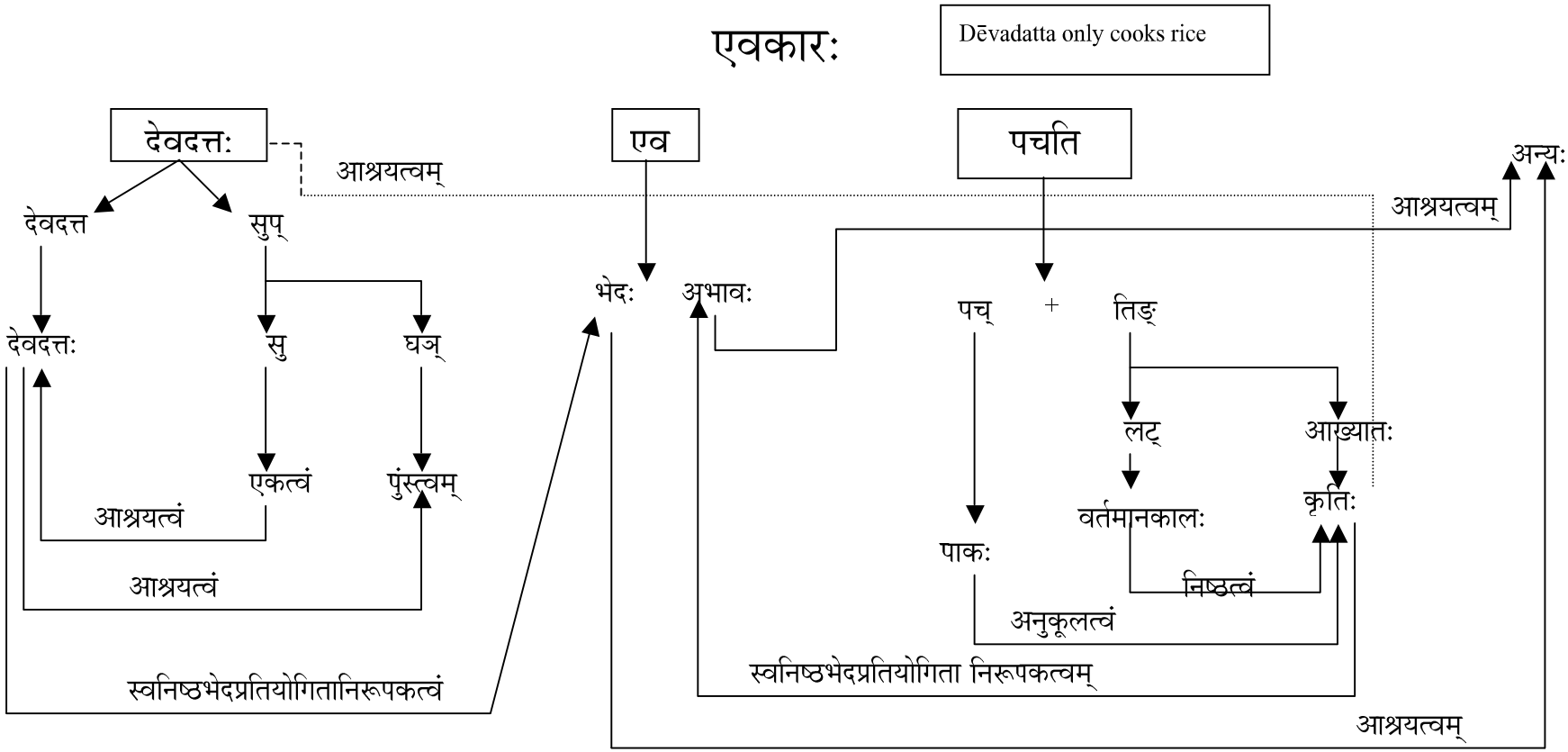


शाब्दबोधः - गुरुनिष्ठकर्मतानिरूपक नमननिरूपितोत्तरकालवृत्तित्वाश्रयलेखनानुकूलवर्तमानकालवृत्तिकृत्याश्रयैकत्वाश्रयः पुंस्त्वाश्रयः (उदाहरणार्थम् - एष शब्दः लिङ्गत्रयेऽपि समानः) अहम्

नियमाः - १. केवलधात्वर्थस्य अवाद्युपसर्गपूर्वक-धात्वर्थस्य च क्रमेण क्त्वा-ल्यबर्थयोः सार्धं (सर्वत्र) निरूपितत्वं सम्बन्धः । २. क्त्वा-ल्यपोः सर्वत्र (स्व) उत्तरकालवृत्तित्वम् अर्थः । ३. वाक्येषु सर्वत्र प्रथमान्तपदाभावे सति तदनुरोधेन अध्याहार्यम् । ४. क्त्वा-ल्यबर्थस्य उत्तरकालवृत्तित्वस्य आख्यातार्थकृतौ आश्रयत्वं सम्बन्धः ।

Naiyāyika-Śābda-bodhaḥ - Prathama-staraḥ- Kartṛ-Upasṛṣṭa-dhātu-yutaḥ - Kramika-kriyā-dvayam (Pūrvakāla-kriyārthakaḥ)

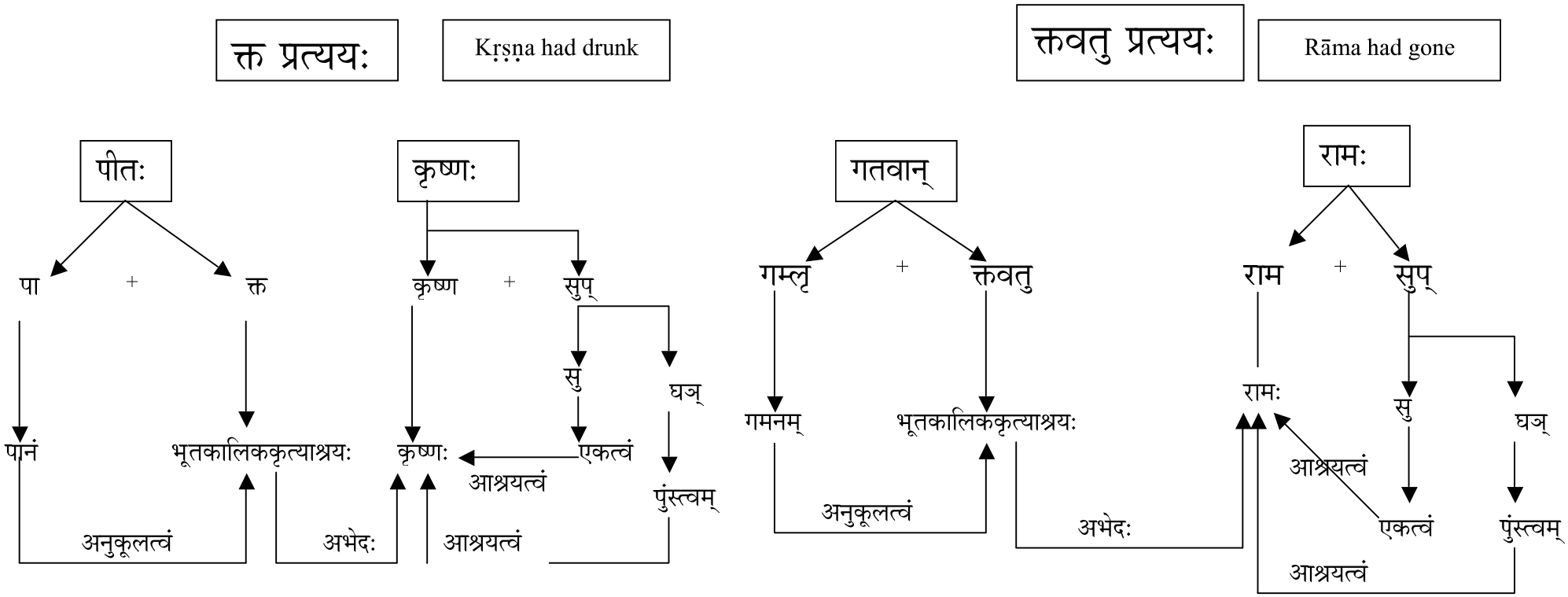
Figure 9.51



शाब्दबोधः - पाकानुकूल-कृत्याश्रयः एकत्वाश्रयः पुंस्त्वाश्रयः देवदत्तः or देवदत्तनिष्ठ-प्रतियोगितानिरूपक-भेदाश्रयः पाकानुकूलकृतिनिष्ठ-प्रतियोगितानिरूपकाभावाश्रयः अन्यः (अपेक्षाबुध्या ग्राह्यः) ।
नियमाः - १. एवकारस्य सर्वत्र भेदः, अभावः इति अर्थद्वयम् । २. एवकारघटिते च वाक्ये शाब्दबोधावसरे वाक्यद्वयं प्रतिभासते । प्रकृतवाक्यस्य प्रथमान्तविशेष्यकः एकः, अन्यपदार्थविषयकः अपरः ।
३. प्रकृतवाक्यघटकप्रथमान्तपदस्य एवकारार्थैकदेशभेदेन सार्धं स्वनिष्ठभेदप्रतियोगिताकत्वं सम्बन्धः ।
४. एवकारार्थयोः भेदाभावयोः आश्रयतासम्बन्धेनैव अन्यपदार्थे अन्वयः । (भेदाभावयोः अन्यपदार्थस्य च आश्रयत्वं सम्बन्धः)

Naiyāyika-Śābda-bodhaḥ - Prathama-staraḥ- Kartṛ-Kriyāpada-yutaḥ - Avadhāraṇārthakaḥ)

Figure 9.52



शाब्दबोधः पानानुकूल भूतकालिक कृतिमदभिन्नैकत्वाश्रयपुंस्त्वाश्रयः कृष्णः

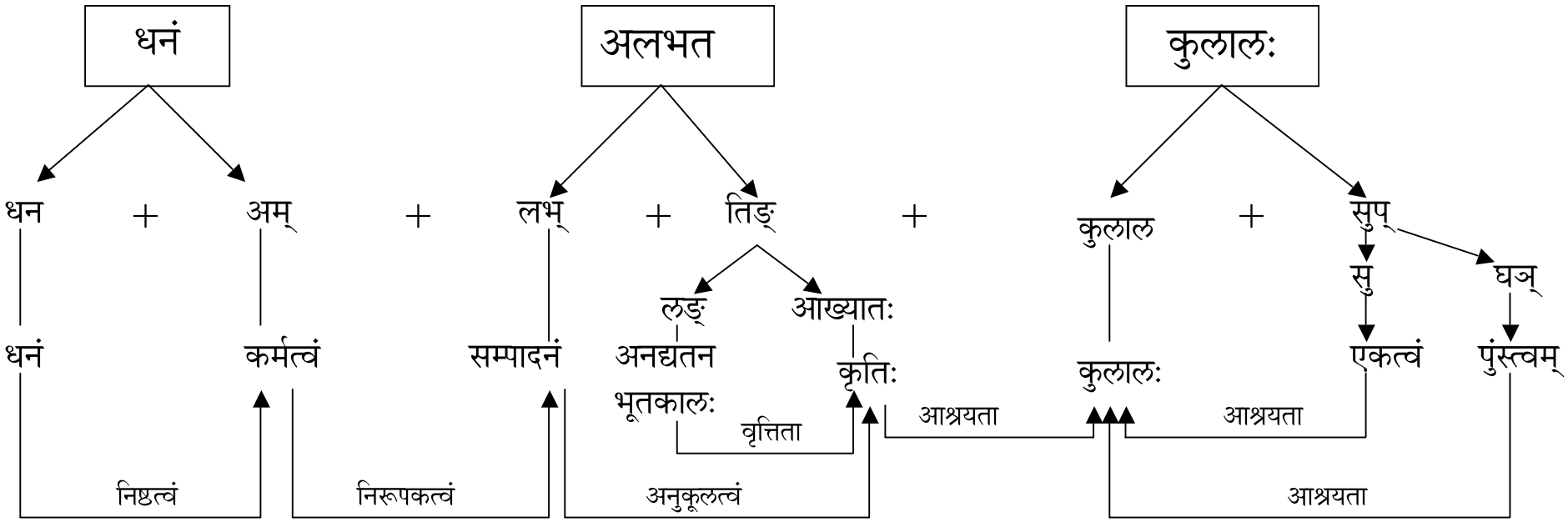
शाब्दबोधः - गमनानुकूलभूतकालिककृत्याश्रयाभिन्नैकत्वाश्रयः पुंस्त्वाश्रयः रामः

नियमाः - १. क्तक्तवत्वोः सर्वत्र भूतकालिककृत्याश्रयः अर्थः । २.क्तक्तवत्वर्थयोः प्रथमान्तार्थेन सार्धं अभेदः सम्बन्धः (अभेदान्वयं प्रति समानविभक्तिकत्वस्य तन्त्रत्वात्)

Naiyāyika-Śābda-bodhaḥ - Prathama-staraḥ- Kartṛ-Kṛdanta-yutaḥ - Bhūtakālikārthakaḥ)

Figure 9.53

(The) Potter got money (imperfect tense)



शाब्दबोधः - धननिष्ठ-कर्मता-निरूपक-सम्पादनानुकूल-अनद्यतन-भूतकाल-वृत्ति-कृत्याश्रयैकत्वाश्रय-पुंस्त्वाश्रयः कुलालः

Naiyāyika-Śābda-bodhaḥ - Dvitīya-staraḥ - Kartṛ-Karma-kriyā-Pada-yutaḥ - Karma-kārakam (Bhūta-kālikaḥ)

Figure 9.54

PART V

Chapter 10 : Remarks

10.1 Conclusion

An attempt has been made to survey the contents of Andian Indian technical treatises on Language comprehension and adapt them for NLP research work. Necessary extensions to Information Interchange Standards have been evolved and Computer interfaces through the Indian Scripts (covering Vedic texts also) are devised.

A preliminary algorithm has been designed for Morphological Analysis for Classical and Vedic Sanskrit. Necessary databases are also built (and a trial implementation has been achieved). The Vedic extension rulebases are testable against the Vedic databases built. A Parser with syntactic and semantic processing provisions and ontological knowledgebase is designed.

Many explanations and concepts derived from Sanskrit have not been effectively translated due to lack of expertise and in some cases, Sanskrit terms have been retained. Original forms of these, useful for those proficient in Sanskrit, are accessible from the author.

10.2 Future Scope

Extension to speech research of the Vedic module is obvious. Other Indian languages may also be tested with this approach. A few language-independent features like Kāraka - Vibhakti Mappings, Activity characterisations are applicable even in Machine Translation tasks, though with a lot more sophistication.

As set out in the Motivation section, the parser design could be tried out in various analytical spheres like text processing, poetry, Sūtra-s, conversations etc. Web-based language accessing and interaction can also be a desirable outcome. Development of a universal Knowledge Representation formalism and speech processing system for Indian languages could also benefit.

The algorithm can be extendable to multiple split as well, wherein split the word into all the ${}_nP_r$ possibilities and look for constituents of compound words or sentence units (words), additional heuristics are needed to reduce the explosive combinations (as in the case of compound words, unpunctuated sentences, literary/figurative usages). Dealing with verses/poetry, Sūtras (terse aphorisms) also require such control and larger rulebase.

10.3 References

1. Veda Vāridhi P. Rāmānujan, "Śikṣā Śāstra and Experimental Phonetics", National Seminar on the topic at NIAS, Bangalore, by ASR, May '96.

2. Prof. Subhash C. Kak [1987], ''The Pāṇinian approach to Natural Language Processing'', Int'l. J. of Approximate Reasoning 1987, 1: 117-130.

3. Rick Briggs [1984], "Knowledge representation in Sanskrit and Artificial Intelligence", AI Magazine, Spring 1985, pp. 32-39.

4. Rick Briggs [1985], " Śāstric Sanskrit as Machine Translation Interlingua".

5. K.L. Kanthan [1989], ''Formal Language System of Pāṇini'', Pre-Conference Papers of National Conference on 'Samskritam and Computers', New Delhi, 1989.

6. Dr. Vineet Chaitanya and Dr. K. V. Ramakrishnamacharyulu [1987], "SARALA SAMSKRITAM as a knowledge representation language".

7. Dr. Rajeev Sangal and Dr. Vineet Chaitanya [1986], "An intermediate language for Machine Translation : An approach based on Sanskrit using Conceptual Graph notation", Computer Science and Informatics, Jour. of Comp. Soc. of India, 17, 1, pp. 9 -21, 1987.

8. Akshar Bharati, Vineet chaitanya and Rajeev Sangal [1989], "A Kāraka based approach to Parsing of Indian languages", TRCS-89-88, I.I.T, Kanpur.

9. Akshar Bharati, Vineet chaitanya and Rajeev Sangal [1995], "Natural language Processing : A Paninian Perspective", Prentice-Hall india Ltd.

10. Akshar Bharati, Vineet chaitanya and Rajeev Sangal [1993c], "Anusaraka or Language Accessor: A Short Introduction", TRCS-93-205, Dept. of CSE, I.I.T, Kanpur.

11. Dr E. Schredl [1990], "Sanskrit Computational", ['SIMURG' project], Germany.

12. J. Ahlfors [1990], "Two level morphophonology of Sanskrit", Helsinki, Finland.

13. A.Verboom [1990], "Morphological analysis of Sanskrit by Computer", Netherlands.

14. George Cardona, "Pāṇinian Database as information Archives for Sanskrit grammar", M.I.T, USA.

15. Anand V Hudli, "A KRL based on Navya Nyāya", Ph.D. Thesis, 1990, Purdue Univ, Nebraska, USA.

16. S.M. Katre [1989], "Aṣṭādhyāyī of Pāṇini", Motilal Banarsidoss.

17. S.N. Srihari, W.J. Rapaport and D. Kumar [1987], ''On Knowledge Representation using Semantic Networks and Sanskrit'', Tech. Rep. 87-03 of SUNY, Buffalo, USA.

18. P. Rāmānujan [1990], "A scheme for knowledge representation in Samskritam", paper for Abhinava Vidya Bharati, USA awards, 1991.

19. B. Sabarinath and Saji K. David [1991], "Natural Language Understanding System for Sanskrit", M.Sc. (Comp. Sc.) Thesis, Univ. of Kerala, 1991.

20. P. Rāmānujan [1992], "Computerisation of Vedic texts", Proc. of National Conference on Vedas and Śāstra-s, Tirupati.

21. P. Rāmānujan [1992], "मस्तिष्क यन्त्रे शाब्द-बोध प्रतिरूपणम्", Proc. of National Conference on Vedas and Śāstra-s, Tirupati.

22. M.D. Pandit, A Concordance of Vedic Compounds Interpreted by Veda, Vol. 1, CASS, Pune University Press, (1989).

23. Vineet Jain [1993], "Knowledge Representation and Natural Language Processing", M.Tech. (Comp. Sc. & Engg) Thesis, I.I.T, Bombay, 1993.

24. Shanmukhappa A. Angadi [1994], "Design of a Morphological Analyzer for Sanskrit", M.Tech. (Comp. Engg.) Thesis, Univ. of Mysore, 1994.

