Contents

I picked up Professor Madhav Deshpande's book in November 2020. These are my notes from his book. While studying from his book, I explored a variety of relevant resources available on the WWW. These notes rely heavily on those wonderful resources. These notes are a unification of what I could find on various topics as seen from my perspective.

1 Introduction

The book begins with an overview of Sanskrit that we shall cover below. It touches upon the language's history and evolution, writing system, sounds, grammar, and syntax.

1.1 Grammar

Sanskrit (संस्कृतम्), like Greek and Latin, is an inflected language. This means it shows alteration in form especially by adding affixes. The bulk of grammatical information is carried by morphology (i.e. the rules for forming admissible words).

A morpheme which is the minimal meaningful language unit, is of one of these types:

1. nominal stem (adjectives, pronouns, and indeclinables(अव्ययम्))
    - primary
    - secondary
        - derived from other nominals via affixation (e.g. कुरु + अ = कौरव, नर + त्व = नरत्व)
        - derived from verbal roots via affixation (e.g. गम् + अन = गमन, कृ + तृ = कर्तृ)
        - compounds (e.g. नर + पति = नरपति, चक्र + पाणि = चक्रपाणि)
2. verbal root (धातुः)
    - primary
    - secondary
3. indeclinables (अव्ययम्)
    - particles (e.g. उपरि)
    - pre-positions (e.g. अधि, परि, अनु)
    - post-positions
    - adverbs (e.g. सततम्)
    - connectives (e.g. च, वा)
    - (occasionally) nouns

The nominal stem is characterized by gender as an intrinsic property and it is grammatical, usually unrelated to semantics (though the living beings are usually masculine or feminine). There are three genders:

- masculine,
- feminine, and
- neuter

Between masculine and feminine, the former is generic, meaning it takes precedence. For pronouns, neuter is the most generic.

Declension of nouns (as we shall later see, declension serves the same purpose that prepositions serve in English) is affected by several factors such as their

- gender (masculine, feminine, neuter),
- final sound or sounds of the stem (e.g. अकारान्त, न्-कारान्त),
- number (singular, dual, and plural), and
- case (प्रथमा : nominative – I, द्वितीया : accusative – II, तृतीया : instrumental – III, चतुर्थी : dative – IV, पञ्चमी : ablative – V, षष्ठी : genitive – VI, सप्तमी : locative – VII, सम्बोधनम् : vocative – VIII). The following list may help describe the usual purpose of cases:

    1. nominative – serving as or indicating the subject of the verb (कर्ता)
    2. accusative – serving as or indicating the (direct) object of the verb (कर्म)
    3. instrumental – serving or acting as a means or aid (साधन, करण)
    4. dative – serving as the (indirect) object or the recipient (beneficiary) of the action of the verb (सम्प्रदानम्)
    5. ablative – indicating the source or separation of the agent, instrument, or location (अपादानम्)
    6. genitive – expressing ownership (–)
    7. locative – designating the place or state or action denoted by the verb (अधिकरणम्)
    8. vocative – identifying the person being addressed (सम्बोधनम्)

Here is the declension of a masculine अकारान्त word देव :

The verbal system is more complex and in the vedic system it is even more so [than the classical system]. The book describes complexities of the vedic verbal system and mentions that classical verbal system gradually got rid of a lot of constructs from the former.

Here is an overview of the complexities of the grammar of the vedic language:

1. Verb roots (धातवः) are generally of two types: athematic and thematic.

Singular (एकवचनम्)	Dual (द्विवचनम्)	Plural (बहुवचनम्)	
देवः	देवौ	देवाः	प्रथमा
देवं (देवम्)	देवौ	देवान्	द्वितीया
देवेन	देवाभ्यां (देवाभ्याम्)	देवैः	तृतीया
देवाय	देवाभ्यां (देवाभ्याम्)	देवेभ्यः	चतुर्थी
देवात्	देवाभ्यां (देवाभ्याम्)	देवेभ्यः	पञ्चमी
देवस्य	देवयोः	देवानां (देवानाम्)	षष्ठी
देवे	देवयोः	देवेषु	सप्तमी
हे देव	हे देवौ	हे देवाः	सम्बोधनम्

(a) Athematic: Variable accent[1] and variable stem to which terminations are directly attached (e.g. अस् + ति = अस्ति)

(b) Thematic: Invariable accent and invariable stem followed by the thematic vowel अ before the termination (e.g. बुध् + अ + ति = बोधति). This helps make various verb forms more regular as the complex interactions between the root-final consonant and suffix-initial vowel are prevented by the intervening अ .

