Chapter 3

Frameworks

The corpus of Aṣṭādhyāyī formulates the grammatical system specified in the previous chapter in a special framework. I will mention some of the salient features of the Pāṇinian formulation in the following section. The present chapter, however, introduces a new `formal framework` which facilitates a *re-presentation* of the Aṣṭādhyāyī. The framework I propose is different from the one in which the Aṣṭādhyāyī is formulated.

The most significant difference between Pāṇinian and formal representation is that of the medium. Without entering into the debate about whether script was known to Pāṇini or whether he made use of it when compiling his grammar, based on the manner in which the grammar is formulated—e.g. the use of accents or phonemes as markers—it can be asserted that Aṣṭādhyāyī is composed, transmitted and applied in an oral framework. Moreover, it is composed in Sanskrit, although it employs a number of special meta-linguistic conventions. A formal representation, on the other hand, is meant to render it in terms of logical relations that can be implemented on a computer.

The reason for proposing a new framework is that the Pāṇinian formulation, although special on several counts, is not adequately formal for a direct implementation on computer. Apart from its oral and verbal character, Aṣṭādhyāyī is meant to be used by a person who has acquired skills to apply it through the study of grammar. The rules of grammar are not listed in the order in which they are to be applied for derivation of any linguistic expression. Instead, rules having common contexts are grouped together to avoid repeated and redundant enunciation of these contexts. A formal representation, however, is meant to be applied by a computer program with the aid of a user. For this, a mechanism needs to be developed by which the required information scattered in the grammatical corpus can be collected together using various indices encoded in a complex meta-language.

An important challenge in Pāṇinian formulation is the presence of ambiguities, inconsistencies and the incomplete nature of the Aṣṭādhyāyī. I will discuss some of these issues in section 3.2. While the commentaries and other explanations available in the later grammatical literature help in clarifying them, a device is still needed to incorporate these explanations and suggestions. Instead of directly modeling the corpus of Aṣṭādhyāyī, the present approach models the categories and processes outlined in the previous chapter about the grammatical system. It also facilitates incorporation of information necessary for application of the grammar but not explicitly mentioned in the grammatical corpus.

3.1 Salient features of the Pāṇinian formulation

The formulation of the grammatical system in the Aṣṭādhyāyī of Pāṇini is in the form of concise statements—the *sūtra*s.[1] Typically, a *sūtra* consists of one or more inflected expressions. In the following, I shall call them `elements` of a *sūtra*. These elements are inflected just like any other expression in Sanskrit. For example, in a *sūtra* like *veḥ pādaviharaṇe*[2] the element *veḥ* is the ablative singular of vi and *pādaviharaṇe* is locative singular of *pādaviharaṇa*. Pāṇini's description of Sanskrit, therefore, is in Sanskrit.

Yet not all the elements within the grammatical corpus belong to the common language. There are many elements that appear only in the grammatical corpus. In the above *sūtra*, the second element *pādaviharaṇe* belongs to the common speech and the first one *veḥ* is an entity which exists only within the grammatical system. There is thus a clear distinction between the object language which is being described i.e. Sanskrit and the language of the Aṣṭādhyāyī, which is a special language of description. From Pāṇini onwards the tradition clearly recognizes two types of languages: the perennial utterances of the Vedas and the established language of everyday communication on the one hand and on the other hand the artificial language of grammatical instructions. The language of grammar is for the specific purpose of providing a description of the standard usage and is artificially created by the grammarians. In this sense, Pāṇini is not the creator (*kartṛ*) but the one who re-collects (*smartṛ*) the object language.[3]

---

[1] Louis Renou (1963 p. 165-216) provides a detailed survey on the genre of *sūtra* in Sanskrit literature. See also (Staal 1992 p. 303-314) and (Wezler 2001a p. 351-366).

[2] वेः पादविहरणे ॥ १.३.४१ ॥ ▶ ātmanepada suffixes are attached after kram(u) if it comes together with the preverb vi to express nice strides (*pādaviharaṇa*).

[3] A detailed discussion of the traditional points of view on this topic is provided by R. S. Bhattacarya (1966 p. 212-237). Hartmut Scharfe (1971 p. 2-6) also gives a summary of the main arguments.