25. Pushpak Bhattacharya, "Expert System for understanding Sanskrit - A first step", KRIS Conference, Bangalore, 1996.

Paradigm analysis of Nominal stems in Sanskrit

The methodology of Pāṇinian grammar may be illustrated by its generalisation of the nominal declensional and verbal inflexional affixes. By Aṣṭādhyāyī 4.1.2, it gives the following archetypal affixes for the nominal declension :

	Singular	Dual	Plural
Nominative	सु	औ	जस्
{Vocative	सु	औ	जस्}
Accusative	अम्	औट्	शस्
Instrumental	टा	भ्याम्	भिस्
Dative	ङे	भ्याम्	भ्यस्
Ablative	ङसि	भ्याम्	भ्यस्
Genetive	ङस्	ओस्	आम्
Locative	ङि	ओस्	सुप्

These minimise the set of rules for deriving the allomorphs from the declensional affixes with the least effort. The table below shows the basic forms with their allomorphs :

1. Nom. Sing.	स् (उ)
	0 after consonantal stems, feminine stems ending in long [ई, ऊ] (6.1.68)
	[अम्] after neuter stems ending in [अ] (7.1.24) and first and second
	personal pronouns [अद्ड्] for डतरादि (7.1.25)
2. Voc. Sing.	स् (उ) → 0 after stems in short vowels and monophthongs [ए, ओ] (6.1.69)
3. Nom-Voc. Acc. Dual	[औ (ट्)]
	[(श्) ई] after feminine stems in [आ] (7.1.18) and all neuter stems (7.1.19)
	[अम्] after first and second personal pronouns (7.1.28)
4. Nom. Voc. Plural	[(ज्) अस्]
	[(श्) ई] after pronomial stems (7.1.17)
	[(श्) इ] after neuter stems (7.1.20)
	[औ (श्)] after [अष्टन्-] (7.1.21)
	[0] after numerals designated by the technical term [षट् 1.1.24] (7.1.22)
	[अम्] after 1st and 2nd pers. pronouns (7.1.28)
5. Acc. Sing.	[अम्]

	0 after neuter stems (7.1.23)
	[अम्] after neuter stems ending in [अ] (7.1.24)
6. Acc. Plur.	[(श्) अस्]
	long vowel + न् in masc. stems (6.1.103)
	[(श्) इ] after neuter stems (7.1.20)
	[औ (श्)] after [अष्टन्] (7.1.21)
	[अम्] after 1st and 2nd pers. prons. (7.1.28)
7. Instr. sing.	[(ट्) आ]
	[ना] after masc. stems defined by घि (1.4.7:ending in short इ, उ) (7.3.120)
	[इन] after masc. and neut. stems ending in short [अ] (7.1.12)
8. Inst. dat. - abl. dual	[भ्याम्]
9. Inst. plur.	[भिस्]
	[ऐस्] after stems ending in short [अ], masc. as well as neut. (7.1.9)
10. Dat. sing.	[(ङ्) ए]
	[य] after stems ending in short [अ], masc. as well as neut. (7.1.13)
	[स्मै] after masc. pronomial stems (7.1.14)
	[स्या (ट्) + (ङ्) ए] after fem. pron. stems (7.3.114)
	[अम्] after 1st and 2nd pers. pronomial stems (7.1.28)
	[आ (ट्) + (ङ्) ए] after fem. stems in both short and long [ई, ऊ] (7.3.112)
	[या (ट्) + (ङ्) ए] after fem. stems ending in long [आ] (7.3.113)
11. Dat. plur.	[भ्यस्]
	[अभ्यम्] after 1st and 2nd pers. pron. (7.1.30)
12. Abl. sing.	[(ङ्) अस् (इ)]
	[आत्] after stems in short [अ] (7.1.12)
	[अत्] after 1st and 2nd pers. pron. stems (7.1.32)
	[आ(ट्)+ (ङ्)अस्(इ)] after fem. stems in short and long [ई, ऊ: नदी 1.4.3ff.] (7.3.112)
	[या (ट्) + (ङ्) अस् (इ)] after fem. stems in [आ] (7.3.113)
13. Abl. plur.	[भ्यस्]
	[अत्] after 1st and 2nd pers. pronominal stems (7.1.31)

14. Gen. sing.	[(ङ्) अस्]
	[स्य] after stems in [अ] (7.1.12)
	[स्या (ट्) + (ङ्) अस्] after pronomial stems (7.3.114)
	[आ (ट्) + (ङ्) अस्] after fem. stems in short or long [इ, उ] (7.3.112)
	[या (ट्) + (ङ्) अस्] after fem. stems in [आ] (7.3.113)
15. Gen. Loc. dual.	[ओस्]
16. Gen. plur.	[आम्]
	[न् (उट्) + आम्] after stems ending in short vowels and fem. stems in [इ, उ, long and short, and आ] (7.1.54)
	[स् (उट्) + आम्] after pron. stems (7.1.52)
	[आकम्] after 1st and 2nd pers. pronomial stems (7.1.33)
17. Loc. sing.	[(ङ्) इ]
	[आम्] after fem. stems in [आ, ई, ऊ] (7.3.116) and after fem. stems in short [इ, उ] (7.3.117)
	[औ (त्)] after stems denoted by the technical term घि (1.4.7) (7.3.118)
	[आ (ट्) + आम्] after stems denoted by the संज्ञा [नदी 1.4.3ff] (7.3.112)
	[या (ट्) + आम्] after fem. stems in long [आ] (7.3.113)
18. Loc. plur.	[सु (प्)]

[Ref. Introduction in *Aṣṭādhyāyī of Pāṇini*, Roman Transliteration and English Translation by S M KATRE].

This has been applied to all the different bases obtaining in Amarakośa, Pāṇini's Liṅgānuśāsana etc. and similar works like Śabda-ratnāvalī, Prof. Borvankar's and Prof. Junnarkar's categories etc. as also with a view to apply for morphological analysis and 'parsing' including Vedic texts, (even treating cerebralisation as a distinct type etc.) the following list has been arrived at :

सुबन्त-शब्द-प्रभेदाः - कतिपय-आदर्श-शब्द-सहिताः

पुल्लिङ्गः - *अ कारान्तः*

1. राम - अकूपार, अक्ष, अक्षर, अग्निहोत्र, अङ्कुर
2. देव - अकाल, अक्षत, अग, अगोप, अघ
3. सर्व - अन्यतर, इतर, कतर, ततर, यतर
4. विश्व - अन्य, कतम, ततम, त्व, यतम
5. पूर्व - अधर, अन्तर, अपर, अवर, पर
6. दक्षिण - स्व
7. प्रथम - अर्ध, अल्प, कतिपय, द्वितय
8. चरम - पूर्वापर, वर्णाश्रमेतर
9. नेम
10. द्वितीय - तृतीय
11. उभ - उभक
12. उभय
13. एक

14. निर्जर

15. पाद

16. दन्त

17. मास

18. यूष

19. द्यह्न - व्यह्न, सायाह्न

आ कारान्तः

20. विश्वपा - गोपा, धूम्रपा, शङ्खध्मा, सोमपा

21. हाहा

इ कारान्तः

22. हरि - अद्रि, इन्द्रारि, ऊर्मि, ऋषि, कुक्षि

23. श्रीपति - अग्नि, अञ्जलि, अतिथि, अतिद्वि, अतिसखि

24. पति

25. सखि

26. औडुलोमि

27. अतिस्त्रि

28. प्रियत्रि

29. सुसखि - अतिसखि, परमसखि

30. द्वि

31. त्रि - परमत्रि

32. क्षि

ई कारान्तः

33. बहुश्रेयसी

34. अतिलक्ष्मी

35. कुमारी

36. नी - अग्रणी, उन्नी, ग्रामणी, सेनानी

37. सखी

38. वातप्रमी - पपी, ययी

39. वातप्रमी - जलपी, प्रधी, वेगी

40. सुधी - पक्की, परमधी, यवक्री, शुद्धधी, शुष्की

41. सुखी - क्षामी, प्रस्तीमी, लूनी, सखी, सुती

उ कारान्तः

42. गुरु - इषु, ऋभु, कुरु, कुसुमेषु, दिदृक्षु

43. वायु - अशु, अंशु, ऋतु, ओतु, कण्डु

44. क्रोष्टु

45. सानु

46. क्ष्णु - दु, द्रु, प्लु, यु, रु, श्रु, स्तु, स्नु, हु

ऊ कारान्तः

47. हूहू

48. अतिचमू

49. स्वभू - आत्मभू, कटप्रू, काराभू, दृग्भू, प्रतिभू

50. वर्षाभू - उल्लू, खलपू, दृन्भू, दृम्भू, पुनर्भू

ऋ कारान्तः

51. धातृ - आराधयितृ, उद्गातृ, कर्तृ, क्षत्तृ, त्वष्टृ

52. नृ

53. पितृ - जामातृ, देवृ, भ्रातृ, सव्येष्टृ

ॠ कारान्तः

54. कॄ - तॄ

ऌ कारान्तः

55. शक्ल - गम्ल

ए कारान्तः

56. से - स्मृते

च कारान्तः

57. जलमुच् - पयोमुच्, सुवाच्

58. सुवृश्च्

59. प्राञ्च् - अदद्र्यञ्च्, अदमुयञ्च्, अन्वञ्च्, अमुमुयञ्च्,

60. प्राच् - अर्वाच्, अवाच्, पराच्

61. प्रत्यच् - अदद्र्यच्, अदमुयच्, अमुमुयच्, देवद्र्यच्

62. अन्वच् - विष्वच्

63. उदच्

64. तिर्यच्

65. अच्

ज कारान्तः

66. भिषज् - ऊर्ज्, ऋत्विज्, क्रतुभुज्, तनुत्यज्, सुखभाज्

67. युञ्ज्

68. विभ्राज् - सुयुज्

69. विश्वराज्

70. भृज्ज्

71. खञ्ज्

72. राज् - देवेज्, परिमृज्, परिव्राज्, विभ्राज्, विश्वसृज्

ण कारान्तः

73. सुगण् - सुगाण्

त कारान्तः

74. ददत् - अम्बुभृत्, इन्द्रजित्, क्ष्माभृत्, गरुत्, चकासत्,

75. महत्

76. भवत् - अदत्, अधीयत्, उदयत्, कुर्वत्, चिन्तयत्

77. धीमत् - अग्निमत्, इच्छत्, उदन्वत्, एतावत्, गोमत्

थ कारान्तः

78. अग्निमथ्

द कारान्तः

79. दिविषद् - अतित्यद्, क्रव्याद्, राजनुद्, शास्त्रविद्, सुहृद्

80. त्यद्

81. यद्

82. एतद्

83. अत्यस्मद् - (त्वां मां वा अतिक्रान्तः इति विग्रहे)

84. अतियुष्मद् - (त्वां मां वा अतिक्रान्तः इति विग्रहे)

85. अत्यस्मद् - (युवाम् आवां वा अतिक्रान्तः इति विग्रहे)

86. अतियुष्मद् - (युवाम् आवां वा अतिक्रान्तः इति विग्रहे)

87. अत्यस्मद् - (युष्मान् अस्मान् वा अतिक्रान्तः इति विग्रहे)

88. अतियुष्मद् - (युष्मान् अस्मान् वा अतिक्रान्तः इति विग्रहे)

89. सुपाद् - त्रिपाद्, द्विपाद्

90. तद्

ध कारान्तः

91. बुध्

न कारान्तः

92. सुत्रामन् - क्रावन्, गरिमन्, ग्रावन्, तक्षन्, वृषन्

93. मूर्धन् - अणिमन्, महिमन्, सहस्रधामन्, सुदामन्

94. आत्मन् - पाप्मन्, यज्वन्

95. ब्रह्मन् - अनर्वन्, विश्वकर्मन्, सुतर्मन्, सुपर्वन्

96. भ्रूणहन् - अमीवहन्, गोहन्

97. वृत्रहन् - पूरुषहन्, ब्रह्महन्, रक्षोहन्

98. गुणिन् - अधीतिन्, कृतिन्, गामिन्, गृहीतिन्, तपस्विन्

99. करिन् - अनुसारिन्, छत्रिन्, दण्डधारिन्, दोषिन्, पक्षिन्

100. पूषन् - अर्यमन्

101. पथिन् - मथिन्

102. राजन् - प्रियपञ्चन्

103. प्रतिदिवन्

104. मघवन्

105. श्वन्

106. अर्वन्

107. युवन्

108. ऋभुक्षिन्

109. दीर्घाहन्

110. मघवन्

111. प्रियाष्टन् - परमाष्टन्

प कारान्तः

112. गुप्

म कारान्तः

113. किम्

114. इदम्

115. इदकम्

116. प्रशाम्

117. प्रियचतुर्

रेफान्तः

118. चतुर् - परमचतुर्, प्रियचतुर्

श कारान्तः

119. विश् - भ्रश्, मृश्

120. नश्

121. पुरोडाश्

122. तादृश् - घृतस्पृश्, यादृश्, स्पृश्

ष कारान्तः

123. रत्नमुष् - देहपुष्, द्विष्, नीलत्विष्, सुरद्विष्

124. दोष्

125. दधृष्

126. पिपठिष्

127. प्रियषष्

128. चिकीर्ष्

129. विविक्ष् - तक्ष्

130. गोरक्ष् - तक्ष्, दिधक्ष्, पिपक्ष्, विवक्ष्

स कारान्तः

131. वेधस् - अमृतान्धस्, चन्द्रमस्, दिवौकस्, बिडौजस्, सुमनस्

132. उशनस्

133. पुम्स्

134. अदकस्

135. सुपिस्

136. सुतुस्

137. विद्वस् - दाश्वस्, मीढ्वस्, साह्वस्

138. तस्थिवस् - उपेयिवस्, ऊचिवस्, सेदिवस्

139. अनेहस् - पुरुदंसस्

140. श्रेयस् - कनीयस्, गरीयस्, ज्यायस्, प्रेयस्

141. सुहिम्स् - जिघाम्स्

142. स्रस् - ध्वस्

143. सुवस् - पिण्डग्रस्, पिण्डग्लस्

144. उदर्चिस् - गृहीतधनुस्, ज्योतिस्, त्रिभुवनवपुस्, दीर्घायुस्

145. अदस्

ह कारान्तः

146. लिह् - दामलिह्, महीरुह्

147. दुह्

148. द्रुह्

149. अनडुह्

150. तुरासाह्

151. भूवाह्

152. मुह् - नह्, स्निह्, स्नुह्

153. विश्ववाह् - भारवाह्, श्वेतवाह्, हव्यवाह्

154. श्वेतवह्

ल कारान्तः

155. कमल्

ङ कारान्तः

156. लिङ् - लङ्, लुङ्, लृङ्

स्त्रीलिङ्गः - *आ कारान्तः*

157. रमा - अधीता, अनादरा, उदारा, क्षणप्रभा, गोत्रा

158. अम्बिका - अङ्गना, अचला, अजा, अतुला, अनन्ता

159. सर्वा - अधरा, अन्तरा, अन्यतरा, अन्या, अपरा

160. पूर्वोत्तरा - उत्तरपूर्वा, पूर्वोत्तरा

161. द्वितीया - तृतीया

162. अम्बा - अक्का, अल्ला

163. जरा

164. नासिका

165. निशा

166. पृतना

167. एका

168. उभया

169. उभा

इ कारान्तः

170. मति - अङ्गुलि, अद्मनि, अमनि, अरणि, अवनि

171. ऊर्मि - अनुपलब्धि, अश्रि, कृषि, खारि, दर्वि

172. प्रियत्रि

173. अतिस्त्रि

174. त्रि

ई कारान्तः

175. गौरी - अनृभुक्षी, उर्वी, कुरुचरी, कौबेरी, क्रोष्ट्री

176. सखी - अटवी, आग्नेयी, आग्रहायणी, ऐन्द्री, ऐरावती

177. स्त्री

178. अवी - प्रधी

179. लक्ष्मी - तन्त्री, तरी, महालक्ष्मी, स्तरी

180. श्री - ह्री

181. प्रधी - धी, भी

उ कारान्तः

182. धेनु - करेणु, कर्कन्धु, किष्कु, कु, कुहु

183. सरयु - इषु

ऊ कारान्तः

184. वधू - अलाबू, आडू, कच्छू, कण्डु, कासू

185. रम्भोरू - आरू, कर्षू, कशेरू, कुरू, खर्जू

186. वर्षाभू - पुनर्भू, मू

187. भू

188. भ्रू - सुभ्रू

ऋ कारान्तः

189. स्वसृ - ननान्द, यातृ

190. मातृ - दुहितृ

च कारान्तः

191. वाच् - ज्योच्, त्वच्, शुच्, स्फिच्, स्रुच्

ज कारान्तः

192. स्रज् - भिषज्, रुज्

त कारान्तः

193. हरित् - चत्वारिम्शत्, तडित्, त्रिम्शत्, पञ्चाशत्, पुत्

द कारान्तः

194. मुद् - आपद्, परिषद्, प्रतिपद्, विपद्, शरद्

195. तद्

196. त्यद्

197. एतद्

198. यद्

ध कारान्तः

199. क्षुध् - क्रुध्, युध्, वीरुध्, समिध्

न कारान्तः

200. दामन् - सीमन्

201. बहुराजन्

प कारान्तः

202. अप्

भ कारान्तः

203. ककुभ् - त्रिष्टुभ्

म कारान्तः

204. इदम्

205. किम्

206. इदकम्

रेफान्तः

207. गीर्

208. द्वार् - वार्

209. पूर् - जूर्, तूर्, धूर्, मूर्, हूर्

210. चतुर्

श कारान्तः

211. दिश् - दृश्, धृश्, विश्, सदृश्

212. विपाश्

ष कारान्तः

213. प्रावृष् - तृष्, त्विष्, द्विष्, रुष्, विप्रुष्

214. सजुष्

स कारान्तः

215. अदस्

216. आशिस्

217. भास्

218. सुमनस् - अप्सरस्, उषस्

219. छदिस् - अर्चिस्

ह कारान्तः

220. उपानह्

221. उष्णिह्

222. कामदुह्

नपुंसकलिङ्गः - *अ कारान्तः*

223. ज्ञान - अङ्क, अङ्कुश, अङ्ग, अङ्गण, अङ्गद

224. गृह - अक्ष, अग्र, अङ्गुलित्र, अजस्र, अजिर

225. कतम - अन्य, ततम, यतम

226. कतर - अन्यतर, इतर, ततर, यतर

227. सर्व - अधर, उत्तर, एकतर, पर, पूर्व

228. विश्व - दक्षिण, स्व

229. पूर्व - अधर, अन्तर, अपर, अवर, पर

230. दक्षिण - स्व

231. एक

232. उभय

233. अजर

234. हृदय

235. आस्य

236. उदक

237. मांस

238. उभ

इ कारान्तः

240. शुचि - उद्भासि, पादपाणि, शोभि

241. अक्षि

242. प्रधि

243. सुधि

244. अतिस्त्रि

245. प्रियत्रि

246. दधि

247. अस्थि - अतिदधि, दधि, प्रियदधि, प्रियास्थि, सक्थि

248. अतिरि - अनादि, प्ररि, सखि

249. त्रि

उ कारान्तः

250. मधु - अम्बु, कमण्डलु, जतु, तालु, त्रपु

251. गुरु - अश्रु, उरु, कशेरु, चारु, श्मश्रु

252. प्रियक्रोष्टु

253. सानु

ऋ कारान्तः

254. कर्तृ - ज्ञातृ, धातृ, पातृ, शास्तृ, हर्तृ

च कारान्तः

255. सुवाच्

256. गवाञ्च्

257. तिर्यञ्च् - अदद्र्यञ्च्, अदमुयञ्च्, अन्वञ्च्, अमुमुयञ्च्

258. प्राच् - अर्वाच्, अवाच्, पराच्

259 प्रत्यच् - अदद्र्यच्, अदमुयच्, अमुमुयच्, देवद्र्यच्

260. अन्वच् - विष्वच्

261. उदच्

262. तिर्यच्

ज कारान्तः

263. असृज्

264. ऊर्ज्

265. बहूर्ज्

266. खञ्ज्

त कारान्तः

267. धनवत् - अदत्, उदश्वित्, गुणवत्, जगत्, पृषत्

268. महत्

269. पचत् - गच्छत्, चिकीर्षत्, चोरयत्, दीव्यत्, भवत्

270. ददत् - जक्षत्, जाग्रत्, दधत्, शासत्

271. तुदत् - उल्लसत्, दास्यत्, नुदत्, भात्, यात्

272. यकृत् - शकृत्

द कारान्तः

273. तद् - त्यद्, यद्

274. बेभिद् - चेच्छिद्

275. एतद्

न कारान्तः

276. ब्रह्मन् - कर्मन्, चर्मन्, नर्मन्, भर्मन्, वर्मन्

277. जन्मन् - धन्वन्, वर्त्मन्

278. प्रेमन् - बहुपूषन्, बह्वर्यमन्

279. सुराजन्

280. सुपथिन्

281. बहुवृत्रहन् - बहुवृत्रहन

282. अहन् - दीर्घाहन्

283. नामन् - दामन्, धामन्, भस्मन्, रोमन्, व्योमन्

284. दण्डिन् - गुणिन्, वाग्मिन्

285. स्रग्विन्

प कारान्तः

286. स्वप्

म कारान्तः

287. किम्

288. इदम्

289. इदकम्

रेफान्तः

290. चतुर्

291. वार्

व कारान्तः

292. विमलदिव् - सुदिव्

श कारान्तः

293. तादृश् - एतादृश्, यादृश्

ष कारान्तः

295. पिपठिष्

स कारान्तः

298. अदस्

300. सुपुम्स्

302. स्रस् - ध्वस्

304. धनुस् - अरुस्, आयुस्, चक्षुस्, जनुस्, तनुस्

305. स्वनडुह्

307. प्रधी

308. रै

309. गो - द्यो, सुद्यो, स्मृतो

310. ग्लौ - नौ, सुनौ

311. अवयाज्

312. सुदिव् - दिव्

313. द्वि

314. कति

315. युष्मद्

न कारान्तः

318. अष्टन् - परमाष्टन्

319. षष् - परमषष्

320. उक्थशास्

294. रत्नमुष् - नीलत्विष्

296. दोष्

297. श्रेयस् - अणीयस्, अन्धस्, अर्णस्, आगस्, उरस्

399. विद्वस्

301. सुहिम्स् - जिघाम्स्

303. हविस् - अर्चिस्, उदर्चिस्, छर्दिस्, ज्योतिस्

ह कारान्तः

306. लिह्

पुं/स्त्रीलिङ्गकः - *ई कारान्तः*

ऐ कारान्तः

ओ कारान्तः

औ कारान्तः

ज कारान्तः

व कारान्तः

स्त्री/नपुंसकलिङ्गकः - *इ कारान्तः*

त्रिलिङ्गकः - *इ कारान्तः*

द कारान्तः

316. अस्मद्

317. पञ्चन् - चतुर्दशन्, दशन्, द्वादशन्, नवन्, परमपञ्चन्

ष कारान्तः

स कारान्तः

Note : *Legend* : 1st Col. = #(cum.) Sl. No. of word-type; 2nd Col. = Illustrative Head-word of the type; 3rd Col. = a few similar words of the particular type; Typically, when single entries are found in word-types, the head-words may be unique words of that type.