2. A verb has a single root, however, in Vedic Samskritam it may have many stems[2]. Consider the following forms (root, stem, affixes) with respect to a single verbal root गम् which means "to go". Forms depend upon voice, number, and person of the agent (कर्ता), but for brevity, we show the form only[3] for the singular (एकवचनम्), active (कर्तरि) voice, and third person (प्रथम पुरुषः):

(a) Tenses:

i. Present (लट् – गच्छति)
ii. Aorist[4] (expressing action (especially past action) without indicating its completion or continuation) (लुङ् – अगमत्)
iii. Imperfect Past (लङ् – अगच्छत्)
iv. Perfect (लिट् – जगाम)
v. Future (ऌट् – गमिष्यति)

(b) Moods:

i. Indicative (simple declarations) (लट् – गच्छति)
ii. Injunctive (could indicate intention) (जिगमत्)
iii. Subjunctive (typically used to express various states of reality such as wish, emotion, judgment etc.)
iv. Optative (wish, request, or command) (गच्छेत्)

[1] Accent or Stress in linguistics is the relative emphasis or prominence given to a certain syllable (unit of speech sounds) in a word or to a certain word in a phrase or sentence
[2] A stem is that part of a word that inflectional affixes attach to
[3] See here for a programmatic interface to all the forms of गम्
[4] There are several aorist tense forms in Samskritam

v. Imperative (command or request) (गच्छतु)

(c) Meanings of the present tense stems:

i. Indicative (simple declarations)(लट् – गच्छति)

ii. Intensive (denotes stronger, more forceful, or more concentrated action relative to the root on which the intensive is built) (जङ्गम्यते)

(d) Participle forms indicating tenses and voices (these forms depend on the gender of the nominal they modify):

i. Past Passive (गत, गता)

ii. Past Active (गतवत्, गतवती)

iii. Present Active (गच्छत्, गच्छन्ती)

iv. Present Middle (गच्छमान, गच्छमाना)

v. Present Passive (गम्यमान, गम्यमाना)

vi. ...

3. Verb conjugations could also be one of

(a) Primary

(b) Causative

(c) Intensive

(d) Desiderative (a verb formed from another and expressing a desire to do the act denoted by the root verb)

4. All of this clearly means that the number of derived forms was very large. See Table 8 for a more complete list of verb conjugations and related forms.

The verbal form complexity was greatly reduced in the classical language. The so-called लकाराः show a glimpse of a plethora of forms that a verbal root leads to:

लट् वर्तमाने लेट् वेदे भूते लङ् लुङ् लिटस्तथा ।
विध्याशिषोस्तु लिङ्लोटौ लुट् लृट् लृङ् च भविष्यति ॥

The language evolved to favor nominal sentences over verbal sentences.

The most remarkable feature of the classical language is the compounds (especially their phenomenal length). Here is an example from Jayadeva's गीतगोविन्द :

चन्दनचर्चितनीलकलेवरपीतवसनवनमाली।
केलिचलन्मणिकुण्डलमण्डित गण्डयुगः स्मितशाली॥

The author believes that several changes occurred to the vedic Sanskrit that Panini grammarized. There was also the influence of local languages. In spite of that, because of Panini's efforts, the language established itself as an "elite language". In such evolution, the language's surface forms[5] were retained.

[5]Surface form of a word is the form of a word as it appears in the text (e.g. "goes" is a surface form of the verb "go"). Contrast it with the lexical form which consists of things such as the root, the part of speech etc.

2 The संस्कृत Alphabet

2.1 Basics

2.1.1 Vowels (when not combined with consonants)

There are 13 vowels of which 5 are short (ह्रस्व) and 8 are long (दीर्घ). Not combined with consonants, here they are:

Table 1: Sanskrit Vowels

अ	आ
इ	ई
उ	ऊ
ऋ	ॠ
ऌ	
ए	ऐ
ओ	औ

In addition to these 13 independent vowels, there are two "add-on signs" that, like other vowels, conjugate with consonants:

1. ं – अनुस्वारः
2. ः – विसर्गः

One last vowel-like letter is "अवग्रह" which elongates the pronunciation of a preceding vowel: ऽ . The independent vowels conjugate with consonants and create "add-on signs" that we will see below.