The artificiality of the language of Aṣṭādhyāyī also lies in the fact that it follows only partially the rules of grammar that it lays down. In many cases it is guided by a special set of rules, namely the meta-linguistic conventions. For example, although the elements of the grammatical corpus consist of inflectional suffixes like any other linguistic expression, the meaning which they denote is different compared to what they express in common usage. While the rules under kāraka-section relate to common usage,[4] in case of elements of the grammatical corpus several meta-linguistic conventions provide for other denotations. Thus, for example, the convention *ṣaṣṭhī sthāneyogā*[5] specifies that the sixth case (ṣaṣṭhī) or the genetive case-ending denotes the place where a substitution takes place (i.e. the substituendum). The right-context is expressed through locative endings and this is specified by the meta-rule *tasminnitinirdiṣṭe pūrvasya*[6]. Similarly, the convention *tasmādityuttarasya*[7] specifies that the left-context is denoted through ablative endings.[8]

Considering the special meta-linguistic conventions followed in the formulation of the grammatical corpus, and the fact that several of these conventions are explicitly mentioned in the same corpus, the statements of Aṣṭādhyāyī can be grouped into two basic types:

1. Operational rules that introduce the grammatical components, characterize them and provide for their combinations.

2. Meta-linguistic conventions about the formulation of the operational rules.

The first group deals with what is to be done and the second one with how it is presented or formulated in the Aṣṭādhyāyī.

In the previous chapter, I introduced three kinds of constituents of grammar: (i) the components of linguistic expressions (ii) the attributes that characterize them and (iii) the meaning-expressions specifying the semantic information. As mentioned before, within the grammatical corpus, these constituents occur in an inflected form. The combination of the inflectional suffix with the base follows the normal rules of suffix additions and phoneme combinations

---

[4] The section on kāraka is from 1.4.23-1.4.55. For a detailed discussion on the kāraka's see (Cardona 1974) and the Kārakāhnika of the Mahābhāṣya (Joshi and Roodbergen 1975).

[5] षष्ठी स्थानेयोगा ॥१.१.४९॥ ► ṣaṣṭhī stands for 'introduction in place of'.

[6] तस्मिन्नितिनिर्दिष्टे पूर्वस्य ॥१.१.६६॥ ► locative case indicates that the grammatical operation is to be applied to the immediately preceding component.

[7] तस्मादित्युत्तरस्य ॥१.१.६७॥ ► ablative case indicates that the grammatical operation is to be applied to the immediately following component.

[8] It should be noted that the meta-linguistic conventions are not universally applied in a consistent and consequent manner throughout the Aṣṭādhyāyī. See the discussion in section 3.2.

laid down in the Aṣṭādhyāyī. For example, the attribute vṛddhi is stated as
vṛddhiḥ in the rule *vṛddhirādaic*[9]. Similarly, the meaning-expression *svatantraḥ*
in rule *svatantraḥ karttā*[10] is mentioned together with the inflectional suffix
for nominative singular.

Rules of phoneme combination and compounding are applied as well.
The components have a special presentational scheme. While the meaning-
expressions and many (but not all) attributes are from the natural language,
the components do not belong to the common usage. Their artificial character
is reinforced by the use of marker-sounds that are often attached to them
either at the beginning or at the end. Thus, in a component like ti(p), the final
sound p is a marker. In this book, markers are pointed out by including them
within brackets ( ). Similarly, the component (ś)a(p) has a marker ś attached
at the beginning and another marker p at the end.

Apart from markers, the components sometimes have a vowel which is
needed just for facilitating its pronunciation. Thus, n[u](m) consists of the
vowel u which is not part of the component, but is added in order to be able
to pronounce this component. Such extra vowels are noted within square
brackets [ ] in this book.