Further details regarding the classification with applicable rules etc. can be had from the author or at http://www.soft.net/cdac/html/solutions/ih.html.

This classification is mainly for the purposes of exhaustive morphological analysis using the Computer. Multiple word identifications are a major gain of this process. Accent-related sub-types are covered in the algorithm using Sup-Svara rules. Yāska's Nighaṇṭu may be adopted for lexicon in accented word analysis.

This list is about 95 % complete. Word-types with Tadrāja suffixes (Añ, Aṇ, Ñyaṅ, Ṇya, lñ, Ñya, Ñyaṭ, Ṭeṇyaṇ, Cha, Yañ) etc. which, while denoting plurality (in non-feminine stems) get these suffixes elided. (2.4.62ff.) Certain special (*e.g,* pronomial) compound formations, words formed with suffixes like *Akac* etc. are to be accounted for. Vedic types also require handling as per details found in Bhaṭṭa-bhāskara-bhāṣya etc.

Vedic Specialities
Accent-related meaning Changes - More examples/discussion

We could cite many examples of the same words being nouns and verbs in isolation, derived through different root-suffix combinations, determined on the basis of accents (e.g, तस्मान्नानादेवत्यानि सन्ति वायव्यान्युच्यन्ते (तै. सं. ६.४.६.७); [*nānadēvatyāni*] *santi* - is a noun here;

अङ्गिरसः सुवर्गं लोकं यतः पुरोडाशः कूर्मो भूत्वानुप्रासर्पत् (तै. सं. ५.२.८.४); *yataḥ* [*purōḍāśaḥ*] - *yataḥ* is a noun here [mas. acc. plural in śatṛ ending], only *yataḥ* derived from *yad* pronoun, as its suffix *tasil* is *lit* - यतः - changes verb accents;

yathā - तं नेमिमूभवो यथाऽऽनमस्व सहूतिभिः (तै. सं. २.६.११.१); आ सवꣳ सवितुर्यथा भगस्येव भुजिꣳ हुवे (तै. सं. ३.१.११.८); when meaning similarity or analogy, *yathā* is unaccented - यथा - [as in these two cases] and does not accent verbs and only *yathā* derived from *yad* pronoun, as its suffix *thāl* is *lit* - यथा - changes verb accents).

Pāṇini has enumerated various circumstances for the change of default accents for nouns, verbs, particles, compounds, affixes etc. in Aṣṭādhyāyi. Few examples are :

1. The word ***viśvāhā*** occurring at the beginning of the foot of a verse with four different accent patterns [with the first shown form meaning *'one who destroys all* (hostile)', while the other three forms having the same meaning as *'always, at all times'* or lit. *'on all days'*] -

i). इन्द्रो नस्तत्र वृत्रहा **विश्वाहा** शर्म यच्छतु (तै. सं. ४.६.४.५); पदपाठे - विश्वाहेति विश्व - हा ।

इङ्क्यं समस्तं पदम् ।

ii). ते गृहासो मधुश्चुतो **विश्वाहा**ऽस्मै शरणास्सन्त्वत्रꣳ; and तिलवथ्सा ऊर्जमस्मै दुहाना **विश्वाहा** सन्त्वनपस्फुरन्तीः (तै. आ. ४.७.१) । पदपाठे - (?) पदद्वयं यद्वा आरण्यकशिक्षया स्वरस्य साधुत्वम् ?

iii). तत्रा रथमुप शग्मꣳ सदेम **विश्वाहा** वयꣳ सुमनस्यमानाः (तै. सं. ४.६.६.३) । पदपाठे - विश्वाहेति, विश्वा - अहा, - समस्तं इङ्क्यं पदम् ।

iv). सिञ्चामहा अवटमुद्रिणं वयं **विश्वाहा**ऽदस्तमक्षितम् (तै. सं. ४.२.५.५) । पदपाठे - विश्वा । अहा । इति व्यस्तं अनिङ्क्यं पदद्वयम् ।,

2. the word ***indrāgnī*** (three accent patterns),

i). तास्सृष्टा **इन्द्राग्नी** अपागूहताम् (तै. सं. २.२.१.१); पदपाठे - इन्द्राग्नी इतीन्द्र - अग्नी ।

ii). उभा वामिन्द्राग्नी आहुवध्यै (तै. सं. १.१.१४.१); पदपाठे - इन्द्राग्नी इतीन्द्र - अग्नी ।

iii). **इन्द्राग्नी** आ गतम् (तै. सं. १.४.१५.१); - पदपाठे - इन्द्राग्नी इतीन्द्र - अग्नी ।

3. the word ***dyāvāpṛthivī*** (three accent patterns),

i). इदं **द्यावापृथिवी** भद्रमभूत् (तै. सं. २.६.९.५); - पदपाठे - द्यावापृथिवी इति द्यावा - पृथिवी ।

ii). अन्तस्तें दधामि **द्यावापृथिवी** अन्तरुर्वन्तरिक्षम् (तै. सं. १.४.३.१); पदपाठे - द्यावापृथिवी इति द्यावा - पृथिवी ।

iii). इन्द्राग्नी **द्यावापृथिवी** आप ओषधीः (तै. सं. १.२.१.२); - पदपाठे - द्यावापृथिवी इति द्यावा - पृथिवी

4. and ***vājayantaḥ*** (accent change causing it to be a single word or a compound word)

i). अश्याम वाजमभि **वाजयन्तो**ऽश्याम द्युम्नं (तै. सं. १.३.१४.३); पदपाठे - वाजयन्तः । व्यस्तं पदम् (क्यजन्तम् ?)

ii). त्वया वाजं **वाजयन्तो** जयेम (तै. सं. १.४.४६.३); **वाजयन्तो** हवामहे (तै. ब्रा. २.४.६.४); पदपाठे - वाजयन्त इति वाज - यन्तः । इङ्क्यम् समस्तं पदम् ।

Verbal accents as sentence delimiters

The author's study of about 30000 verbal forms occurring in Taittirīya Saṃhitā have revealed many interesting aspects of word-meaning relations thru' accents. Verbal accents are dealt with in great detail in Aṣṭādhyāyi and are quite significant in semantic extraction at sentence level. All words are

unaccented except one syllable (anudāttaṃ padamekavarjam - Aṣṭādhyāyī. 6.1.158) is the default rule. Then, verbs are all unaccented except at the beginning of a foot of a hymn (anudāttaṃ sarvamapādādau - Aṣṭādhyāyī. 8.1.18) or when in association with the pronomial forms of yad (yadvṛttānnityam - Aṣṭādhyāyī. 8.1.66) or with certain particles designated nipāta etc (nipātairyadyadihanta ... - Aṣṭādhyāyī. 8.1.30) or when followed by another verb (tiṅṅatiṅaḥ - Aṣṭādhyāyī. 8.1.28).

By default the verb is unaccented as in - परा॑ऽस्य॒ भ्रातृ॑व्यो भवति (तै. सं. ३.४.६.२) - पदपाठे - भ॒व॒ति॑.

A **few illustrations** of accented verbs are :

पादादौ - वेदा॒हमे॒तं पु॑रुषं म॒हान्ता॑म् (तै. आ. ३.७.७);

अ॒सि॒ य॒मो अस्यादि॒त्यः (तै. सं. ४.६.७.१);

तिङ्तिङः - ता वि॑ष्णो पाहि पा॒हि य॒ज्ञम् (तै. सं. १.१.११.२) (प्रथमं - पा॒हि॒ ; द्वितीयं - पा॒हि) ;

सर्वा॑ पा॒प्मान॑ं त॒रति॒ तर॑ति ब्रह्महत्यां (तै. सं. ५.३.१२.१-२);

स्वेन॑ भाग॒धये॒नोप॑ धावति॒ धाव॑ति वाज॑म् (तै. सं. २.२.४.६);

पिब॑न्तु॒ मद॑न्तु॒ विय॑न्तु॒ सोम॑म् (तै. ब्रा. २.६.११.१०);

तस्मा॑द्यज्ञोपवी॒त्ये॑वाधी॑यीत याजयेद्यजे॑त वा य॒ज्ञस्य॒ प्रसृ॑त्यै (तै. आ. २.१.१) - तिङन्तत्रयं ;

यस्मा॑द्द्वीषाऽवे॑पिष्ठाः प॒लायि॑ष्ठास्स॒मज्ञा॑स्थाः (तै. ब्रा. ३.७.८.२) - तिङन्तत्रयं एकमुपसृष्टं ;

यथ्स॒ञ्चर॑त्य॒पवि॑शत्य॒त्ति॒ष्ठ॒ते च (तै. आ. ६.६४.१); - उपसृष्टं तिङन्तत्रयं ;

आश्रा॑वयास्तु॒श्रौष॑ड्यजये॒यजा॑महे वषड्कु॒रः (तै. सं. १.६.११.२); - तिङन्तपञ्चकम् ? .

यद्वृत्तात् - अग्ने॒ यं य॒ज्ञम॑ध्व॒रं वि॒श्वतः॑ परि॒भूर॑सि॒ (तै. सं. ४.१.११.१);

यस्य॑ समु॒द्रꣳ र॒सया॑ स॒हाऽऽहुः (तै. सं. ४.१.८.४-५) । (पदकाले आ॒हुः) ;

अन्यत्र - ऋष॑य॒ एन॑साऽऽहुः (तै. सं. ३.२.८.१-२); - पदकाले आ॒हुः ;

प्राच्यां॑ दि॒शि त्वमि॑न्द्रासि॒ राजो॒तोदी॑च्यां वृत्रहन्वृत्र॒हाऽसि॑ (तै. सं. २.४.१४.१); ।

पदपाठे - वृ॒त्र॒हन्नि॑ति॑ वृत्र - ह॒न् । वृ॒त्र॒हेति॑ वृत्र - हा । अ॒सि॒ ।।

Here, as accented and unaccented forms both give rise to the same surface (combined) form, the study of accent change is necessary for proper vākyānta vēṣṭana during Krama-pāṭha, *i.e*, वृ॒त्र॒हा + असि = वृ॒त्र॒हाऽसि ; वृ॒त्र॒हा + अ॒सि॒ = वृ॒त्र॒हाऽसि ;

निपातैर्यद्यदि, चवायोगे .. - यदि॑ बिभी॒याद्ध्रु॒श्मा भविष्यामीति (पदपाठे - बि॒भी॒यात् - तै. सं. २.२.१०.३);

अव॑ चो॒क्षति॒ प्र च॑ किरति॒ शुद्ध्यै॑ (पदपाठे - उ॒क्षति॑ - तै. सं. ६.२.७.४);

हि प्राप्त्या - आपो॒ हि ष्ठा म॑यो॒भुव॑ः (तै. सं. ४.१.५.१); पदकाले - स्थ इति उदात्तं पदम् ।

दे॒वा ह्यै॑तमै॒न्ध॑त (तै. सं. २.५.९.१); पदकाले - ऐ॒न्ध॑त ।

इन्द॑वो वामु॒शन्ति॒ हि (तै. सं. १.४.४.१). पदकाले - उ॒शन्ति॑ । (यद्धितुपरं छन्दसि ८.१.५६)

In long prose form of Brāhmaṇa texts, sentence demarcation is easily done with the accent study. Kalpasūtras and Mīmāṃsā Śāstra also deal with these considerably.

Key board overlay for Vedic

A character on the Vedic overlay is selected by pressing it along with Extension key. The Extension key can vary in location according depending on the type of the keyboard and implementation. On an 84-key IBM PC-AT keyboard, the extension key cab be the same as the SYS-REQ key. On the 101-key IBM Enhanced Keyboard, the right ALT key can be used as an Extension key. The right right -ALT key, being much closer, is more convenient than the SYS-REQ key; this makes the keyboard more convenient for speedy typing of Vedic text.

The Svarita symbols go above, below or after the a character. Only one symbol can be attached at a time. The character can be a full Consonanat, Pure Consonant (with Halant), Vowel or another stand-alone Vedic symbol.

Extended ISCII Code for Vedic

The Extension character (EXT), defined in ISCII, can be followed by another character in the range of A1h to EEh, to yield addl. characters. Fig. is the list of Vedic diacritic markers in Sanskrit (Devanagari) font.

Code Hex. Dec	ISCII Char	Vedic Char	Name
A1 161	ँ	÷	Visarga 1
A2 162	ं	◌	Filler
A3 163	ः	◌	Yajurvedic Anusvāra 1
A4 164	अ	◌	Yajurvedic Anusvāra 2
A5 165	आ	◌	Śukla Yajurvedic Anusvāra
A6 166	इ	◌	Yajurvedic Anusvāra 3
A7 167	ई	◌	Yajurvedic Anusvāra 4
A8 168	उ	◌	Visarga 2
A9 169	ऊ	◌	Visarga 3
AA 170	ऋ	◌	Visarga 4
AB 171	ऌ	◌	Visarga 5
AC 172	ए	×	jihvāmūlīya
AD 173	ऐ	◌	Kṛṣṇa Yajurvedic Anusvāra
AE 174	ऍ	◌	Kṛṣṇa Yajurvedic Long Anusvāra
AF 175	ऑ	◌	Yajurvedic Anusvāra 5
B0 176	ओ	◌	Yajurvedic Anusvāra 6
B1 177	औ	◌	Yajurvedic Anusvāra 7
B2 178	ऑ	१	Short Kampa
B3 179	क	३	Long Kampa
	SWAR		
B4 180	ख		Jātya Svarita-Atharva Veda
B5 181	ग	◌	Svarita
B6 182	घ	◌	Long Svarita
B7 183	ङ	◌	Kampa
B8 184	च	◌	Anudātta
B9 185	छ	◌	Jātya Svarita (Śukla Yajurveda)
BA 186	ज	◌	Jātya Svarita (Maitrāyaṇīya)
BB 187	झ	◌	Sentence Ending Udātta
BC 188	ञ	◌	Jātya Svarita (Non-Taittirīya Yajurveda)
BD 189	ट	◌	Svarita (Maitrāyaṇīya)
BE 190	ठ	◌	Anudātta (Kāṭhaka)
	MISCELLANEOUS		
BF 191	ड	०	Abbreviation sign

SAMPLE OF ṚGVEDIC TEXT (ACCENTED)

ऋग्वेद संहिता

। अथ प्रथमं मण्डलम् ।

। अथ प्रथमोऽष्टकः ।

प्रथमोऽध्यायः ।। (वर्गाः १-३७)

(९) (म.१, अ.१, सू.१)

ऋषिः:- मधुच्छन्दाः वैश्वामित्रः

छन्दः:- गायत्री

देवताः- अग्निः

ॐ अग्निमीळे पुरोहितं यज्ञस्य देवमृत्विजम् । होतारं रत्नधातमम् ।। १ ।।

अग्निः पूर्वेभिर्ऋषिभिरीड्यो नूतनैरुत । स देवाँ एह वक्षति ।। २ ।।

अग्निना रयिमश्नवत्पोषमेव दिवेदिवे । यशसं वीरवत्तमम् ।। ३ ।।

अग्ने यं यज्ञमध्वरं विश्वतः परिभूरसि । स इद्देवेषु गच्छति ।। ४ ।।

अग्निर्होता कविक्रतुः सत्यश्चित्रश्रवस्तमः । देवो देवेभिरा गमत् ।। ५ ।। १ ।।

यदङ्ग दाशुषे त्वमग्ने भद्रं करिष्यसि । तवेत्तत्सत्यमङ्गिरः ।। ६ ।।

उप त्वाग्ने दिवेदिवे दोषावस्तर्धिया वयम् । नमो भरन्त एमसि ।। ७ ।।

राजन्तमध्वराणां गोपामृतस्य दीदिविम् । वर्धमानं स्वे दमे ।। ८ ।।

स नः पितेव सूनवेऽग्ने सूपायनो भव । सचस्वा नः स्वस्तये ।। ९ ।। २ ।।

(९) (म.१, अ.१, सू.२)

ऋषिः:- मधुच्छन्दाः वैश्वामित्रः

छन्दः:- गायत्री

देवताः- वायुः १-३ इन्द्रवायू ४-६ मित्रावरुणौ ७-९

वायवा याहि दर्शतेमे सोमा अरंकृताः । तेषां पाहि श्रुधी हवम् ।। १ ।।

वाय उक्थेभिर्जरन्ते त्वामच्छा जरितारः । सुतसोमा अहर्विदः ।। २ ।।

वायो तव प्रपृञ्चती धेना जिगाति दाशुषे । उरूची सोमपीतये ।। ३ ।।

इन्द्रवायू इमे सुता उप प्रयोभिरा गतम् । इन्दवो वामुशन्ति हि ।। ४ ।।

वायविन्द्रश्च चेतथः सुतानां वाजिनीवसू । तावा यातमुप द्रवत् ।। ५ ।। ३ ।।

वायविन्द्रश्च सुन्वत आ यातमुप निष्कृतम् । मक्ष्वि १ त्था धिया नरा ।। ६ ।।

SAMPLE OF KṚṢṆA-YAJURVEDIC TEXT (ACCENTED)