There are 33 consonants and 2 special consonant clusters. The arrangement is according to the location and mechanism of sound production:

Table 2: Sanskrit Consonants

Velar (Guttral) (कण्ठ्य)	क	ख	ग	घ	ङ
Palatal (मूर्धन्य)	च	छ	ज	झ	ञ
Cerebral (तालव्य)	ट	ठ	ड	ढ	ण
Dental (दन्त्य)	त	थ	द	ध	न
Labial (ओष्ठ्य)	प	फ	ब	भ	म
Semivowels	य	र	ल	व	
Sibilants (hissing sound)	श	ष	स		
Aspirate (rush of air)	ह				
Special consonant clusters	क्ष	ज्ञ			

Table 3 shows the phonetic analysis of the Sanskrit Alphabet. In that table, 'Voice' feature refers to sonorous vibration, 'Asp' (aspiration) feature refers to a rush of air, and 'Nasal' refers to sound generation with the passage of air through nose (and not mouth[6]). If a feature is present, a '+' preceeds it (e.g. '+Voice' – generation of a sonorous vibration) and if a '-' preceeds it, that feature is absent (e.g. '-Asp' – no rush of air).

Figure 1 (taken from Wikipedia) shows the places where the sound is produced in its audible form[7].

We have already seen the Sanskrit alphabet in Tables 1 and 2. In the written form, when one consonant conjugates with one vowel, we see an alternate representation. For example, when क combines with इ , the resulting conjugate is written as कि and not क.इ .

These are collectively called the "add-on vowel signs" and they look like below (note that each of them has a Unicode codepoint which means that it is recognized as a distinct character in a writing system (in this case देवनागरी)). There is no add-on vowel sign for अ ; when conjugated with अ , the consonant simply loses the विरामः – ◌् . Any consonant should conjugate with these add-ons, the Table 4 below uses ' त् ' as an example.

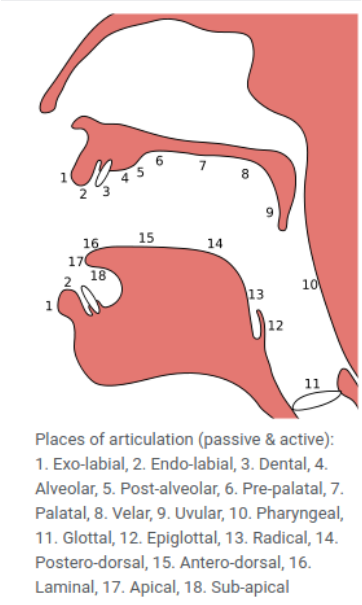
Thus the representation of a single consonant like प is a combination of the base consonant प् and the vowel अ . This document treats प and प् interchangeably

[6]Nearly all nasal consonants are nasal occlusives, in which air escapes through the nose but not through the mouth, as it is blocked (occluded) by the lips or tongue.

[7]Articulation means the aspect of pronunciation that involves bringing articulatory (speech) organs together so as to shape the sounds of speech.

Table 3: Phonetic Analysis

Places of Articulation	Stop -Voice -Asp	Stop -Voice +Asp	Stop +Voice -Asp	Stop +Voice +Asp	Stop +Voice -Asp +Nasal	Semi-vowels +Voice -Asp	Sibilants +Voice +Asp	Sibilants -Voice +Asp
कण्ठ्य	क	ख	ग	घ	ङ		ह	ः
मूर्धन्य	च	छ	ज	झ	ञ	य अनुनासिक य		श
तालव्य	ट	ठ	ड	ढ	ण	र		ष
दन्त्य	त	थ	द	ध	न	ल अनुनासिक ल		स
ओष्ठ्य	प	फ	ब	भ	म	व अनुनासिक व		



Places of articulation (passive & active): 1. Exo-labial, 2. Endo-labial, 3. Dental, 4. Alveolar, 5. Post-alveolar, 6. Pre-palatal, 7. Palatal, 8. Velar, 9. Uvular, 10. Pharyngeal, 11. Glottal, 12. Epiglottal, 13. Radical, 14. Postero-dorsal, 15. Antero-dorsal, 16. Laminal, 17. Apical, 18. Sub-apical

Figure 1: Places of Articulation

unless noted otherwise. The word कारः in Samskritam grammar denotes the sound associated with a letter or (occassionally,) a word. For example, when the vowel आ combines with any consonant an आकारः results, when the vowel उ combines with any consonant an उकारः results, and so on.