3.1.1 Methodology of Pāṇinian formulation

The elements of the grammatical corpus are threaded within a number of
inter-connected *sūtra*s (lit. a string). A *sūtra* constitutes a distinguishable
unit of the grammatical corpus with a specific function.[11] Although *sūtra*
is a distinguishable functional component, it is not always an independent
unit and in most of the cases several *sūtra*s must be read together in order
to execute a grammatical operation. A significant Pāṇinian device which is
employed for this purpose is the carrying over of elements from previous
to subsequent rules (*anuvṛtti*).[12] Moreover, many rules operate within the
domain set by what are known as *adhikāra* rules. In this manner repetitions
are avoided and enunciation of an element at one place suffices for its use

---

[9] वृद्धिरादैच् ॥१.१.१॥ ▶ vṛddhi stands for ā*t* and ai*c*.

[10] स्वतन्त्रः कर्त्ता ॥१.४.५४॥ ▶ that kāraka which is *svatantra* or independent of others is kartṛ.

[11] Tradition speaks of six kinds of *sūtra*s based on their function: *sañjñā* (definition), *paribhāṣā*
(convention), *vidhi* (injunction), *niyama* (restriction), *pratiṣedha* (prohibition) and *adhikāra* (ex-
panse). See (Abhyankar 1974 p. 432).

[12] S. D. Joshi and Saroja Bhate (1984) have comprehensively examined the principles behind
such an organization of the grammatical corpus in a monograph titled "The fundamentals
of *anuvṛtti*".

at other parts of the corpus.[13] The method of *anuvṛtti* functions at the level of grammatical elements. In most of the cases it is only a part of the *sūtra* which is carried over to the subsequent rules. Consider for example the rule *kṛnmejantaḥ*[14], which says that those expressions that end in kṛt suffixes having m or e*c* at the end, are termed avyaya. The term avyaya here is to be taken from the rule *svarādinipātamavyayam*[15] where it is first introduced. So a complete reading would be *kṛnmejantaḥ avyayam*, where the term avyaya is collected from a foregoing rule. Not the entire rule, but only a part of it is carried further.[16]

Since application of a grammatical operation requires elements from more than one *sūtra*s to be collected together, it becomes necessary to specify the boundaries of elements within the grammatical corpus. In fact, grammatical corpus can be considered as a sequence of elements where *sūtra*-boundary is marked as well.

The conciseness of a *sūtra* is firstly because as a unit it is not always a complete and self-contained expression of what is to be said, but only a part of the whole statement. It needs to be completed by incorporating other information. The information lacking may be present in some other part of the corpus and must be fetched to make a complete and applicable statement. Sometimes the missing information is to be provided through conventions and interpretations not explicitly mentioned in the grammatical corpus. A *sūtra* therefore is a unit within an inter-dependent and inter-related network.

The organization of the grammatical corpus in terms of the *sūtra*s consisting of one or more elements, points towards the fact that the formulation of the grammatical corpus follows the same methodology as the grammar itself. As specified in the previous chapter, this approach is to analyze a given whole into constituent components and then provide a rule-based combination of the components to specify the whole. In the case of the corpus of grammar, the *sūtra*s (or parts thereof) are components that need to be combined with other components (i.e. with other *sūtra*s or parts thereof) in order to fetch an applicable complete rule of grammar. How the different components of an applicable rule are to be gathered together is guided by

---

[13] H. V. Nagarajarao (1978 p. 145-176) discusses the scope and necessity of the *adhikāra* of aṅga although Pāṇini does not explicitly states its domain. The meaning-*adhikāra*s in the taddhita section of the Aṣṭādhyāyī are analyzed by Saroja Bhate (1987 p. 81-92). Ashwini Deo (2007 p. 1-37) shows that the taddhita section of the Aṣṭādhyāyī is structured as a default inheritance hierarchy.

[14] कृन्मेजन्तः ॥१.१.३९॥ ► expressions that end with those kṛt suffixes which end in m or e*c* are avyaya.

[15] स्वरादिनिपातमव्ययम् ॥१.१.३७॥ ► svarādi (svar etc.) and nipāta are avyaya.

[16] In his six volume edition of the Aṣṭādhyāyī, Rama Nath Sharma (1990 p. ix-x) has provided *anuvṛtti* elements from preceding *sūtra*s.

a number of meta-linguistic conventions, as well as several interpretations
supplied by the later Pāṇinīyas.[17]

As mentioned in the previous chapter, grammar comprehends the stan-
dard usage by analyzing and identifying the constituent components and
then providing a rule-based combination of them. Similarly, the formulation
of the grammar is achieved in terms of *sūtra*s (including their constituent
elements) and together with conventions to synthesize the operational
statements.