।। श्रीरस्तु।।

कृष्णयजुर्वेदीय तैत्तिरीय संहिता

प्रथमकाण्डे प्रथमः प्रश्नः ।।

हरिः ओ(4)म् ।।

इ॒षे त्वो॒र्जे त्वा॑ वा॒यवः॑स्थोपा॒यवः॑स्थ दे॒वो वः॑सवि॒ता प्रार्प॑यतु॒ श्रेष्ठ॑तमाय॒ कर्म॑ण॒ आ प्या॑यध्वमघ्निया देवभा॒गमू॒र्जस्वती॒ः पय॑स्वती॒ः प्र॒जाव॑तीरनमी॒वा अ॑यक्ष्मा मा वः॑स्ते॒न ई॑शत॒ माऽघश॑ꣳसो रु॒द्रस्य॑ हे॒ति॒ः परि॑ वो वृणक्तु ध्रु॒वा अ॒स्मिन्गोप॑तौ स्यात ब॒ह्वीर्यज॑मानस्य प॒शून्पा॑हि ।। 1

इ॒षे त्रिच॑त्वारिꣳशत् । (1)

य॒ज्ञस्य॑ घो॒षद॑सि॒ प्रत्यु॑ष्टꣳ रक्ष॒ः प्रत्यु॑ष्टा अरा॑तय॒ः प्रेयम॑गाद्विषणा॑ बर्॒हिरच्छ॒ मनु॑ना कृ॒ता स्व॒धया॒ वित॑ष्टा त आव॑हन्ति क॒वय॑ः पु॒रस्ता॑द्दे॒वेभ्यो॒ जुष्ट॑मि॒ह ब॒र्॒हिरा॒सद॑े दे॒वाना॑म्परिषू॒तम॑सि वा॒र॒ष॒वृ॒द्धम॑सि॒ देव॑बर्॒हिर्मा त्वाऽन्वङ्मा ति॒र्यक्पर्व॑ ते राध्यासमाच्छेत्ता ते॒ मा रि॑षन्देव॑बर्॒हिश्श॒तव॑ल्॒शाँव्वि रो॑ह स॒हस्र॑वल्॒शा [2] वि वा॒यꣳ रु॑हेम पृथि॒व्यास्स॒म्पृच॑ः पाहि सुस॒म्भृता॑ त्वा॒ सम्भ॑राम्यदि॒त्यै रास्ना॑ऽसीन्द्रा॒ण्यै स॒न्नह॑नम्पू॒षा ते॑ ग्र॒न्थिङ्ग्र॑थ्नातु॒ स ते॒ माऽऽस्था॒दिन्द्र॑स्य त्वा बा॒हुभ्या॒मुद्य॑च्छे बृह॒स्पते॑र्मू॒र्ध्ना ह॑राम्यु॒र्वन्तरि॑क्षमन्वि॑हि देवङ्गम॑मसि ।। [3]

स॒हस्र॑वल्॒शा अ॒ष्टात्रिꣳ॑शच्च । (2)

शु॒न्ध॑ध्व॒न्दैव्या॑य॒ कर्म॑णे देवय॒ज्यायै॑ मात॒रिश्व॑नो घ॒र्मोऽसि॒ द्यौर॑सि पृथि॒व्यसि॑ वि॒श्वध॑राया असि पर॒मेण॒ धाम्ना॒ दृꣳह॑स्व॒ मा ह्वा॑र्॒सूनाम्प॒वित्र॑मसि शत॒धार॑व्वँसू॒नाम्प॒वित्र॑मसि स॒हस्र॑धारꣳ हुतस्तो॒को हु॒तो द्रप्सो॑ऽग्नये॑ बृह॒ते नाका॑य॒ स्वाहा॒ द्यावा॑पृथि॒वीभ्या॒ꣳ सा वि॒श्वायु॒स्सा वि॒श्वव्य॑चा॒स्सा वि॒श्वक॑र्मा सम्पृ॒च्यध्वमृतावरीरू॒र्मिणी॑र्मधु॒मत्तमा॒ मन्द्रा धन॑स्य सा॒तये॒ सोमे॑न॒ त्वाऽऽत॑न॒च्मीन्द्रा॑य॒ दधि॒ विष्णो॑ ह॒व्यꣳ र॑क्षस्व ।। [4]

सोमे॑ना॒ष्टौ च॑ । (3)

SAMPLE OF KṚṢṆA-YAJURVEDIC TEXT (ACCENTED)

कृष्णयजुर्वेदीय तैत्तिरीय ब्राह्मणम्

।। प्रथमाष्टके प्रथमप्रपाठक प्रारंभः।।

।। हरिः ओ(३)म् ।।

ब्रह्म॒ संध॑त्तं॒ तन्मे॑ जिन्वतम् । क्षत्र॑ꣳ संध॑त्तं॒ तन्मे॑ जिन्वतम् । इष॒ꣳ संध॑त्तं॒ तां मे॑ जिन्वतम् । ऊर्ज॒ꣳ संध॑त्तं॒ तां मे॑ जिन्वतम् । र॒यिꣳ संध॑त्तं॒ तां मे॑ जिन्वतम् । पुष्टि॒ꣳ संध॑त्तं॒ तां मे॑ जिन्वतम् । प्र॒जाꣳ संध॑त्तं॒ तां मे॑ जिन्वतम् । प॒शून्थ्संध॑त्तं॒ तान्मे॑ जिन्वतम् । स्तु॒तोऽसि॒ जन॑धाः । दे॒वास्त्वा॑ शुक्र॒पाः॒ प्रण॑यन्तु ।।१।। सु॒वीरा॑ः॒ प्र॒जाः॒ प्र॑ज॒न॒य॒न्परी॑हि । शु॒क्र॒श्शु॒क्रशो॑चिषा । स्तु॒तोऽसि॒ जन॑धाः । दे॒वास्त्वा॑ मन्थि॒पाः॒ प्रण॑यन्तु । सु॒प्र॒जाः॒ प्र॒जाः॒ प्र॑ज॒न॒य॒न्परी॑हि । म॒न्थी म॒न्थिशो॑चिषा । सं॒ज॒ग्मा॒नौ दि॒व आपृ॑थि॒व्याः॑ । संध॑त्तं॒ तन्मे॑ जिन्वतम् । प्रा॒णꣳ संध॑त्तं॒ तं मे॑ जिन्वतम् । अ॒पा॒नꣳ संध॑त्तं॒ तं मे॑ जिन्वतम् ।।२।। व्या॒नꣳ सन्ध॑त्तं॒ तं मे॑ जिन्वतम् । चक्षु॒स्संध॑त्तं॒ तन्मे॑ जिन्वतम् । श्रोत्र॒ꣳ संध॑त्तं॒ तन्मे॑ जिन्वतम् । मन॒स्संध॑त्तं॒ तन्मे॑ जिन्वतम् । वाच॒ꣳ संध॑त्तं॒ तां मे॑ जिन्वतम् । आयु॑स्स्थ॒ आयु॑र्मे धत्तम् । आयु॑र्य॒ज्ञाय॑ धत्तम् । आयु॑र्य॒ज्ञप॑तये धत्तम् । प्रा॒णस्थ॑ाः॒ प्रा॒णं मे॑ धत्तम् । प्रा॒णं य॒ज्ञाय॑ धत्तम् ।।३।। प्रा॒णं य॒ज्ञप॑तये धत्तम् । चक्षु॑स्स्थ॒श्चक्षु॑र्मे धत्तम् । चक्षु॑र्य॒ज्ञाय॑ धत्तम् । चक्षु॑र्य॒ज्ञप॑तये धत्तम् । श्रोत्र॑ꣳ स्थ॒श्श्रोत्रं॑ मे धत्तम् । श्रोत्रं॑ य॒ज्ञाय॑ धत्तम् । श्रोत्रं॑ य॒ज्ञप॑तये धत्तम् । तौ दे॑वौ शुक्रामन्थिनौ । क॒ल्पय॑तं॒ दैवी॒र्विशः॑ । क॒ल्पय॑तं॒ मानु॑षीः ।।४।। इष॒मूर्ज॑म॒स्मासु॑ धत्तम् । प्रा॒णान्प॒शुषु॑ । प्र॒जां मयि॑ च॒ यज॑माने च । निर॑स्त॒श्शण्डः॑ । निर॑स्तो॒ मर्कः॑ । अप॑नुत्तौ शण्डा॒मर्कौ॑ स॒हामुना॑ । शु॒क्रस्य॑ स॒मिद॑सि । म॒न्थिन॑स्स॒मिद॑सि । स प्र॑थ॒मस्संकृ॑तिर्वि॒श्वक॑र्मा । स प्र॑थ॒मो मि॒त्रो व॑रु॒णो अ॒ग्निः । स प्र॑थ॒मो बृ॒ह॒स्पति॑श्चिकि॒त्वान् । तस्मा॒ इन्द्रा॑य सु॒तमा॑जु॒होमि ।।५।।
न॒य॒न्त्व॒पा॒नꣳ सन्ध॑त्तं॒ तं मे॑ जिन्वतम् प्रा॒णं य॒ज्ञाय॑ धत्तं॒ मानु॑षीर॒ग्निर्द्वे च॑ ।।१।।

ब्रह्म॑ क्ष॒त्रं तदिष॒मूर्ज॑ꣳ॑ र॒यिं पुष्टिं॑ प्र॒जां तां प॒शून्तान् ।। संध॑त्तं॒ तत्प्रा॒णम॑पा॒नं व्या॒नं तं चक्षु॑ श्श्रोत्रं॑ मन॒स्तद्वाच॑ं॒ ताम् । इ॒षादि॒ पञ्च॑के॒ वाचं॒ तां मे॑ प॒शून्थ्सन्ध॑त्तं॒ तान्मे॑॑ प्रा॒णादि॒ त्रित॑ये॒ तं मे॒ऽन्यत्र॒ तन्मे॑॑ ।।

कृ॒त्ति॑का॒स्व॒ग्निमा॑द॒धीत । ए॒तद्वा अ॒ग्नेर्न॑क्ष॒त्रम् । यत्कृ॒त्ति॑काः । स्वाया॑मे॒वैनं॑ दे॒वता॑याम॒धाय॑ । ब्र॒ह्म॒व॒र्च॒सी भ॑वति । मुख॑ं॒ वा ए॒तन्न॑क्ष॒त्राणाम् । यत्कृ॒त्ति॑काः । यꣳ कृ॒त्ति॑का॒स्व॒ग्निमा॑धत्ते । मुख्य॑ ए॒व भ॑वति । अथो॒ खलु॑ ।। ६।। अ॒ग्निन॑क्ष॒त्रमित्यप॑चायन्ति । गृ॒हान् ह॒ दाहु॑को भवति । प्र॒जाप॑ती रोहि॒ण्याम॒ग्निम॑सृजत । तं दे॒वा रो॑हि॒ण्यामा॑द॑धत ।

SAMPLE OF SĀMAVEDIC TEXT (ACCENTED)

सामवेद - संहिता

पूर्वार्चिकः (छन्द आर्चिकः)

आग्नेयं काण्डम् ।

अथ प्रथमोऽध्यायः ।

अथ प्रथमप्रपाठके प्रथमोऽर्धः ।

[1]

(१-१०) १-२,४,७,९ भरद्वाजो बार्हस्पत्यः, ३ मेधातिथिः काण्वः, ५ उशनाः काण्वः, ६ सुदीतिपुरुमीढा वाङ्गिरसौ तयोर्वान्यतरः, ८ वत्सः काण्वः, १० वामदेवः ।। अग्निः ।। गायत्री ।।

अग्न आ याहि वीतये गृणानो हव्यदातये। नि होता सत्सि बर्हिषि ।।१

त्वमग्ने यज्ञानां होता विश्वेषां हितः । देवेभिर्मानुषे जने ।।२

अग्निं दूतं वृणीमहे होतारं विश्ववेदसम्। अस्य यज्ञस्य सुक्रतुम् ।।३

अग्निर्वृत्राणि जङ्घनद् द्रविणस्युर्विपन्यया । समिद्धः शुक्र आहुतः ।।४

प्रेष्ठं वो अतिथिꣳ स्तुषे मित्रमिव प्रियम्। अग्ने रथं न वेद्यम् ।।५

त्वं नो अग्ने महोभिः पाहि विश्वस्या अरातेः। उत द्विषो मर्त्यस्य ।।६

एह्यू षु ब्रवाणि तेऽग्न इत्थेतरा गिरः । एभिर्वर्धास इन्दुभिः ।।७

आ ते वत्सो मनो यमत्परमाच्चित्सधस्थात्। अग्ने त्वां कामये गिरा ।।८

त्वामग्ने पुष्करादध्यथर्वा निरमन्थत। मूर्ध्नो विश्वस्य वाघतः ।।९

अग्ने विवस्वदा भरास्मभ्यमूतये महे । देवो ह्यसि नो दृशे ।।१०

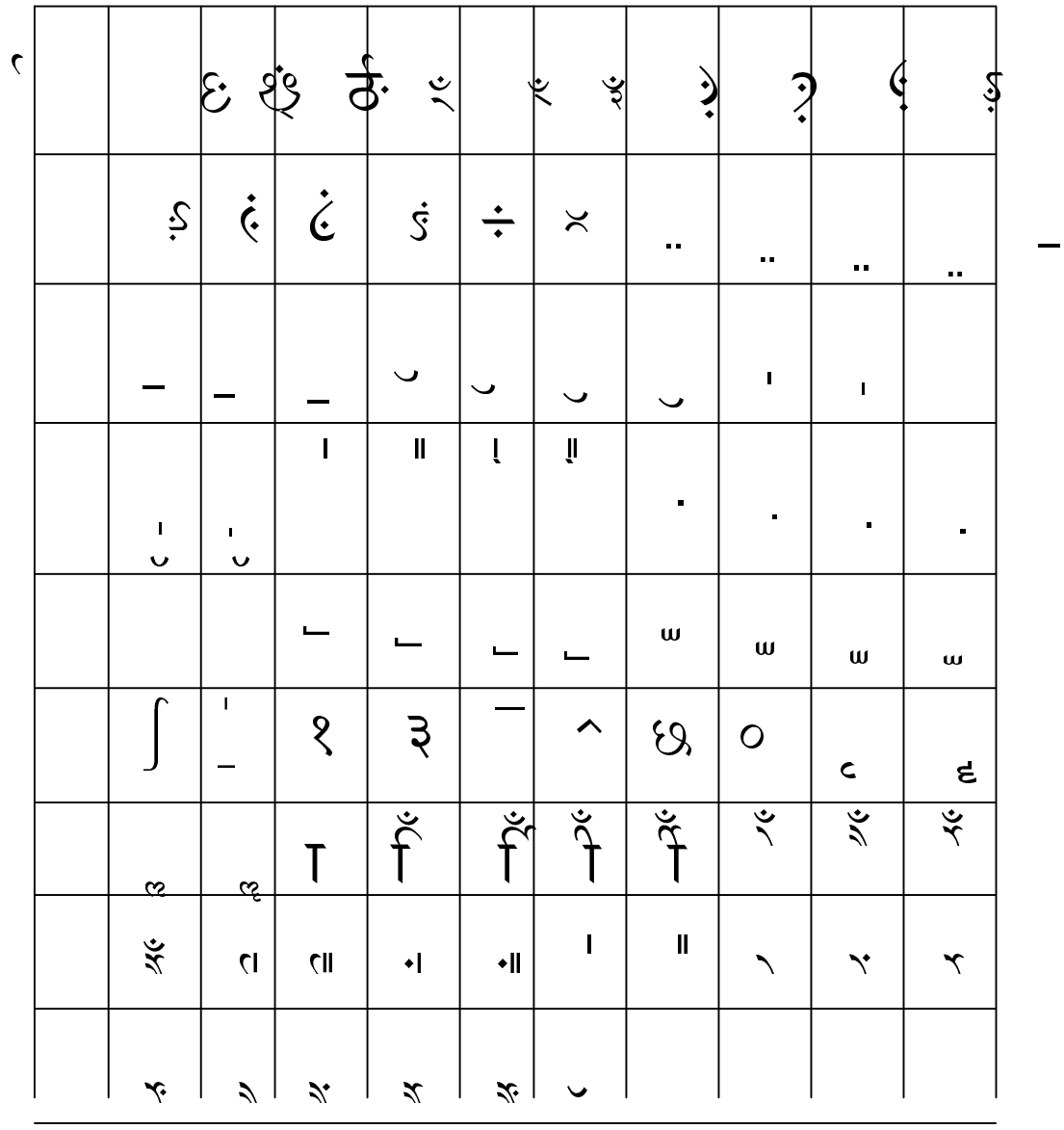
इति प्रथमा दशतिः ।।१।। प्रथमः खण्डः ।।१।। [स्वरिताः ९ । उ० ना० । धा ३७ । वे।।]

[2]

(१-१०) १ आयुङ्क्ष्वाहिः (ऋ० विरूप आङ्गिरसः), २ वामदेवो गौतमः, ३,८-९ प्रयोगो भार्गवः, ४ मधुच्छन्दा वैश्वामित्रः, ५,७ शुनःशेप आजीगर्तिः, ६ मेधातिथिः काण्वः, १० वत्सः काण्वः ।। अग्निः ।। गायत्री ।।

नमस्ते अग्न ओजसे गृणन्ति देव कृष्टयः । अमैरमित्रमर्दय ।।११

KEYMAP OF SANSKRIT (DEVANAGARI) FONT – VEDIC DIACRITIC MARKERS



KEYMAP OF SANSKRIT (DEVANAGARI) FONT – VEDIC DIACRITIC MARKERS

List of verbal root meanings

001 अङ्गान्निस्सरणम् - (waving/moving member of the body)
002 अक्षरविन्यासः - (writing)
003 अग्नि संयोगः - (blowing air/fuelling)
004 अग्रगमनम् - (leading)
005 अङ्कनम् - (marking/defining)
006 अतिक्रमः - (exceeding)
007 अतिसर्जनम् - (giving)
008 अदनम् - (eating)
010 अदर्शनम् - (disappearance)
011 अधः पतनम् - (falling)
012 अधाष्ट्र्यम् - (fear)
013 अनवस्थानम् - (non-stationary/mobile)
014 अनादरः - (ignore/neglect)
015 अनुशिष्टिः - (instruction/order/command)
016 अनृतभाषणम् - (lying)
017 अन्तकर्मन् - (loss)
018 अन्विच्छा - (searching)
019 अन्वेषणम् - (searching)
020 अपनयनम् - (taking away/removal)
021 अपवारणम् - (veiling/covering)
022 अपहरणम् - (stealing)
023 अप्रदानम् - (not giving/refusal)
024 अप्रीतिः - (anger/not being pleased)
025 अभिकांक्षा - (desire)
026 अभिगमनम् - (to approach/go upto)
027 अभिप्रीतिः - (to be pleased)
028 अभिभवः - (to win/humble)
029 अभियोगः - (to meet)
030 अभिवर्धनम् - (to hail)
031 अभिवादनम् - (to salute/to regard/greet)
032 अभिषवः - (bathe/dismembering)
033 अभ्यवहारः - (eating)
034 अभ्यासः - (practice/repeat/memorise)
035 अर्जनम् - (earning)
036 अर्दनम् - (to hurt)
037 अलंकारः - (to decorate/adorn)
038 अलमर्थः - (to suffice)
039 अलाभः - (not getting/to miss)
040 अलीकम् - (to be defective/faulty)
041 अल्पीभावः - (shrink/reduce)
042 अवकल्कनम् - (mix/consider)
043 अवक्षेपणम् - (deride)
044 अवखण्डनम् - (cut)
045 अवगमनम् - (make known/teach)
046 अवगमः - (knowing)
047 अवगाहनम् - (submerge/immerse)
048 अवतंसनम् - (to sound/to make noise)
049 अवदारणम् - (dig)
050 अवध्वंसनम् - (damage)
051 अवबोधनम् - (enlighten/know/think)
052 अवयवः - (to form a part)
053 अवसादनम् - (to decay/to perish)
054 अवस्थानम् - (residing/dwelling)/to hold/carry/bear up)
055 अवस्यन्दनम् - (oozing/trickling)/to dance/to act)
056 अवस्रंसनम् - (falling off/dropping down)
057 अवाप्तिः - (to get)
058 अविशब्दनम् - (to sound)
059 अव्यक्ता वाक् - (to speak barbarously)
060 अव्यक्तः शब्दः - (to hum/to coo/to whistle/to make any inarticulate sound/to roar/to neigh/..)
061 अश्रु विमोचनम् - (to cry)
062 असंस्कारः - (to speak ill/to deceive)
063 असनम् - (throwing/sending)
064 असर्वोपयोगः - (to leave as a residue/remaining/deficient)
065 आक्रोशः - (to curse/to imprecate)
066 आक्षेपः - (to blame)
067 आख्यानम् - (speaking/making known/to know)
068 आघातः - (to strike/to beat)
069 आघ्राणम् - (to smell)
070 आच्छादनम् - (covering/surrounding/hiding)
071 आदरः - (to regard/to care for/to have regard)
072 आदानम् - (taking/receiving)
073 आदेशः - (to instruct/advice)
074 आधारः - (holding/containing)
075 आध्यानम् - (to recollect)
076 आप्यायनम् - (the act of making full)
077 आप्रवणम् - (to raise/to lift)