Table 4: Add-on Vowel Signs with the consonant ‘ त् ’

Add-on sign codepoint (decimal)	Add-on sign	Its conjugation with ‘ त् ’	Corresponding vowel
	–	त	अ
2366	ा	ता	आ
2367	ि	ति	इ
2368	ी	ती	ई
2369	ु	तु	उ
2370	ू	तू	ऊ
2371	ृ	तृ	ऋ
2372	ॄ	तॄ	ॠ
2402	ॢ	तॢ	ऌ
2375	े	ते	ए
2376	ै	तै	ऐ
2379	ो	तो	ओ
2380	ौ	तौ	औ
2306	ं	तं	अं
2307	ः	तः	अः

Consonants combine with the vowels to form the conjugate forms as in Table 4. Thus, with another consonant, ग् , the forms will be written: ग, गा, गि, गी, गु, गू, गृ, गॄ, गॢ, गे, गै, गो, गौ, गं, गः . There are few exceptions in writing these forms:

- र् + उ = रु
- र् + ऊ = रू
- र् + ऋ = र्ऋ
- र् + ॠ = र्ॠ
- ह् + ऋ = ह्ऋ

When two or more consonants occur successively without any intervening vowels, they are written in a conjoined form:

- Horizontal cluster (read left → right): च्चे
- Vertical cluster (read top → bottom): ह्↓

The rules for making consonant clusters depend (mostly) on whether the first consonant has a vertical line from top to bottom or a short central stem from which character is suspended:

1. In a horizontal cluster, the final vertical lines of all but the last consonant are dropped and then the remaining parts are joined together: व् + य = व्य , ष् + प् + य = ष्प्य , स् + त् + वा = स्त्वा

2. Characters suspended from a central stem have several forms: ट् + ट = ट्ट, द् + द = द्द, ड् + ड = ड्ड, ल् + ल = ल्ल

3. Other combinations are handled in various ways: ट् + य = ट्य, द् + य = द्य, क् + य = क्य, क् + त = क्त . When ल् is the second consonant, it may be placed next to (ग् + ल = ग्ल) or below (क् + ल = क्ल) the first consonant.

4. Special forms: श् and र्

   (a) For श् , alternate combining form may be used: श्च, श्न, श्र, श्व etc. or श् may be retained: श्क, श्त, श्म etc. But with र it is always श्र , never as श्र.

   (b) र् has several combining forms.

      i. When र् is the second consonant:

         A. First consonant has a vertical line: प् + र = प्र, क् + र = क्र, ब् + र = ब्र , but there are special forms: त् + र = त्र, क् + र = क्र.

         B. First consonant has a central stem: ट् + र = ट्र, ड् + र = ड्र etc. Note the exception: द् + र = द्र .

      ii. When र् is the first consonant: This gives rise to रेफः . Examples: र् + क = र्क, र् + म = र्म, र् + को = र्को .

There are defined consonant clusters (different from the special consonant clusters क् + ष = क्ष, ज् + ञ = ज्ञ that are part of the alphabet) like: क्क, क्ख, क्च, क्ण, ..., क्ष्म, ङ्क्, ङ्ख्, ..., द्र, द्ध, द्व, द्म, द्य्य . Upon careful observation, one can see several similarities and some anomalies in the way letters of the देवनागरी script are written. One reason is, of course, evolution of the language and the script due to socioeconomic circumstances (e.g. ease of writing). Still, one generally makes no mistake in identifying a piece of text written in some script. In other words, how humans identify a certain script and letters in it is a matter of scientific research. Describing unambiguously how to write a symbol of a script is challenging. Think of teaching a computer how to write, for example, क and you will perhaps appreciate the difficulty involved.

Generally speaking, a document (like this one) with देवनागरी letters that has been typeset on a computer (or in a traditional printing press), the letters (especially the combined ones) are dictated by the font used. Thus, all of the above letters that you see are defined by the artist(s) who created the font. The देवनागरी letters in this document use a font named Shobhika(शोभिका) which is developed by the Indian Institute of Technology Bombay, Mumbai, India.

There aren't very many punctuation marks in Samskritam. The few that are used consistently are:

1. विरामः – ् that denotes a bare consonant (without the terminal vowel अ).
2. दण्डः – । that marks the end of a sentence or the end of half a stanza: रामः गृहं गच्छति।
3. दण्डौ – ॥ that marks the end of a stanza.

According to Prof. Deshpande, in modern editions of Sanskrit texts, other English punctuation marks like ‘,’, ‘!’, and ‘?’ are found.