078 आप्लाव्यम् - (to bathe)
079 आभण्डनम् - (to see/to look at)
080 आभीक्ष्ण्यम् - (continued/repetition)
081 आमन्त्रणम् - (to sound/to call/to invite)
082 आमर्दनम् - (crushing/squeezing)
083 आमर्शनम् - (to touch)
084 आमर्षणम् - (to be angry)
085 आयामः - (stretching/extending)
086 आर्जवम् - (to make/be straight)
087 आर्द्रीभावः - (to make wet/moist)
088 आलस्यम् - (to resist/to oppose/idleness)
089 आलिङ्गनम् - (to embrace)
090 आलोचनम् - (seeing)
091 आवरणम् - (cover)
092 आविष्कारः - (make/bring forth)
093 आशा - (desire/hope/speculation)
094 आशीः - (to bless/to wish good to)
095 आशुगमनम् - (to run)
096 आसेचनम् - (wetting/the act of pouring into)
097 आस्कन्दनम् - (an attack/to ooze)
098 आस्वादनम् - (testing)
099 आह्लादनम् - (gladdening)
100 आह्वानम् - (to challenge/a call or invitation)
101 इच्छा - (desire)
102 इज्या - (teaching/initiating/introducing a pupil)
103 इन्द्रिय-प्रलयः - (to fail in faculties/to go numb/to become hard or stiff)
104 ईप्सा - (desire)
105 ईर्ष्यार्थः - (envious/to feel impatient at others success)
106 ईषद्धसनम् - (to smile/laugh{gently})
107 उच्छ्रायः - (height/exalted)
108 उञ्छः - (gleaning)
109 उत्क्लेशः - (sickness)
110 उत्क्षेपः - (to move/throw up/shake)
111 उत्क्षेपणम् - (throwing upwards)
112 उत्सर्गः - (emission/pouring out)
113 उत्साहः - (effort/increase of energy/ perseverance)
114 उद्गिरणम् - (ejecting/vomitting)
115 उद्यमः - (continued effort)
116 उद्यमनम् - (effort/raising/elevation)
117 उद्वेगः - (to tremble/to shake)
118 उन्दनम् - (to wet/to moisten)
119 उन्दः - (to wet/to moisten)
120 उन्मथनम् - (killing/slaughter)
121 उन्मादः - (madness/extravagance)
122 उन्मादनम् - (to be mad)
123 उन्मानम् - (measuring)
124 उपकरणम् - (to assist/to aid/to favour)
125 उपक्षयः - (waste/decay/expenditure)
126 उपघातः - (to hurt/damage/injury)
127 उपचयः - (accumulation/increase/addition)
128 उपतापः - (sickness/disease)
129 उपदेशः - (instruction/advice/teaching)
130 उपदेहः - (to cover)
131 उपधारणम् - (to study/to practice repeatedly/to exercise)
132 उपनयनम् - (to invest with the sacred thread)
133 उपयाच्ञा - (request/prayer/soliciting)
134 उपरमः - (to stop/to restrain)
135 उपलेपनम् - (smearing an ointment)
136 उपवेशनम् - (sitting down)
137 उपशमः - (becoming quite/assuagement)
138 उपसेवा - (service/worship/honour)
139 उपस्कारः - (beautifying/ornamenting)
140 उपादानम् - (taking/receiving)
141 उपार्जनम् - (acquiring/gaining/earning)
142 ऊर्जनम् - (making animate)
143 एकचर्या - (unmarried)
144 ऐश्वर्यम् - (power/sway/supremacy)
145 कत्थनम् - (boasting)
146 कदाचिद्दर्शनम् - (rare sight)
147 कम्पनम् - (trembling/shaking)
148 करणम् - (doing/performing/executing)
149 कर्णभेदनम् - (piercing ear-lobe)
150 कर्मन् शुभम् - (to do holy)
151 कर्मसमाप्तिः - (extremity/end)
152 कलहकर्मन् - (quarrel)
153 कल्कनम् - (deceiving/overreacting)
154 कल्याणम् - (holy/auspicious)
155 कांक्षा - (wish/desire)

156 कान्तिः - (wish/desire)
157 कान्तिकरणम् - (beautifying/illuminating)
158 कामः - (desire)
159 कार्कश्यम् - (hardness)
160 कालोपदेशः - (denoting time)
161 किञ्चित् चलनम् - (to move)
162 कुटिला गतिः - (to move torturously)
163 कुटुंब धारणम् - (raising/supporting family)
164 कुट्टनम् - (cutting/pounding/abusing)
165 कुत्सनम् - (abuse/contempt/abusive language)
166 कुत्सा - (abuse/contempt/abusive language)
167 कुत्सा गतिः - (running torturously)
168 कुत्सितः शब्दः - (contemptible/mean/hateful words)
169 कु(त्सित)स्मयनम् - (laughing in a bad sense)
170 कृच्छ्र जीवनम् - (living in distress)
171 कृपा गतिः - (slack movement/limping due to sickness etc.)
172 कैतवम् - (deceit/falsehood)
173 कोपः - (passion/wrath/anger)
174 कौटिल्यम् - (deceit)
175 क्रिया - (doing/performing/executing)
176 क्रीडनम् - (playing/sporting)
177 क्रोधः - (anger/not being pleased)
178 क्लेदनम् - (to moisten/to become wet)
179 क्लेशनम् - (pain/anguish/trouble/suffering)
180 क्षमा - (patience/to be patient/to endure)
181 क्षयः - (loss/decay/diminutive)
182 क्षरणम् - (the act of drooping/flowing/oozing)
183 क्षरणार्थः - (to ooze/to drop/to drip)
184 क्षान्तिः - (patience/forgiveness/forbearance)
185 क्षेपः - (to throw)
186 क्षेपणम् - (throwing/sending)
187 क्षोदः - (pounding)
188 खण्डनम् - (cut)
189 खदनम् - (to hurt)
190 खादनम् - (eating)
191 खेदः - (to be distressed)
192 गतिः - (going/moving/knowing)
193 गतिचातुर्यम् - (to go quickly/to run/to trot)
194 गतित्वरणम् - (to go quickly/to make haste)
195 गतिनिवृत्तिः - (to cease to move/to stand still)
196 गतिप्रतिघातः - (to be idle/to be lame/to be impeded)
197 गतिवैकल्यम् - (lame/crippled)
198 गत्यर्थः - (to go/to move)
199 गत्याक्षेपः - (obstruct movement)
200 गत्यादिः - (going etc.)
201 गन्धः - (smell)
202 गन्धनम् - (hurting/killing)
203 गन्धोपादानम् - (to smell)
204 गर्वः - (to be proud)
205 गहनम् - (hide/die)
206 गात्रप्रक्षरणम् - (to sweat/perspire)
207 गात्रप्रस्रवणम् - (to sweat/perspire)
208 गात्रविक्षेपः - (to dance)
209 गात्रविचूर्णनम् - (to injure)
210 गात्रविनामः - (yawning)
211 गार्ध्ये - (to desire eagerly/to long for something)
212 गुणवचनम् - (describe/praise/extol)
213 गुदक्रीडा एव - (to play/sport)
214 गुप्त भाषणम् - (confidential communication)
215 गोपनम् - (to guard/to protect)
216 ग्रन्थः - (to compile/to string or weave together)
217 ग्रन्थनम् - (to tie/connect together)
218 ग्रहणम् - (take)
219 ग्लानिः - (fatigued/tired)
220 ग्लेपनम् - (to move slowly)
221 घनत्वम् - (thickness/robustness)
222 घातनम् - (striking/killing)
223 घोरवाशितम् - (harsh sound/like frogs)
224 चक्यर्थः - (to satisfy/please)
225 चयनम् - (piling/heaping)
226 चलनम् - (shaking)
227 चिकित्सा - (to give medicine)
228 चित्रीकरणम् - (to cause wonder/to be an object of wonder)
229 चिन्ता - (to think of/to recollect)
230 चुम्बनम् - (kissing)
231 चेतना - (to know)
232 चेष्टा - (motion/action)
233 चोदनम् - (throwing)

234 चौर्यम् - (to steal)
235 छद्मगतिः - (go crookedly/deceit)
236 छादनम् - (a cover/screen)
237 छेदनम् - (cutting/dividing)
238 जननम् - (to give birth)
239 जयः - (conquest/triumph/victory)
240 जिघांसा - (desire of killing/injuring/acting as an enemy against)
241 जिह्वोन्मथनम् - (to loll the tongue)
242 जीवनम् - (to be weary)
243 जुगुप्सा - (dislike/separate)
244 ज्ञानम् - (to know/understand)
245 ज्ञीप्सा - (the desire of knowing)
246 तत्क्रिया - (pain/pleasure)
247 तनूकरणम् - (to be thin/small/emasculate)
248 तन्तुसन्ताने - (to weave)
249 तपः - (to make efforts/to exert oneself)
250 तरणम् - (crossing over)
251 तर्जनम् - (threatening)
252 तर्पणम् - (pleasing/satisfying)
253 ताडनम् - (beating/whipping)
254 तितिक्षा - (endurance/patience/to bear)
255 तुष्टिः - (satisfaction/gratification/pleasure)
256 तृप्तिः - (satisfaction)
257 तृप्तियोगः - (to be glad/to be pleased)
258 तेजनम् - (sharpening)
259 तोडनम् - (hurting/injuring)
260 तोषणम् - (to know/satisfaction/contentment)
261 त्यागः - (quit/give up/abandon)
262 त्रासः - (frightening)
263 त्वचनम् - (to wound/cut/chisel)
264 त्वरणम् - (hurry/move with speed)
265 दंशनम् - (the act of hitting)
266 दण्डनिपातः(नम्) - (punishing)
267 दन्दशूकः - (snake)
268 दम्भः - (deceit/fraud)
269 दर्पः - (pride/arrogance)
270 दर्शनम् - (observing/looking)
271 दशनम् - (biting)
272 दानम् - (giving)
273 दाहः - (burning)
274 दीप्तिः - (brightness/splendour/lustre)
275 दुर्गतिः - (to be poor/ail)
276 दुर्गन्धः - (bad odour)
277 दृष्ट्युप(संहारः)घातः - (to blind/obstruct vision)
278 देवनम् - (to sport/play/gamble)
279 देवपूजा - (to worship)
280 देवशब्दः - (thunder)
281 दैन्यम् - (sorrow/grief/low-spiritedness)
282 दैर्घ्यम् - (length/longness)
283 दौर्बल्यम् - (weakness)
284 द्युतिः - (shine/lustre/brightness)
285 द्रवीकरणम् - (melt/liquefy)
286 द्रव्यविनिमयः - (buy/purchase)
287 द्वैधीकरणम् - (to cut into two pieces)
288 धान्यम् - (grain/corn)
289 धान्यावरोधः - (to hoard grain)
290 धारणम् - (holding/carrying/preserving)
291 धाष्ट्यम् - (violence/arrogance/impudence)
292 नयः - (guiding/leading/managing)
293 नाशनम् - (death/destruction/ruin/removal/expulsion)
294 निकारः - (effort/sending)
295 निकेतनम् - (an abode)
296 निगरणम् - (swallowing)
297 निद्राक्षयः - (to be awake)
298 निन्दा - (to censure/blame/defamation)
299 निमज्जनम् - (bathing/diving)
300 निमीलनम् - (shutting the eyelids)
301 निमेषणम् - (to close the eyes)
302 नियमः - (a religious observance)
303 नियो(जनम्)गः - (command/order/instruction)
304 निरसनम् - (expelling/ejecting)
305 निवासः - (a house/an abode)
306 निशानम् - (sharpening)
307 निशामनम् - (seeing/sight)
308 निष्कर्षः - (drawing out/extraction)
309 निष्पत्तिः - (completion/termination/consummation)
310 निष्पाकः - (to boil/to digest)
311 नीचैः गतिः - (going slow/moving slowly)
312 नृतिः(त्तिः) - (dance/dancing)
313 नोच्यते - (general activity)

314 नोपलभ्यते - (to seek/to obtain)
315 न्यक्करणम् - (humiliation/degradation)
316 पदः - (define)
317 पद्यर्थः - (attain/go to)
318 परप्रैष्यः - (to send out)
319 परमैश्वर्यम् - (to rule supreme)
320 परिकल्कनम् - (to cheat/to deceive)
321 परिकूजनम् - (to rejoice)
322 परिग्रहः - (taking/grasping)
323 परिघातः - (killing/striking)
324 परितर्कणम् - (guess/hurt)
325 परितर्पणम् - (satisfaction/pleased)
326 परितापः - (to suffer pain)
327 परिदेवनम् - (lamentation)
328 परिबृंहणम् - (prosperity/welfare)
329 परिभाषणम् - (speaking)
330 परिमाणः - (measuring/weighing)
331 परिवर्तनम् - (turning back or around)
332 परिवेषणम् - (serving eatables)
333 परिवेष्टनम् - (surrounding/covering)
334 परिष्वंगः - (an embrace)
335 परिहाणम् - (decrease/loss/deficient)
336 परिहासः - (to jest/joke/laugh at)
337 पर्याप्तिः - (preserving/guarding/warding off)
338 पवनम् - (purification)
339 पाकः - (cooking/baking)
340 पादविक्षेपः - (to walk/step/jump)
341 पानम् - (drinking)
342 पारुष्यम् - (roughness/hardness)
343 पालनम् - (guarding/protecting/fostering)
344 पिपासा - (feeling thirsty)
345 पुरीषोत्सर्गः - (the voiding of excrement)
346 पुष्टिः - (nourishing/supporting/increasing)
347 पूजनम् - (worshipping/honouring)
348 पूजा - (to worship)
349 पूतीभावः - (cleaning/purifying)
350 पूरणम् - (filling up/completing)
351 पूर्वनिकेतनम् - (earlier place)
352 पृथक्कर्मन् - (to divide/to distribute)
353 पृथग्भावः - (to divide/to separate)
354 पैशुन्यम् - (wickedness/backbiting/roguery)
355 पोषणम् - (nourishing/fostering/supporting)
356 प्रकथनम् - (to tell/to communicate)
357 प्रकाशनम् - (to shine/making bright/illuminating)
358 प्रक्षेपः - (throwing/casting)
359 प्रक्षेपणम् - (throwing/casting)
360 प्रख्यानम् - (to become well known/famous)
361 प्रजनम् - (birth/delivery)
362 प्रणिदानम् - (to exchange/barter)
363 प्रतापनम् - (inflicting/punishment)
364 प्रतिघातः - (to strike against/to resist)
365 प्रतिज्ञानम् - (to promise)
366 प्रतिदानम् - (to restore)
367 प्रतिबन्धः - (stopping/arresting/opposing/ obstacle)
368 प्रतियत्नः - (wish/desire)
369 प्रतिष्टंभः - (obstruction)
370 प्रतिष्ठा - (fixity/strength/firm/foundation)
371 प्रतिहर्षः - (joy/happiness/delight)
372 प्रतीघातः - (resistance)
373 प्रपूरणम् - (satisfying)
374 प्रमर्दनम् - (crushing/destroying)
375 प्रमादः - (a blunder/mistake/to err)
376 प्रमोचनम् - (setting free/liberating)
377 प्रयत्नः - (taking efforts)
378 प्रलंभनम् - (deceiving/cheating)
379 प्रवेशनम् - (entrance/penetration)
380 प्रवेशः - (entrance/penetration)
381 प्रसवः - (source/origin)
382 प्रसहनम् - (defeating/overcoming)
383 प्रसह्यकरणम् - (taking forcibly)
384 प्रसादः - (to be pleased)
385 प्रस्रवणम् - (flowing/oozing/dripping)
386 प्रह्वत्वम् - (bowing humbly/stooping)
387 प्रागल्भ्यम् - (confidence/boldness)
388 प्राणत्यागः - (to die)
389 प्राणधारणम् - (to live/to be alive)
390 प्राणनम् - (breathing/living)
391 प्राणिगर्भविमोचनम् - (to give birth)
392 प्राणिप्रसवः - (birth)
393 प्रादुर्भावः - (arising)
394 प्राधान्यम् - (superiority/predominance)
395 प्रापणम् - (conveying/leading into)
396 प्राप्तिः - (acquisition/gain/attainment)
397 प्रीणनम् - (pleasing/satisfying)

398 प्रीणनार्थः - (pleasing/satisfying)
399 प्रीतिः - (liking/fondness/affection)
400 प्रेक्षणम् - (to see)
401 प्रेरणम् - (sending/despatching)
402 प्लवगतिः - (leaping/jumping)
403 प्लवनम् - (floating/swimming)
404 प्लुतगतिः - (a particular pace of a horse)
405 प्लुतिः - (hopping/jump)
406 बन्धः - (union/connection)
407 बन्धनम् - (fastening/binding/tying)
408 बन्धुः - (a relation)
409 बलम् - (strength/vigour)
410 बाधनम् - (annoyance/oppression)
411 बाल्यम् - (childhood/boyhood)
412 बीजजन्मन् - (to grow/to increase)
413 बीजसन्तानम् - (to sow/to plant)
414 बुभुक्षा - (hunger/desire of eating)
415 बोधनम् - (instruction/teaching)
416 भक्षः - (eating/consuming)
417 भक्षणम् - (eating)
418 भंगः - (breaking/destroying)
419 भयः - (fear/alarm/dread)
420 भरणम् - (bearing/supporting/nourishing)
421 भर्जनम् - (roasting/baking)
422 भर्त्सनम् - (threatening)
423 भषणम् - (the barking of a dog)
424 भस्मीकरणम् - (to make into ashes/burning)
425 भावकरणम् - (to act/to make)
426 भाषणम् - (speaking)
427 भाषार्थः - (speech)
428 भासनम् - (shining/glittering)
429 भिक्षणम् - (asking for alms/begging)
430 भीमार्थः - (frightful/terrible)
431 भूः - (becoming)
432 भूतप्रादुर्भावः - (to come forth/to appear)
434 भूषणम् - (decorating/adorning)
435 भूषा - (jewel/ornament)
436 भृतिः - (bear/support/maintain/nourish/service for wages)
437 भेद(ः)नम् - (divide/break)
438 भोजनम् - (eating)
439 भमणम् - (to turn around)
440 मण्डनम् - (the act of decorating/adorning)
441 मदः - (to be proud)
442 मदनम् - (make arrogant)
443 मन्दा गतिः - (to move slowly/to creep)
444 मन्थः - (churning/agitating)
445 मर्दनम् - (crushing/grinding)
446 मर्षणम् - (endurance/patience)
447 महत्त्वम् - (great/much/extensive)
448 मांगल्यम् - (auspicious/prosperity/welfare)
449 मानम् - (pride/haughtiness)
450 मानसम् - (mutter/whisper)
451 मारणम् - (killing/slaughter/destruction)
452 मार्गणम् - (an arrow)
453 मार्गसंस्कारः - (paving)
454 मार्जनम् - (cleaning)
455 मिश्रणम् - (mixing/combining)
456 मूर्तिभावः - (to become hard)
457 मेधा - (intelligence/the power of memory)
458 मैथुनम् - (copulation)
459 मोक्षणम् - (resigning/rescuing/setting at liberty)
460 मोचनम् - (freeing/discharging)
461 मोदः - (delight/gladness)
462 मोदनम् - (joy/pleasure)
463 मोहः - (swoon/fainting)
464 मोहनम् - (arrogance)
465 मौण्ड्यम् - (to initiate a pupil)
466 म्रक्षणम् - (heaping up)
467 म्लेच्छनम् - (speaking barbarously)
468 याचनम् - (begging)
469 याच्ञा - (begging/solicitation)
470 युद्धः - (war)
471 योगः - (junction/union)
472 रक्षणम् - (guarding/protecting/watching)
473 रक्षा - (defence/protection)
474 रागः - (dying/color/hue)
475 राभस्यम् - (to begin/commence)
476 रुजा - (disease/illness)
477 रुजाविशरणम् - (suffer sickness)
478 रूपक्रिया - (figuring/beholding)
479 रेषणम् - (howling/neighing/roaring/yelling)
480 रोगः - (disease)
481 रोगापनयनम् - (to cure)
482 रोदनम् - (weeping/crying)

483 रोधनम् - (to stop/to restrain)
484 रोषः - (anger/wrath)
485 रोषकृतम् - (to be angry)
485 रोषणम् - (angry/shout)
486 रोहणम् - (the act of mounting/growing)
487 लक्षणम् - (a mark/sign)
488 लज्जा - (feeling of shame)
489 लंभनम् - (attainment/acquiring)
490 लवनम् - (cutting)
491 लाभः - (obtaining/gaining/acquisition)
492 लिप्सा - (obtaining/to search for/to seek/to inquire for)
493 लोडनम् - (disturbing/agitating)
494 लौल्यम् - (fickleness/unsteady)
495 वक्त्रसंयोगः - (to kiss)
496 वज्रनिर्घोषः - (a clap of thunder)
497 वदनैकदेशः - (the whole side of the face including the temple)
498 वमनम् - (vomiting)
499 वयोहानिः - (to become old/aging)
500 वरणम् - (to choose/select)
501 वर्जनम् - (abandoning/losing/giving up)
502 वर्णः - (color)
503 वर्णक्रिया - (to paint)
504 वर्तनम् - (being)
505 वर्षा - (raining)
506 वसनम् - (clothing/dressing)
507 वाक्यप्रबन्धः - (connected composition)
508 वादित्रग्रहणम् - (to play on an instrument)
509 वारणम् - (opposition/resistance)
510 विकसनम् - (blossoming/opening/expanding)
511 विक्रान्तिः - (heroism/prowess)
512 विक्षेपः - (casting/throwing/projecting)
513 विचारणम् - (discussion/understanding)
514 विजिगीषा - (desire to conquer)
515 विज्ञानम् - (knowledge/science)
516 वितर्कः - (to consider/think)
517 वित्तसमुत्सर्गः - (to expend)
518 विदारणम् - (to break/split/crack)
519 विद्योपादानम् - (learn/to acquire knowledge)
520 विधानम् - (to rule/govern)
521 विधूननम् - (agitation/tremor)
522 विनिन्दनम् - (blame/reproach/censure)
523 विपरीत मैथुनम् - (intercourse)
524 विप्रयोगः - (disunion/separation)
525 विबाधनम् - (pain/agony/anguish)
526 विबाधा - (pain/agony/anguish)
527 विभागः - (division/apportioning)
528 विभाजनम् - (distributing)
529 विमोचनम् - (liberation/emancipation)
530 विमोहनम् - (tempting/seducing)
531 वियोजनम् - (division/separation)
532 विरेचनम् - (evacuation/to empty)
533 विलसनम् - (sporting)
534 विलासः - (to sport)
535 विलेखनम् - (splitting/dividing)
536 विलोडनम् - (agitation/shaking/churning)
537 विवासः - (banishment/exile/expulsion)
538 विशब्दनम् - (to announce/declare)
539 विशरणम् - (splitting)
540 विशेषणम् - (distinguishing)
541 विश्राणनम् - (distribute)
542 विश्वासः - (trust/confidence)
543 विसर्गः - (confidence)
544 विस्तारः - (let go/let loose)
545 विस्तार वचनम् - (expansion/extension)
546 विस्तृतिः - (spread out/extended)
547 विस्मापनम् - (illusion/deceit)
548 विस्मृतिः - (forgetfulness)
549 विहायसा गतिः - (flying up)
550 विहारः - (sport/play)
551 वृणोत्यर्थः - (choose/select)
552 वृद्धिः - (increase)
553 वृद्ध्यर्थः - (to increase)
554 वेतिना तुल्यम् - (weaving)
555 वेष्टनम् - (surrounding/encircle)
556 वैकल्यम् - (agitation/flurry)
557 वैकृत्यम् - (change/alter)
558 वैक्लव्यम् - (confusion/grief)
559 वैचित्त्यम् - (grief/mental distraction)
560 व्यक्ता वाक् - (to speak/tell)
561 व्यक्तिः - (distinction)
562 व्यक्तीकरणम् - (explaining/clarifying)
563 व्यथनम् - (giving pain)

564 व्यवहारः - (business/profession)
565 व्याकुलत्वम् - (perplexing/confusion)
566 व्याजीकरणम् - (deception/deceive/defraud)
567 व्याप्तिः - (extending/widely spread)
568 व्यायामः - (stretching out/to bring out of)
569 व्रतादेशः - (investiture with the sacred thread)
570 व्रीडः - (shame)
571 शक्तिः - (ability/power/strength)
572 शक्तिबन्धनम् - (to be capable of generating)
573 शक्य(ा)र्थः - (being capable)
574 शंका - (doubt/uncertainty)
575 शंकुबन्धनम् - (forcibly/violently)
576 शब्दः - (sound/noise)
577 शब्दः तारः - (to utter a shrill cry)
578 शब्द कुत्सा - (to laugh at/ridicule/mock)
579 शब्दक्रिया - (to sound/make any noise)
580 शब्दार्थः - (to sound)
581 शयः - (sleep)
582 शातनम् - (withering/decaying)
583 शासनम् - (order/command/ruling)
584 शास्त्रम् - (order/instruction)
585 शिल्पयोगः - (shine/glitter)
586 शीघ्रार्थः - (quick/speedy)
587 शुद्धिः - (purification)
588 शैथिल्यम् - (looseness/laxity)
589 शोकः - (sorrow/grief/anguish)
590 शोधनम् - (cleaning/purifying)
591 शोभार्थः - (to look beautiful)
592 शोषणम् - (drying up)
593 शौचः - (cleansing/purification)
594 शौचकर्मन् - (purification)
595 श्रद्धा - (composure)
596 श्रवणम् - (to hear)
597 श्लाघा - (praise/self praise)
598 श्लेषणम् - (embrace/cling)
599 श्वैत्यम् - (whiteness)
600 संकोचनम् - (contract/shrink)
601 संक्लेशनम् - (suffering/pain)
602 संख्यानम् - (counting)
603 संगतिकरणम् - (worship/sacrifice)
604 संगमः - (meeting/union)
605 संघर्षः - (emulation/rivalry)
606 संघातः - (multitude)
607 संचयः - (collection/gathering)
608 संचलनम् - (trembling/shaking)
609 संचूर्णनम् - (grind/crush)
610 संचेतनम् - (the mind)
611 संचोदनम् - (sending/directing)
612 संज्ञानम् - (knowledge)
613 संतर्जनम् - (blame/censure)
614 संतानः - (extending/spreading)
615 संतानक्रिया - (extending/spreading)
616 संतापः - (heat/fire)
617 संदर्भः - (weaving/collecting)
618 संदीपनम् - (inflaming/kindling)
619 संदेश(वच)नम् - (information/news)
620 संनिकर्षणम् - (approximating/approaching)
621 संपर्कः - (contact)
622 संपर्चनम् - (to join/mix)
623 संपे(प्रे)षणम् - (to strike against)
624 संप्रस्रवणम् - (to sprinkle)
625 संप्रहारः - (striking/wounding/beating)
626 संबन्धः - (connection)
627 संभक्तिः - (to honour/worship)
628 संभमः - (haste)
629 संयमनम् - (to bring into contact)
630 संराद्धिः - (to cook/prepare)
631 संवरणम् - (covering)
632 संवेष्टनम् - (surrounding/encircling)
633 संशब्दनम् - (telling/narrating)
634 संशयः - (doubt/uncertainty)
635 संश्लेषणम् - (to obstruct)
636 संसर्गः - (mixture/union/contact)
637 संसिद्धिः - (complete accomplishment)
638 संस्त्यानम् - (assemblage)
639 संस्पर्श(नम्)ः - (contact)
640 संगः - (joining/uniting/come in contact)
641 सत्ता - (existence/being)
642 समवायः - (conjunction/union/cohesion)
643 समाघातः - (war/battle/killing/slaughter)
644 समाधिः - (perfect absorption of thought into the object of meditation)
645 समुच्छ्रायः - (height/elevation)
646 समृद्धिः - (prosperity/well-being)
647 सम्यगवभाषणम् - (to speak ill)

648 सहनम् - (enduring/bearing)
649 सातत्यम् - (continuity)
650 सातत्यगमनम् - (to go constantly)
651 सान्त्वनम् - (reconcile)
652 साम(न्त्व)प्रयोगः - (reconcile)
653 सामर्थ्यम् - (force/power/ability)
654 सुखम् - (joy/pleasure)
655 सुखनम् - (make happy/please)
656 सेचनम् - (sprinkling)
657 सेवनम् - (serving/attending)
658 सेवा - (service/attendance)
659 स्खदनम् - (cutting/hurting/harassing)
660 स्तंभः - (to fix/to make rigid/to paralyse)
661 स्तंभनम् - (paralysing/numbing/stunning)
662 स्तवनम् - (praising)
663 स्तुतिः - (praise)
664 स्तेयः - (theft/robbery)
665 स्तेयकरणम् - (stealing)
666 स्थानम् - (standing/staying/being stationary)
667 स्थैर्यम् - (firmness)
668 स्थौल्यम् - (bulkiness/being big)
669 स्नेह(ः)नम् - (anointing/lubricating)
670 स्पर्धा - (envy/rivalry/emulation)
671 स्पर्शः - (touch/contact)
672 स्पर्शनम् - (touching/handling)
673 स्मरणम् - (remembering)
674 स्मृतिः - (memory/remembrance)
675 स्रवणम् - (oozing/flowing)
676 स्वप्नः - (sleep/sleeping)
677 स्वाम्यर्थः - (mastership/lordship)
678 हन्त्यर्थः - (killing)
679 हरणम् - (partake/stealing/destroying/carrying)
680 हरितभावः - (greening)
681 हर्षः - (joy)
682 हर्षक्षयः - (to grow weary/despond/be dispirited)
683 हसनम् - (laughing)
684 हानिः - (decrease/damage/abandon)
685 हावकरणम् - (express coquettishly)
686 हिंसनम् - (hurting/injuring/killing)
687 हिंसा - (hurting/injuring/killing)
688 हिंसार्थः - (hurting/injuring/killing)
689 हूर्छनम् - (deceiving/arrogating)
690 ह्वरणम् - (to be crooked/to deceive)

Some root meanings referred to in Pāṇini's Aṣṭādhyāyī [*Format* : (Adhyāya,pāda,sūtra#, Sūtram)

1,3,015 न गतिहिंसार्थेभ्यः
1,3,087 निगरण चलनार्थेभ्यः
1,4,025 भीत्रार्थानाम् भयहेतुः
1,4,027 वारणार्थानाम् ईप्सितः
1,4,033 रुच्यर्थानाम् प्रीयमाणः
1,4,037 क्रुधद्रुहेर्ष्याऽसूयार्थानाम् यं प्रति कोपः
1,4,052 गतिबुद्धिप्रत्यवसानार्थ शब्दकर्माकर्मकाणामणि कर्ता स णौ
2,3,012 गत्यर्थकर्मणि द्वितीयाचतुर्थ्यौ चेष्टायामनध्वनि
2,3,052 अधीगर्थदयेशां कर्मणि
2,3,054 रुजार्थानां भाववचनानामज्वरेः
3,2,148 चलनशब्दार्थादकर्मकाद्युच्
3,2,151 क्रुधमण्डार्थेभ्यश्च
3,2,188 मतिबुद्धिपूजार्थेभ्यश्च
3,3,129 छन्दसि गत्यर्थेभ्यः
3,3,146 किंकिलास्त्यर्थेषु लृट्
3,3,157 इच्छार्थेषु लिङ्लोटौ
3,3,160 इच्छार्थेभ्यो विभाषा वर्तमाने
3,4,048 हिंसार्थानां च समानकर्तृकाणाम्
3,4,065 शकधृषज्ञाग्लाघटरभलभक्रमसहार्हास्त्यर्थेषु तुमुन्
3,4,072 गत्यर्थाकर्मकश्लिषशीङ्स्थासवसजनरुहजीर्यतिभ्यश्च
3,4,076 क्तोऽधिकरणे च ध्रौव्यगतिप्रत्यवसानार्थेभ्यः
6,2,080 उपमानं शब्दार्थप्रकृतावेव
8,1,025 पश्यार्थैश्चानालोचने
8,1,051 गत्यर्थलोटा लृण्नचेत्कारकं सर्वान्यत्

Pāṇinian Rules

Words in Sanskrit have mainly two parts, a base and a suffix. The base can take various suffixes/declensions to denote different functional relationships (cases or persons) and numbers (singular, dual and plural). The grammar lays down rules for declining the word bases depending on various parameters.

Rule Structure:

The grammar rules are formally structured by Pāṇini in his *Aṣṭādhyāyī*. Pāṇini's Aṣṭādhyāyī is arranged in eight chapters (*Adhyāya*), each chapter containing four sections or quarters (P*āda-s*) and every section containing several number of rules (*Sūtra-s*). *Sūtra-s* are not sentences - but like formulae or program code, it would run out to sentences after certain algebraic operations were carried out. In all, Pāṇini's formal language formulates about 4000 rules. These rules are used for deriving the word forms, verbs, nouns case endings, compound words and to explain syntax and semantics of the language.

Pāṇinian rules are essentially meant to exhaustively cover the generative aspects of Sanskrit language. The style of the rules is descriptive and covers a generalisation of all the prevalent valid word forms as observed by Pāṇini. The grammar rules are of various types with a section on *metarules* which layout the various categories of rules.

The main aim of the rule structure was to achieve utmost brevity without sacrificing clarity. The rules are of the following types :-

1) **Definitional (*saṁjñā*) Rules** : These describe certain *technical terms* used to denote certain longer characters, strings etc. Whenever, these technical terms occur in the grammar rules we will have to replace them with their denoted forms. There are 84 such technical terms defined by Pāṇini. For eg., *gha* denotes the two suffixes *tarap* and *tamap*. Similarly, feminine words ending in *i* or *u* are termed as *nadī*. Operational rules applicable to the denoted terms are greatly facilitated by the use of technical terms. They are called *saṁjñā.*

2) **Interpretive (paribhāṣā) Rules** : Most of the operational rules describe one of the three character/string processes of *addition* (augmentation), *deletion* (elision) or *replacement* (substitution). Obviously, these three processes can have one or more operands. The applicable set of operands, number of them, their sequence i.e., LHS/RHS operand and the position where the process should take place etc., are denoted by the *case endings* used with the operands. The interpretive rules define the way these case endings occuring in the operational rules are to be understood.

3) **Jurisdictional (adhikāra) Rules** : Rules are arranged on topical basis and the scope of this is defined as *Adhikāra*. All operational rules coming under a particular topic will inherit the *Adhikāra sūtra* words. The above two types may be called *meta-rules*, *i.e*, rules about rules.

4) **Inheritence (anuvṛtti) Rules** : For sake of convenience, operands may be grouped together in such a way that certain words of operational rules get carried forward, till certain specific number of rules. This is called *Anuvṛtti* (partial inheritence) which is used for achieving brevity.

5) **Operational (vidhi) Rules** : These define one of the three above said processes for particular operands. These are *Injunctive* in nature i.e., say, *if this condition is satisfied then do the following*. The

next type will be the exception to these processes fully or partially. These are respectively called *niṣedha* (exception) and *niyama* (restriction).

6) **Miscellaneous (atideśa) Rules** : There are also a few rules which list certain *finished words* as valid, as in these cases, the generative process is irregular or diverse.

The types of rules mentioned above are arranged in different hierarchy while applying them. When more than one operational rules are applicable in a given case, interpretive rules are referred to for determination of priority.

There are four types of supersessions based on the relative priority in *aṣṭādhyāyī* namely :-

i) Latter rule supercedes the former.
ii) Rules applicable to a particular member or subset suprecede those of a generic class or set.
iii) First 29 quarters (*pāda-s*) supercedes the last three quarters.
iv) Exceptions supercedes the general rules.

Even between these four types of supersessions there may be a conflict between two types, in which case the latter type is given precedence to the former.

General algorithm for using a Pāṇinian rule :

1) Choose the particular rule and note down its position or reference number in terms of chapter, quarter and *sūtra* number.

2) Check if the rule contains any *Technical terms* and if so take its denoted term.

3) Check for *Adhikāra* under which jurisdiction the rule under question falls and supply the *Adhikāra* rule words.

4) Find out if any word(s) are to be inherited from other rules through *Anuvṛtti*. If so, add those words.

5) Syntactically analyse each word of the chosen rule and get the base and case.

6) Determine the process involved, operands, position of operation etc., by invoking the interpretive rules. Decide on priorities if the rule is of more than one type.

7) Construct the *long hand form* of the selected rule from the details obtained through these steps.

8) Make a mention of databases involved for reference like *dhātupāṭha* (list of verbal roots), *gaṇapāṭha* (list of nominal words sharing certain common syntactic properties), *uṇādi-sūtra* (list of suffixes) etc., if any.

We illustrate this procedure through an example :

Let the rule chosen be *ād guṇaḥ*.

1) The reference number of this rule is 6.1.87.

2) This rule contains a technical term *guṇaḥ* (refer 1.1.2). This technical term refers to the three characters *a*, *e* and *o*. We now replace the word *guṇaḥ* with *a*, *e* and *o*.

3) This rule comes under the *Adhikāra* (jurisdiction) of *Saṃhitāyām* (6.1.72) i.e., in continuous utterance.

4) By *Anuvṛtti* we inherit the words *ekah* and *pūrva-parayoh* from 6.1.84 and *aci* from 6.1.77.

5) This rule has two words *āt* and *guṇaḥ*. After syntactic analysis, the base and case of the two words are respectively (*a*,5) and (*guṇaa*,1). For *Saṃhitāyām* the base is *Saṃhitā* and case 7, for *aci* it is (*ac*,7), for *ekah* it is (*eka*,1) and for *pūrva-parayoh* it is (*pūrva-para*,6).

6) Applying the interpretive rules, the base *a* of *āt* denotes the vowel-phoneme class *a* by 1.1.69. Case 5 denotes the LHS operand by 1.1.67. Case 7 denotes the RHS operand by 1.1.66. Case 6 denotes the place of operation by 1.1.49. In case, the substituend is to be found from a list of candidates, the one that is the closest is chosen by 1.1.50. By traditional interpretation, case 1 denotes the substituend. Also *pūrva* denotes LHS and *para* the RHS. *eka* denotes singularity.

7) From the above steps we arrive at the *sūtra* in long hand form (meaning) as :"in continuous utterance, the vowel-phoneme class *a* and a following vowel (*ac*) are both replaced by *a*, *e* or *o*". The actual replacement among *a*, *e* and *o* is the one that is closest to the RHS vowel.

8) This rule does not require any database for interpretation.