2.2 On Transliteration

Perhaps of all the creations of man language is the most astonishing.

–Giles Lytton Strachey [1]

Associated with every human language are sounds and symbols. A distinguishable sound is called phoneme whereas a distinguishable symbol is called grapheme or character. Every language has a script which is used to represent the characters and words in that language. The basic set of symbols of a script is its alphabet. English words, for instance, are represented by a sequence of symbols of the augmented English alphabet. The alphabet is augmented to include punctuation signs, numbers, and other symbols. These symbols are collectively referred to as the “character set” of a writing system.

A sequence of English characters (from ‘a’ to ‘z’ and from ‘A’ to ‘Z’) to represent a word (e.g. “student”) is called the spelling of that word. Spelling or native spelling of a word tells us how to write the sounds of any word in a language using the (native) script of that language so that the written communication can become possible.

Transliteration refers to conversion of graphemes or characters or symbols from one script to another. Although the focus of transliteration is on symbols (and not on sounds), sounds are given due consideration and symbols that sound similar symbols are chosen for conversion. For example, the देवनागरी symbol ‘र’ is typically transliterated to the English symbol ‘r’.

In general, as shown in Figure 2, transliteration refers to conversion of text from a source script to a target script. Thus, in transliteration, two different scripts or “character sets” are involved: a source character set and a target character set. Typically, these sets are disjoint, that is, there is no character (excluding trivial characters like a blank – ‘ ’) that is a member of both the character sets. It should be clear that we undertake transliteration so that you can (at least) recognize certain text that was written in an unfamiliar character set. By reading the transliterated text, you can reproduce the sounds that the source text represents, **without knowing the symbols from the source character set**.

You may ask a question: “Why should transliteration be attempted at all?” It is indeed possible to take an approach like “The native script of any text is inseparable from its language, therefore, learn the native script to understand how that text is written and read” and altogether remove the need for transliteration[8], but such an

[8] A similar argument could also be made about the need for translating texts from one language into another

approach is too restrictive; transliteration facility may provide people with additional means to analyze languages and their scripts. After all, transliteration between two scripts is expected to be simpler than translation between associated languages.

We can examine the process of transliteration in more detail. Suppose you want to read the Samskritam sentence in देवनागरी – रामः वनम् अगच्छत्। – but you have not learned the देवनागरी script and you are only familiar with the English alphabet and a few more characters like period – '.', blank or space – ' ', colon – ':' etc. How might you read that sentence then?

One way to do this may be to split the sentence – let's call it the input sentence – into words at word boundaries. Each word in the sequence of words is, in turn, a sequence of characters in the source character set. Once such "character segregation" of the input sentence is done, we can create an unambiguous, one to one mapping from the source character to one or more target characters. Such a system, shown in Figure 2, is a simple, yet functional one-way interpreter of symbols. A one to one mapping means that every symbol from the source character set is associated with a unique sequence of symbols from the target character set. An illustration of such a mapping is shown in Figure 3. Our system may maintain a table of symbol mappings and simply look up the target symbol(s) for the source symbol that it encounters in the input sentence. Once the corresponding characters are emitted by our interpreter, the resulting sentence – let's call it the output sentence – will be automatically constructed.

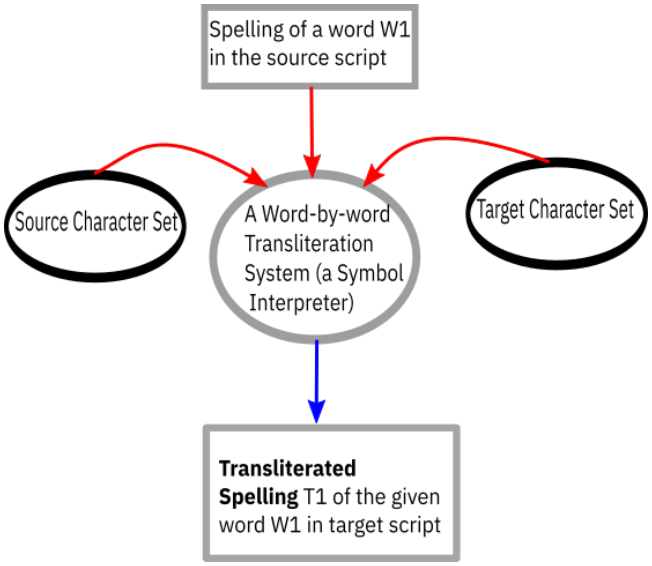


Figure 2: A One-way Transliterator

Let us examine this process for our example sentence:

- Input sentence is रामः वनम् अगच्छत्। . It is made of exactly 18 characters in the following order: 'र', ा, 'म', 'ः',' ', 'व', 'न', 'म', 'ँ', ' ', 'अ', 'ग', 'च', 'ँ', 'छ', 'त', 'ँ', '।'
- Inspired by Figure 3 we devise a mapping in Table 6 that we look up as we deal with the input characters.