For implementation of this procedure we need all the rules of *aṣṭādhyāyī* with the reference number, jurisdictions, inheritances, list of all technical terms, all interpretive conditions and analysis package.

Pāṇini's Kāraka theory

By observation of the language Sanskrit used by native speakers of his times, particularly the educated class, Pāṇini arrived at certain definite semantic features inherent in the inflections, the nominal bases undertook. These are wide categories by default, while many related shades of subtle difference in senses are described under each of these categories. The cases could denote certain default case relations as well as certain specific semantic features or sub-functional categories. Thus, there could be certain specific relation marker for the cases as will be explained.

These case relations could be assigned active functional roles in sentences involving verbal activities. In descriptive sentences, these relations may have a diluted sense. As an effort to unambiguously describe the generative syntax of the language, for any particular sense to be conveyed, syntactic case markers were identified with a well-defined relation to semantics. Intentionality (vivakṣā) is another factor in ascribing active roles to the functors.

Six basic categories or case relations were propounded with their syntactic markers as below.

(1) An independent thing signifying existence of a substance, its gender, quantity or number with no special marker other than just a noun.

When an active role is intended, an animate category of functor is usually used. For example:
'Rāma kills Rāvaṇa by arrows',
'The enemy is getting killed by Rāma',
'I worship Rāma',
'Rāma blesses the devotee',
as against an inanimate thing like
'the automobile goes',
'cancer kills'.

The independence in active mode could be exercised to cause an action directly or through another active or passive agent. For example,
'Daśaratha sends Rama to the forest',
'Rāma goes to the forest',
'Rāma gets the feast cooked by the cooks'.

Volition or absence of it in an independent being can also lead to a different active role. For example:
'the youth died suddenly '
'he shook out of cold'.

Certain reflexive actions do not involve using the independence present in the actor. For example,
'The food is being cooked by Rāma himself',
'the wood breaks by itself'.

Here the utmost ease with which an activity is accomplished makes the choice of reflexive formation marked by "... self". Swatting mosquitoes while asleep is typically the example for reflexive action.

(2) An item most desired to be reached immediately by the action of the agent is the second case relation. This could as well cover a most undesired one. A combination of these is also possible. This includes producing, modifying, or attaining the object of desire or otherwise. Here also no special markers are used. A few examples follow:

(i) 'The weaver weaves the cloth' (producing).
(ii) 'The goldsmith embeds diamond in Rāma's earring' (refining).
(iii) 'I pray Rāma' (attaining -mental).
(iv) 'Out of lust Rāvaṇa steals Sītā' (undesirable).
(v) 'Angered, he consumes poison' (undesirable).
(vi) 'The hungry child eats polluted food unknowingly' (desired and undesired).
(vii) 'The cow is milked by Kṛṣṇa' (passive).

Since these two cases and the relations are not specific to any particular situation, for any active sentence these are required essentially. Intransitive actions don't take any object.

(3) The most effective means for accomplishing the verbal activity is the third functional relation. This corresponds to the syntactic marker denoting instrumentality like by, with, because of, due to, etc..

Reason or cause gets included under the means. The means could be internal (bodily organs etc.) or external. The agent gets this marker in passive sentences. For example,

(i) 'Rāma cuts off Rāvaṇa's head by sharp arrows' .
(ii) 'Rāma reaches Ayodhyā by palanquin with Sītā'.
(iii) 'I mentally pray Rāma'.
(iv) 'Lakṣmaṇa signals Rāma by his eyes'.

(4) The next functional relation is that of the item one has in view through the object (or activity). This corresponds to the syntactic marker to, for, for the sake of, etc.. The recipient in an act of giving or the beneficiary. The person pleased in activities relating to taste etc. The target of praise, curse, envy, anger, etc.. The creditor, object intended, person being inquired about, the prompter in case of promise, sacrificial recitation, the consentor, non-refuser, etc. gets these markers. For example,

(i) 'Give me salvation, O! Lord'.
(ii) 'Rāma grants release to his devotee'.
(iii) 'The hermit brings fruit for Rāma'.
(iv) 'The devotee offer prayers to God'.

(5) The next relation is one of the reference while separation, this could be static or mobile while the separation could be physical or mental. This also covers the cause of fear, the intolerable one, the one intended to be warded of, the source or origin, picking one from a lot, etc.. This case-relation corresponds to the syntactic marker from in general. For example,

(i) 'The leaves fall from the tree'. (fixed reference)
(ii) 'Warrior falls from a running horse'. (mobile reference)
(iii) 'The goats retreat from each other'. (relative reference)
(iv) 'He looks from the balcony'. (transcendental separation)
(v) 'The army fears from snake'. (cause of fear)
(vi) 'The king protects his subjects from theft'. (protection)
(vii) 'The students dread from studies'. (intolerable)
(viii) 'Kautsa learns from Varatantu'. (systematic learning)
(ix) 'Kṛṣṇa hides from his mother'. (veil or cover)
(x) 'Arjuna is the best among the archers'. (picking)
(xi) 'Ganges flows from Himālaya-s'. (source)
(xii) 'Living beings emerge from God'. (material/instrumental cause)

(6) The last functional relation is the locus of the activity. This could be a time, space, topic, proximity etc. This could further be pervasive or partial. This corresponds to the semantic marker in, at, on etc. The examples are:

(i) 'Oil is in the seeds', 'there is light in the day' (pervasive).
(ii) 'Rāma sits on the throne', 'the teacher taught for thirty years' (partial contact).
(iii) 'He is good in mathematics', 'the climate is fine in spring' (topical).
(iv) 'The hamlet is on the river', 'I reside in M.G. Road', 'I am just joining you' (proximity').

These six are the functional relations of which, the last four are optional in a given construction (with a few exceptions). Further the sixth case is not considered a functional relation as it has no direct

connection in terms of expectancies to create a total sense out of the sentences. These words describe certain additional features of the different functors.

The mandatory Kāraka-s (agent and object) may be denoted under certain conditions and certain peculiar speaker intention based usages ('of whom when enraged even the deities fear in the warfield' is a typical usage in Rāmāyaṇa, where fifth case ought to have been used instead of the sixth) are the only instances where functional relations could be denoted.

Activity categories - agent/object word type specification

Format : (#Meaning Group, #Activity Meaning, (semantic type/condition)

(1,3,"अग्रगमनम्","(स्यात् : {संख्या : कर्ता >1})")
(1,5,"अतिक्रमः","(स्थानम् : अतीतः(कर्म))")
(1,9,"अधःपतनम्","(दिक् : अधः)")
(1,24,"अभिगमनम्","(कर्म विधः चेतनः)")
(1,77,"आप्रवणम्","(विधः : मुहुः)(दिक् : ऊर्ध्वम्))")
(1,78,"आप्लाव्यम्","(विधः : मुहुः)(दिक् : ऊर्ध्वम्))")
(1,95,"आशुगमनम्","(विधः : त्वरितः)")
(1,97,"आस्कन्दनम्","(कर्ता : द्रव्यम् : अश्वः : तिर्यक् : मूर्तम्:)")
(1,157,"कुटिलागतिः","(कर्ता : द्रव्यम् : मूर्तम् : तिर्यक् : सर्पः)")
(1,162,"कुत्सागतिः","(विधः : स्खलिता)")
(1,166,"कृपगतिः","(कर्ता अवस्था दीना)")
(1,187,"आगमनम्","(उपसर्ग : आ)(पूर्व स्थानम् कर्म)")
(1,187,"निर्गमनम्","(उपसर्ग : निर्)(पूर्व क्रिया आगमनम्)")
(1,188,"गतिचातुर्यम्","(कर्ता : अश्वः)")
(1,190,"गतिप्रतिघातः","(विधः : निरुद्धा)")
(1,191,"गतिवैकल्यम्","(कर्ता अवस्था खञ्जः)")
(1,225,"छद्मगतिः","(कर्ता इच्छा निलयनम्)")
(1,241,"तरणम्","(अधिकरणम् : अन्यतरः(वायुः, जलम्))")
(1,330,"पादविक्षेपः","(करणम् : पादः)")
(1,366,"(प्रवेशः","")
(1,432,"मन्दागतिः","(विधः मन्दः)")
(1,472,"रोहणम्","(दिक् ऊर्ध्वम्)")
(1,533,"विहायसागतिः","(अधिकरणम् : वायुः)")
(1,550,"व्याप्तिः","(अधिकरणम् अवस्था व्याप्ता)")
(1,601,"संनिकर्षः","(स्थानम् निकटम्(कर्म))")
(1,634,"सातत्यगमनम्","(कर्म : अनिर्दिष्टः)")

Activity specifications (in alphabetical order of root meanings)

Format : (#Activity Group, #Activity, Root_Meaning, (functor type/semantic condition)

(2,1,"अक्षरविन्यासः","(Object=Topic),(Instrument=writing_material)")
(9,2,"अग्निसंयोगः","(Instrument type pipe),(Output oven burn)")
(1,3,"अग्रगमनम्","(Agent number greater(1)) (Agent=first_agent)(Agent t*)
(2,4,"अङ्कनम्","(output attribute definition)")
(3,4,"अङ्कनम्","(Object secondary topic)")
(1,5,"अतिक्रमः","(Object position (beyond normal)")
(9,6,"अतिसर्जनम्","(#1 Agent has Object), (#2 Recipient has Object)")
(9,7,"अदनम्","(Object type edible)")
(10,8,"अदर्शनम्","(output Object lost)")
(1,9,"अधःपतनम्","(Agent movement downwards)")
(10,10,"अधारृष्ट्यम्","(Object state afraid)")
(2,11,"अध्ययनम्","(Object= Topic), (Activity type continuous")
(2,12,"अनवस्थानम्","(knowledge type false)")

(10,12,"अनवस्थानम्","(Agent movement rotary)")

(5,13,"अनादरः","(Attention nil)")

(3,14,"अनुशिष्टः","(Object secondary command)")

(3,15,"अनृतभाषणम्","(Message type false)")

(8,16,"अन्तकर्मन्","(Object becomes non-existent)")

(2,17,"अन्विच्छा","(present=Object),(location object domain)")

(3,19,"अपनयनम्","(Messge emphasis negative)")

(4,19,"अपनयनम्","(#precond Object),(#postcond location)")

(4,20,"अपवारणम्","(Result = Object : not seen)")

(4,21,"अपहरणम्','(#1 Possession with owner) (#2 Possession with Agent)

(2,22,"अप्रीतिः","(Emotive type negative)")

(2,23,"अभिकाङ्क्षा","(Emotive type negative)")

(1,24,"अभिगमनम्","(Object type sentient)")

(10,26,"अभिभवः","(Object state victorious)")

(8,27,"अभियोगः","(Activity type aggressive)")

(3,28,"अभिवर्धनम्","(Message emotive positive)")

(5,29,"अभिवादनम्","(Object is superior)")

(8,30,"अभिषवः","(Object lose juice)")

(10,30,"अभिषवः","(Object state cleaning)")

(2,32,"अभ्यासः","(Object =Topic),(Activity type repeated)")

(4,33,"अमिश्रणम्","(#precond contact with other (object)),(#postcond **)

(6,34,"अर्जनम्","(Effort present), (output desired Object)")

(7,36,"अलंकरणम्","(output look pleasent)")

(6,39,"अलाभः","(effort present),(output absent)")

(3,40,"अलीकः",")")

(2,42,"अवकल्कनम्","(Object =Topic)")

(3,43,"अवक्षेपणम्","(message emotive negative)")

(10,46,"अवगाहनम्","(medium type water)")

(8,48,"अवदारणम्","(Object becomes pieces)")

(2,50,"अवबोधनम्","(Object primary sentient),(Object secondary topic)")

(4,53,"अवस्थानम्","(#precond any location), (#postcond = #1)")

(6,56,"अवाप्तिः","(Output desired Object)")

(10,58,"अव्यक्तःशब्दः","(Sound type unintelligible)")

(10,60,"अश्रुविमोचनम्","(Output eyes tearful)")

(7,61,"असंस्कारः","(Object look unpleasent)")

(4,62,"असनम्",'(#precond any location), (#postcond object moved away)")

(5,63,"असर्वोपयोगः","(Activity utilizing), (output not_zero)")

(2,68,"आघ्राणम्","(Activity type sensory)"), (instrument = nose)")

(5,70,"आदरः","(Attention full)")

(6,71,"आदानम्",'(Accept object)")

(1,77,"आप्रवणम्","(Agent movement and (repeated, upwards))")

(1,78,"आप्लाव्यम्","(Agent movement and (repeated, upwards))")

(2,79,"आभंडनम्","(Object =Topic), (knowledge type true)")

(3,81,"आमन्त्रणम्","(Message type addressig)")

(8,82,"आमर्दनम्","(30)")

(2,83,"आमर्शनम्","(Activity type sensory), (Instrument=skin)")

(2,84,"आमर्षणम्","(Emotive type negative)")

(9,87,"आर्द्रीभावः","(Object becomes wet)")

(5,89,"आलिङ्गनम्","(Present physical contact), (emotion positive)')

(2,93,"आशा","(Emotive type positive)")

(5,94,"आशी)","(wishing good)")

(1,95,"आशुगमनम्","(Agent movement rate (fast))")

(1,97,"आस्कन्दनम्","(Agent movement movement : horse)")

(2,99,"आह्लादः","(Emotive type positive)")

(5,102,"इज्या","(Object is superior)")

(10,103,"इन्द्रियप्रलयः","(Organ becomes inactive)")

(2,105,"ईष्यार्थः","(Emotive type negative)")

(6,108,"उञ्छः","(Source many), (Object type grain)")

(4,110,"उत्क्षेपः","(Direction movement upwards)")

(4,112,"उत्सर्गः","(#1 Present object), (#2 not (#1)), (movement type voluntary*)

(9,114,"उद्गीरणम्","(Eaten thing comes out)")

(9,115,"उद्यमः",'(Agent puts effort)")

(9,121,"उन्मानम्","(Object compared against reference), (parameter type*)

(7,122,"उपकरणम्","(helps change)")

(7,129,"उपनयनम्",'(Activity taking near), (Activity type preaching)")

(6,130,"उपयाच्ञा",'(Object type donor), (Agent attribute poor)")

(7,132,"उपलेपनम्","(Coating with liquid)")

(5,135,"उपसेवा","(Object is superior), (Activity pleasing)")

(4,137,"उपादानम्","(#1 Possession with any) (#2 possession with agent)")

(7,139,"ऊर्जनम्","(Output object more vigorous)")

(9,141,"ऐश्वर्यम्","(Agent possess control)")

(3,142,"कत्थनम्","(Message type exaggerated)")

(8,145,"कर्णभेदनम्","(Object type ear)")

(9,146,"कर्मसमाप्तिः","(Work state finished)")

(5,149,"कल्याणम्","(Output happiness)")

(7,152,"कान्तिकरणम्','(Object becomes radiant)")

(3,155,"कालोपदेशः","(Object secondary time)")

(1,157,"कुटिलागतिः","(Agent movement movement:snake)")

(9.158,"कुटुम्बधारणम्","(Object type or [family, relatives])")

(8,159,"कुट्टनम्","(Object becomes powder)")

(2,160,"कुत्सनम्','(Emotive type negative)")

(1,162,"कुत्सागतिः","(Agent movement tottering)")

(1,166,"कृपागतिः","(Agent condition weak)")

(2,179,"क्षान्तिः","(Emotive type positive)')

(2,186,"खेदः","(Emotive type negative)")

(1,187,"गतिः","(Upasarga=आ), (source=destination)")

(1,187,"गतिः","(Upasarga= निर्), (previous activity coming)")

(1,188,"गतिचातुर्यम्","(Agent =Horse)")

(1,190,"गतिप्रतिघातः","(Agent movement obstructed)")

(1,191,"गतिवैकल्यम्","(Agent condition cripple)")

(2,197,"गर्वः","(Emotive type negative)")

(8,201,"गात्रविचूर्णनम्","(Object becomes wounded)")

(3,205,"गुप्तपरिभाषणम्","(Message type confidential)")

(2,207,"ग्रन्थः","(Object type topic)")

(4,216,"चयनम्","(#1 location number many), (#2 location number one)")

(5,220,"चुम्बनम्","(Present lips contact), (Emotive positive)")

(9,222,"चेष्टा","(Activity type reflexive)")

****

(2,235,"ज्ञापनम्",(output =knowledge)")

(2,236,"ज्ञीप्सा","(Object primary sentient), (Object secondary topic)")

(5,237,"तत्क्रिया","(Result=or[pain pleasure]))")

(7,238,"तनूकरणम्","(Object thickness reduces)")

(9,239,"तन्तुसंतानम्","(Input type threads), (output type cloth)")

(9,240,"तपस्","(Effort type intense)")

(1,241,"तरणम्","(Medium=or(air, water))")

(3,242,"तर्जनम्","(message emotive negative)")

(8,242,"तर्जनम्","(Activity type physical)")
****

(7,249,"तेजनम्","(Object sharpness increasses)')

(9,252,"त्यागः","(Object is abandoned)")
****

(8,256,"दंशनम्","(Instrument type teeth)")

(9,257,"दण्डनिपातनम्","(Object type sentient), (Object state uncomfortable)")*

(2,261,"दर्शनम्","(Activity type sensory), (Instrument = eyes)")

(8,268,"दृष्ट्युपघातः","(Object becomes blind)")

(5,270,"देवपूजा","(Object is deity)")

(7,275,"द्रवीकरणम्","(Object becomes liquid)")

(4,276,"द्रव्यविनिमयः","(Object exchanged), (Instrument type money)")

(6,278,"धान्यम्","(Object type grain)")

(4,279,"धान्यावरोधनम्","(Object=grain)")

(4,282,"नयः","(Present movement)")

(5,290,"निमन्त्रणम्","(Activity religious)")

(5,293,"नियमः","(Activity religious), (follows rules)")

(4,295,"निरसनम्","(Source=body organ), (Object=excrete)")

((2,298,"निशामनम्","(Activity type sensory), (Instrument=or(eyes,ears))")

(4,299,"निष्कर्षः","(Object=any)")

(9,301,"निष्पाकः","(Object type edible), (Instrument type heat)")

(2,305,"न्यक्करणम्","(Emotive type negative)")

(9,308,"परप्रेष्यम्","(Object=person)")

(9,309,"परमैश्वर्यम्","(Control type absolute)")

(7,314,"परिणामः","(Object state changes)")

(7,318,"परिबृंहणम्","(Object size increases)")

(7,321,"परिवर्तनम्","(Object form changes)")

(9,322,"परिवेषणम्","(Object type edible), (Instrument type crockery)")

(4,323,"परिवेष्टनम्","(Object surrounded/covered)")

(4,325,"परिहाणम्","(#1 Object quantity any), (#2 Object quantity < #1)")

(7,328,"पवनम्","(Object state cleaner)")

(10,329,"पाकः","(Instrument type time)")

(1,330,"पादविक्षेपः","(Agent type animate)",(instrument = feet)")

(9,331,"पानम्","(Object state liquid), (Object type edible)")

(3,332,"पारुष्यम्","(Tone=harsh), (content=harsh)")

(9,333,"पालनम्","(Output object secure)")

(9,336,"पुष्टिः","(Output object healthy)")

(4,340,"पूरणम्","((#1 Object quantity any), (#2 Object quantity or[equal,full])*

(4,341,"पृथक्कर्मन्","((#precond number any), (#postcond number <#1)")

(3,343,"पैशुन्यम्","(Message emotive negative)")

(4,351,"प्रणिदानम्","((Object exchanged), (Instrument type other object)")

(3,354,"प्रतिज्ञानम्","(Message type assurance)")

(4,355,"प्रतिबन्धः","(Movement stopped)")

(4,363,"प्रमोचनम्","(#1 bound), (#2 not (#1))")

(2,365,"प्रलंभनम्","(Emotive type negative)")

(1,366,"प्रवेशः","(Agent type animate), (Object type enclosure)")

(3,368,"प्रसवः","(Message type consent)")