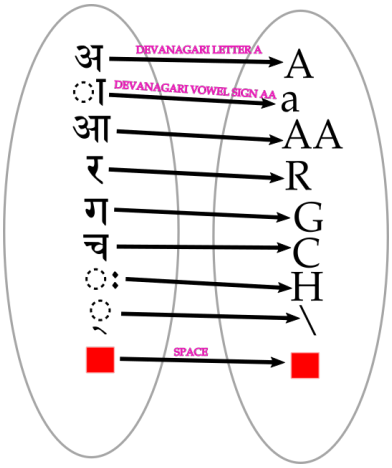


Figure 3: One to One Mapping Between Example Character Sets

- Output sentence then becomes: raaamaH vanam\ Agacha\chhat\|

Table 6: One to One Mappings for the Running Example

Input Symbol	Output Symbol
र	ra
ा	aa
म	ma
ः	H
(blank)	(blank)
व	va
न	na
म	ma
ॅ	\
अ	A
ग	ga
च	cha
छ	chha
त	ta
।	|

A Verb Conjugation Names

Naming verb conjugations is typically challenging. We found a verb conjugation engine that is indeed very useful in this regard. Table 8 shows the देवनागरी names of various verb forms (including conjugations) and their corresponding English names.

The classical language misses some conjugations present in the vedic language.

References

[1] Giles Lytton Strachey and George Rylands. Words and Poetry. New York: Payson and Clarke. Introduction, Page xii.

[9] Benediction: Indicating prayer or invoking divine protection

[10] परस्मैपदी धातुः

[11] आत्मनेपदी धातुः

Table 8: Names of Verb Conjugations

देवनागरी नाम	English Name
तिङन्ताः	Conjugations
अप्रत्ययान्तधातुः	Primary Conjugation
लट् (लकारः)	Present (Tense)
लङ् (लकारः)	Imperfect (Past Tense)
विधिलिङ् (लकारः)	Optative (Mood)
लोट् (लकारः)	Imperative (Mood)
लृट् (लकारः)	Future (Tense)
लृङ् (लकारः)	Conditional (Future Tense)
लुट् (लकारः)	Periphrastic (Future Tense)
लिट् (लकारः)	Perfect (Past Tense)
लुङ् (लकारः)	Aorist (Tense)
आगमाभावयुक्तलुङ् (लकारः)	Injunctive (Mood)
आशिर्लिङ् (लकारः)	Benedictive[9] (Mood)
कृदन्तः (प्रत्ययः)	Participle
क्त (प्रत्ययः)	Past Passive (Participle)
क्तवतु (प्रत्ययः)	Past Active (Participle)
शतृ (प्रत्ययः)	Present Active (Participle)
शानच् कर्मणि (प्रत्ययः)	Present Passive (Participle)
लुडादेश (लुट् + आदेश) पर[10] (प्रत्ययः)	Future Active (Participle)
तव्य (प्रत्ययः)	Future Passive (Participle)
यत् (प्रत्ययः)	Future Passive (Participle)
अनीयर् (प्रत्ययः)	Future Passive (Participle)
यत् (प्रत्ययः)	Future Passive (Participle)
लिडादेश (लिट् + आदेश) पर (प्रत्ययः)	Perfect Active (Participle)
लिडादेश (लिट् + आदेश) आत्म[11] (प्रत्ययः)	Future Passive (Participle)
अव्ययः	Indeclinable
तुमुन् (अव्ययः)	Infinitive
क्त्वा (अव्ययः)	Absolutive (gerund)
क्त्वा (अव्ययः)	Absolutive (gerund)
ल्यप् (अव्ययः)	Absolutive (gerund)
ल्यप् (अव्ययः)	Absolutive (gerund)
णिच् रूपाणि	Causative Conjugation (repeat some of the above)
यङ् रूपाणि	Intensive Conjugation (repeat some of the above)
सन् रूपाणि	Desiderative Conjugation (repeat some of the above)