(2,371,"प्रसादः","(Emotive type positive)")

(4,382,"प्रापणम्","(Present movement), (Object movement dependent agent)")

(9,388,"प्रेरणम्","(Object type any)")

(9,393,"बन्धः","(Object state confined)")

(9,395,"बन्धुः"<"(Object mixing with_people)")

(9,396,"बलम्","(Output object strong)")

(10,399,"बीजजन्मन्","(Agent type plant)")

(9,400,"बीजसंतानम्","(Object type seeds)")

(10,401,"बुभुक्षा","(Agent condition hungry)")

(10,410,"भषणम्","(Agent type dog)")

(3,413,"भावकरणम्","(Message type intention)")

(10,420,"भूतप्रादुर्भावः","(Birth time delayed)")

(9,425,"भोजननिवृत्तिः","(Eating not present)")

(7,431,"मन्थः","(Output essence_of object)")

(1,432,"मन्दगचिः","(Agent movement rate(slow))")

(7,442,"मिश्रणम्","(combine objects), (output single object)")

(7,443,"मूर्तिभावः","(Object state solid)")

(5,445,"मैथुनम्","(Present body contact), (Emotion positive)")

(4,447,"(मोचनम्","(Object like sweat)")

(5,452,"मौण्डः","(Activity religious), (#2 head shaved), (object=head)")

(8,463,"रुजा","(Object condition diseased)")

(10,465,"रेषणम्","(Agent category wolf)")

(7,467,"रोगपनायनम्","(Object becomes healthy), (Object type animate)")

(1,472,"रोहणम्","(Agent movement upwards)")

(2,479,"लौल्यम्","(Mental_state=unsteady)")

(4,492,"वादित्रग्रहणम्","(Object type musical_instrument)")

(4,496,"विक्षेपः","((#1 Location number one), (#2 Location number many)")

(9,501,"वित्तसमुत्सर्गः","(Object type money)")

(2,503,"विद्योपदानम्","(Agent category student)")

(4,516,"विरेचनम्","(((#1 Object quantity equal capacity), (#2 Object quantity nil))*

(9,521,"विवासः","(Object location outside_limits)")

(3,523,"विशब्दनम्","(Pitch=high)")

(2,527,"विश्वासः","(Knowledge type sure)")

(4,528,"विसर्गः","(Object emitting)")

(9,532,"विस्मापनम्","(Object state or[shocked, stunned])")

(1,533,"विहायसागतिः",'(Medium=air)")

(7,540,"वैकृत्यम्","(Object condition bad)")

(8,548,"(व्याकुलत्वम्","(Object state confused)")

(1,550,"व्याप्तिः","(Medium condition occupied)")

(5,552,"व्रतः","((Activity religious), (follows rules)")

(9,566,"शिल्पयोगः","(Activity type art)")

(8,574,"शोषणम्","(Object state dry)")

(2,578,"श्रनणम्","(Activity type sensory), (Instrument=eyes)")

(7,582,"श्वेत्यम्","(Object becomes white)")

(7,583,"संकोचः","(Object size decreases)")

(9,586,"संख्यानम्","(Parameter type size)")

(6,599,"संदर्भः","(Object type or[flowers,books])")

(7,600,"संदीपनम्","(Object becomes brighter)")

(1,601,"संनिकर्षः","(Object distance small)")

(7,602,"संपर्कः","(Object has contact)")

(6,619,"संस्त्यानम्","(Result union)")

(1,634,"सातत्यगमनम्","(Object type unspecified)")

(3,635,"सातत्यम्","(Message emotive negative), (pitch=high)")

(10,635,"सातत्यम्","(Activity is continous)")

(3,636,"सान्त्वनम्","(Message emotive positive)")

(2,644,"सेवा",(Emotive type positive)")

(5,647,"स्तवनम्","(Object type superior), (Activity singing praise)")

(10,667,"हर्षक्षयः","(Agent category animate)")

PRELIMINARY SEMANTIC LEXICON - (ONTOLOGICAL DATABASE)

Format : *lex*(headword, [sub-types], [instances], [suffix-types])

lex("प्रमेय", ["द्रव्य", "गुण", "क्रिया", "संबन्ध"], [], [])

lex("द्रव्य", ["मूर्त", "अमूर्त"], ["ज्ञान", "आत्मन्", "जीव", "ईश्वर", "घट", "पट", "बाष्प"], [])

lex("मूर्त", ["प्रत्यञ्च्", "तैजस", "आप्य", "पार्थिव", "पराञ्च्", "हेतु", "लक्ष्य", "द्रव"], ["घन", "चित्र", "शस्त्र"], [])

lex("अमूर्त", ["वायव्य", "आकाश", "काल", "इन्द्रिय", "विषय"], [], [])

lex("गुण", ["विशेषण", "शब्द", "स्पर्श", "रूप", "रस", "गन्ध", "चित्तवृत्ति", "परिमाण"], ["गुण", "धर्म", "योग", "स्वत्व", "अंश", "संज्ञा"], [])

lex("शब्द", ["सर्वनाम", "कर्मप्रवचनीय", "तुल्यार्थ", "लक्षण"], ["ध्वनि", "कठोर", "व्यक्त", "अविशद", "आम्रेडित", "सान्त्व", "संगत", "हृदयंगम", "निष्ठुर", "परुष", "ग्राम्य", "अश्लील", "सूनृत", "सत्य", "अवाच्य", "श्राव्य", "विस्पष्ट", "वितथ", "अनृत", "ऋत", "सम्यञ्च्", "तथ्य", "कल", "मन्द्र", "गंभीर", "तार"], [])

lex("स्पर्श", [], ["उष्ण", "शीत", "कोष्ण", "कवोष्ण", "मन्दोष्ण", "कदुष्ण"], [])

lex("रूप", ["वर्ण"], ["सामर्थ्य", "तेजस्", "तीक्ष्ण", "घनत्व", "प्रकाश", "शोभा"], [])

lex("रस", [], ["कटु", "अम्ल", "लवण", "मधुर", "तिक्त", "तुवर", "कषाय"], [])

lex("गन्ध", [], ["परिमल", "आमोद", "समाकर्षिन्", "निर्हारिन्", "सुरभि", "घ्राणतर्पण", "इष्टगन्ध", "सुगन्धि", "पूतिगन्धि", "दुर्गन्ध", "विस्र"], [])

lex("चित्तवृत्ति", ["सत्व", "रजस्", "तमस्"], [], [])

lex("सत्व", [], ["प्रीति", "सुख", "स्थैर्य", "हर्ष", "प्रसाद", "प्रमोद", "ज्ञान", "अध्यवसाय", "अभिमान", "भावना", "विवक्षा", "अनुज्ञा", "सौहार्द"], [])

lex("रजस्", [], ["इच्छा", "लज्जा", "प्रयत्न"], [])

lex("तमस्", [], ["भीति", "भय", "द्वेष", "दुःख", "काम", "क्रोध", "कोप", "मद", "मात्सर्य"], [])

lex("वर्ण", [], ["श्वेत", "पीत", "हरित", "रक्त", "श्याम", "पिङ्गल", "कर्बुर", "शुक्ल ", "
"पाण्डर", "अवदात", "सित", "गौर", "वलक्ष", "धवल", "अर्जुन", "हरिण", "पाण्डुर", "पाण्डु", "धूसर", "नील", "असित", "काल", "श्यामल", "मेचक", "पालाश", "हरित", "लोहित", "रोहित", "रक्त", "शोण", "अरुण", "पाटल", "श्याव", "कपिश", "धूम्र", "कपिल", "कद्रु", "चित्र", "किर्मीर", "कल्माष", "शबल"], [])

lex("प्रत्यञ्च्", ["देवता", "मनुष्य", "तिर्यञ्च्"], ["पुरुष", "स्त्री"], [end("त"), gender("पुल्लिङ्गः"), suf("शतृ")])

lex("देवता", [], ["सुर", "देव"], [])

lex("मनुष्य", ["छात्र", "अध्यापक", "बान्धव", "अङ्गविकल"], ["नर", "दीन", "खञ्ज", "क्षुधित", "दरिद्र", "दास", "प्रेष्य", "दातृ", "विजयिन्", "प्रीयमाण", "गच्छत्", "राम", "कृष्ण", "देवासुर", "देवदत्त", "यज्ञदत्त", "अस्मद्", "युष्मद्", "रामा", "मैत्र", "चैत्र", "सीता", "शबरी", "शिव", "ब्राह्मण", "अर्जुन", "लक्ष्मी"], [])

lex("तिर्यञ्च्", ["मृग", "पक्षिन्"], [], [])

lex("छात्र", [], ["शिष्य", "अन्तेवासिन्"], [])

lex("अध्यापक", [], ["उपाध्याय", "आचार्य", "गुरु"], [])

lex("बान्धव", [], ["पितृ", "मातृ", "भ्रातृ", "स्वसृ", "पुत्र", "दुहितृ", "पति", "पत्नी", "कुटुम्ब"], [])

lex("प्राण्यंग", [], ["शिरस्", "मस्तक", "पाद", "हस्त", "कर", "वाच्", "चरण", "चक्षुस्", "अक्षि", "नेत्र", "दृश्", "श्रोत्र", "श्रवण", "शरीर", "दशन", "ओष्ठ", "मुख", "नासिका", "त्वच्", "अक्ष", "इन्द्रिय", "मनस्"], [])

lex("अङ्गविकल", [], ["अन्ध", "काण", "कुणि", "बधिर", "मूक", "पङ्गु"], [])

lex("दोह्य", [], ["गो", "अजा"], [])

lex("मृग", ["दोह्य"], ["अश्व", "सैन्धव", "सर्प", "श्वन्", "वृक"], [])

lex("पक्षिन्", [], ["गृध्र", "गरुड", "काक"], [])

lex("तैजस", [], ["दीप", "नक्षत्र", "सूर्य", "चन्द्र", "ज्योतिस्", "हिरण्य", "रजत", "विद्युत्", "अग्नि"], [])

lex("आप्य", [], ["जल", "नदी", "शिशिर", "हिम", "अप्"], [])

lex("पार्थिव", [], ["पृथिवी"], [])

lex("पराञ्च्", ["स्थावर", "प्राण्यंग", "पुष्प", "वाद्य", "अमेध्य", "हेय", "मध्यमान", "भोज्य"], ["तृण", "नौ", "अन्न", "अक्ष"], [])

lex("स्थावर", [], ["पर्वत", "वृक्ष", "समुद्र"], [])

lex("वायव्य", [], ["वायु", "मरुत्", "प्रकम्पन", "झञ्झावात"], [])

lex("आकाश", ["देश"], [], [])

lex("इन्द्रिय", [], ["पाद", "पायु", "उपस्थ", "पाणि", "वाच्", "चक्षुस्", "श्रोत्र", "घ्राण", "रसना", "त्वच्", "मनस्"], [])

lex("काल", ["नक्षत्र-लुप्"], ["काल", "दिन", "समय", "पक्ष", "मास", "ऋतु", "अयन", "संवत्सर", "वर्ष", "रात्रि", "प्रभात", "अहन्"], [])

lex("देश", ["स्थावर", "दिश्"], ["ख", "वियत्", "गम्यस्थान", "ग्राम", "वन", "गृह", "देवालय", "आश्रम", "देश", "जनपद", "स्थान", "अन्तर", "क्षेत्र", "प्रासाद", "पाटलिपुत्र", "मन्दिर", "आपण", "नगर"], [])

lex("प्रकृति", [], ["तन्तु", "मृत्", "हिरण्य", "अयस्"], [])

lex("प्रकृष्ट", ["अध्यापक", "देवता"], ["पूज्य", "महित", "सत्", "महत्"], [])

lex("इष्ट", ["छात्र", "देवता"], ["प्रिय", "सुहृत्", "भक्त", "शेष"], [])

lex("अमेध्य", [], ["घर्मस्रुति", "पुरीष", "मूत्र", "निष्ठ्यूत"], [])

lex("हेय", [], ["अधर्म", "पाप", "दुश्चरित"], [])

lex("धान्य", [], ["व्रीहि", "यव", "माष", "तिल", "तण्डुल", "बीज"], [])

lex("विषय", ["पुस्तक"], ["वार्ता", "धर्म", "पाठ", "सन्देश", "शासन", "रहस्य"], [])

lex("वाद्य", [], ["वीणा", "मृदङ्ग", "दुन्दुभि", "भेरी", "आतोद्य", "आनक", "पटह"], [])

lex("पुष्प", [], ["कमल", "सुमनस्", "सुम", "पुष्प", "मल्लिका"], [])

lex("पुस्तक", [], ["ग्रन्थ", "काव्य", "पुराण", "श्रुति", "स्मृति", "इतिहास", "आगम", "शास्त्र"], [])

lex("भोज्य", ["हविस्", "धान्य"], ["पयस्", "भोज्य", "अदनीय", "अन्न", "ओदन", "सूप", "अपूप", "फल", "दधि", "मधु", "घृत", "क्षीर", "सैन्धव", "लवण", "मोदक"], [])

lex("द्रव", [], ["पयस्", "जल", "क्षीर", "यवागू", "अमृत", "मधु", "रस", "दुग्ध"], [])

lex("मथ्यमान", [], ["अमृत", "सार", "नवनीत", "सुधा"], [])

lex("कारक", ["कर्तृ", "कर्म", "करण", "संप्रदान", "अपादान", "अधिकरण"], [], [])

lex("कर्तृ", ["प्रत्यङ्"], ["कारक"], [])

lex("कर्म", ["पराङ्", "प्रमेय"], [], [])

lex("संप्रदान", ["प्रत्यङ्", "चित्तवृत्ति", "पुरुषार्थ"], [], [])

lex("अपादान", ["देश", "भयहेतु", "हेतु", "प्रकृति"], ["पात्र"], [])

lex("अधिकरण", ["देश", "काल"], ["जल", "वायु", "पृथिवी", "आकाश", "स्थाली"], [])

lex("करण", ["द्रव्य", "गुण", "हेतु"], ["बाण", "भक्ति", "लेखनी", "अग्नि", "धमनी", "धन", "मान", "पात्र", "काष्ठ", "वित्त", "साधन", "करण", "यान"], [suf("ल्युट्")])

lex("भयहेतु", ["मृग"], ["चोर", "चौर", "मृत्यु", "दैव"], [])

lex("प्रकृति", [], ["तन्तु", "मृत्", "हिरण्य", "अयस्"], [])

lex("पुरुषार्थ", [], ["धर्म", "अर्थ", "काम", "मोक्ष", "प्रीति", "तृप्ति", "धन", "यशस्", "श्री", "अमृत", "श्रेयस्", "कीर्ति", "फल"], [])

lex("कल्प", ["देश", "काल", "अध्वन्"], ["गोदोह", "कुरु", "पाञ्चाल"], [])

lex("अध्वन्", [], ["क्रोश", "मार्ग", "पथिन्", "अध्वन्"], [])

lex("लक्षण", [], ["जटा", "शिखा"], [])

lex("लक्ष्य", [], ["तापस", "परिव्राजक"], [])

lex("हेतु", ["चित्तवृत्ति"], ["धन", "विद्या", "शत", "जाड्य", "प्रज्ञा", "निमित्त", "कारण"], [])

lex("दिश्", [], ["पूर्व", "दक्षिण", "पश्चिम", "उत्तर"], [])

lex("नक्षत्र-लुप्", [], ["आश्विन", "आपभरण", "पुष्य", "मूल"], [])

lex("तुल्यार्थ", [], ["सदृश", "तुल्य", "निभ", "संकाश", "तुला", "उपमा"], [])

lex("हविस्", [], ["समिध्", "चरु", "आज्य", "हविस्", "वपा", "मेदस्", "पशु", "छाग"], [])

lex("सर्वनाम", [], ["सर्व", "विश्व", "उभ", "उभय", "ततर", "ततम", "कतर", "कतम", "यतर", "यतम", "इतर", "अन्य", "अन्यतर", "त्वत्", "त्व", "नेम", "सम", "सिम", "पूर्व", "पर", "अवर", "दक्षिण", "उत्तर", "अपर", "अधर", "स्वम्", "अन्तरम्", "त्यद्", "तद्", "यद्", "एतद्", "इदम्", "अदस्", "एक", "द्वि", "युष्मद्", "अस्मद्", "भवतु", "किम्"], [])

lex("कर्मप्रवचनीय", [], ["अनु", "उप", "अप", "परि", "आ", "प्रति", "अभि", "अधि", "सु", "अति", "अपि"], [])

lex("परिमाण", [], ["द्रोण", "आढक", "खारी", "मानिका", "कार्षापण"], [])

lex("क्रिया", [], [], [])

lex("संबन्ध", ["एकनिष्ठ", "द्विनिष्ठ"], ["प्रतियोगिन्", "अनुयोगिन्", "संबन्धिन्", "कर्तृ", "कर्मन्", "करण", "अधिकरण"], [])

lex("एकनिष्ठ", [], ["आश्रय", "आश्रित", "निरूपक", "निरूप्य", "जनक", "जन्य", "व्यापक", "व्याप्य"], [])

lex("द्विनिष्ठ", [], ["पितृ", "मातृ", "भ्रातृ", "स्वसृ", "पुत्र", "दुहितृ", "पुत्री", "सुत", "भृत्य", "स्वामिन्", "सेवक", "भट", "राजन्", "पति", "पत्नी", "भर्तृ", "भार्या", "आचार्य", "शिष्य", "अन्तेवासिन्", "गुरु", "मैत्र"], [])

lex("क्रियाविशेषण", [], ["शीघ्र", "त्वरित", "लघु", "क्षिप्र", "अर", "द्रुत", "तीव्र", "एकान्त", "नितान्त", "गाढ", "बाढ", "दृढ", "कोष्ण", "कवोष्ण", "तिग्म", "तीक्ष्ण", "खर", "पाप", "पुण्य", "सुख", "शिव", "भद्र", "कल्याण", "मंगल", "शुभ", "प्रकाण्ड"], [])

lex("द्रव्यविशेषण", ["विशेष्यनिघ्न"], ["शीघ्र", "त्वरित", "लघु", "क्षिप्र", "अर", "द्रुत", "तीव्र", "एकान्त", "नितान्त", "गाढ", "बाढ", "दृढ", "कोष्ण", "कवोष्ण", "तिग्म", "तीक्ष्ण", "खर", "पाप", "पुण्य", "सुख", "शिव", "भद्र", "कल्याण", "मंगल", "शुभ", "प्रकाण्ड"], [])

lex("विशेष्यनिघ्न", [], ["धन्य", "महाशय", "प्रवीण", "निपुण", "अभिज्ञ", "निष्णात", "कृतिन्", "कुशल", "पूज्य", "वदान्य", "बहुप्रद", "आयुष्मत्", "परीक्षक", "वरद", "प्रमनस्", "दुर्मनस्", "विमनस्", "सरल", "उदार", "तत्पर", "आसक्त", "प्रथित", "ख्यात", "विज्ञात", "विश्रुत", "आढ्य", "धनिन्", "स्वामिन्", "ईश्वर", "नायक", "प्रभु", "अधिप", "दयालु", "कारुणिक", "श्रीमत्", "स्वतन्त्र", "परतन्त्र", "पराधीन", "अधीन", "निघ्न", "आयत्त", "मन्द", "क्षुधित", "लुब्ध", "लोलुप", "अविनीत", "विनीत", "मत्त", "कामुक", "विधेय", "वश्य", "धृष्ट", "प्रगल्भ", "स्तब्ध", "ज्ञातृ"], [])

lex("विशेषण", ["द्रव्यविशेषण", "क्रियाविशेषण"], ["सुन्दर", "रुचिर", "चारु", "सुषम", "साधु", "शोभन", "मनोहर", "कान्त", "मनोज्ञ", "मञ्जु", "अभीष्ट", "हृद्य", "दयित", "वल्लभ", "प्रिय", "निकृष्ट", "प्रकृष्ट", "दिव्य", "मलिन", "मेध्य"], [])