The purpose of opting for the same methodology is also similar, namely
to provide a compact encoding of the grammatical information as well as
to safeguard it against future loss or corruption. Several techniques are
employed to this effect. The most important method is to enunciate first the
general characteristics and then to specify the exceptions.[18] Components that
share common grammatical characteristics are linked together. Two such
groups are relatively large and enumerated separately. These are the list of
verbal roots (Dhātupāṭha) and of selected nominal stems (Gaṇapāṭha). There
are several sub-groups within these groups and they are frequently referred
to in the main corpus.[19] Formation of groups is attained using marker
sounds as delimiters which facilitate specification of sigla (*pratyāhāra*).[20] The
it-markers as indicators that are directly attached to components serve to
form groups as well.

Meta-linguistic information is also encoded in the intonation of the com-
ponents and *sūtra*s. Thus, according to the convention *svaritenādhikāraḥ*[21],
the circumflex accent (svarita) indicates that the particular *sūtra* (or a part
of it) is a heading rule. Similarly, the verbal roots are specified with accent
markers which aid in deciding the kind of suffixes that should be attached
to them. For example, following the rule *anudāttaṅita ātmanepadam*[22], the
ātmanepada suffixes are attached if the verbal roots are marked by a low
pitched (anudātta) marker vowel.

Pāṇini also uses the principle of correspondence and instead of men-
tioning individual pairs he correlates collectively two lists of equal length.

---

[17] Despite copious literature on this subject the task is by no means finished. For example,
there are cases where scholars have divergent opinions as to the boundary of a *sūtra*.

[18] Patañjali notes this in (P. 53), see (Joshi and Roodbergen 1986 p. 77).

[19] For example, the ajādi group is referred to in the rule अजाद्यतष्टाप् ॥४.१.४॥ ► after ajādi
components and those ending in a*t*, suffix (ṭ)ā(p) is introduced to denote feminine. See the
Gaṇapāṭha in (Katre 1989 p. 1265-1325).

[20] See section 4.1.1.

[21] स्वरितेनाधिकारः ॥१.३.११॥ ► through svarita a domain is marked.

[22] अनुदात्तङित आत्मनेपदम् ॥१.३.१२॥ ► after anudāttet or ṅit components ātmanepada suffixes
are introduced.

The convention *yathāsaṃkhyamanudeśaḥ samānām*[23] specifies this. The rule *iko yaṇaci*[24] is an example. Here i*k* = [i, u, ṛ, ḷ] and ya*ṇ* = [y, v, r, l] are correlated respectively.

For the specification of it-markers, Pāṇini makes use of rules.[25] More-over, the names of the sets of phonemes (*pratyāhāra*) are generated in a rule-based manner. The convention *ādirantyena sahetā*[26] is used to coin the names of the collections in a generic manner. Thus, Pāṇini not only attains brevity by using the acronyms or *pratyāhāra* instead of using the entire list every time, but he also generates their names using just one rule.

The employment of several of the techniques mentioned above to achieve maximum compactness has the consequence that the grammatical corpus attains a complex structure. This requires commentaries and explanations to make explicit the encoded information in an unambiguous manner. The task of formalization of the grammar needs therefore to take this aspect into consideration.

The formulation of the grammatical system does not furnish in an ex-plicit manner the process of derivation which involves operations that are carried out one after the other in a sequential manner. These operations are conditioned. To execute this process, one would ideally expect that the next applicable operation is unambiguously specified. It means that there is a unique operation and all the conditions that are to be satisfied are also available. A *sūtra*, however, is not always prescribing a unique operation and seldom does it have the complete set of conditions mentioned at one place. For example, the *sūtra*: *kartuḥ kyaṅ salopaśca*[27] provides for two operations: introduction of the suffix (k)ya(ṅ) and elision of the final phoneme s. For the conditions other *sūtra*s like *dhātoḥ karmaṇaḥ samānakartṛkādicchāyāṃ vā*[28] and *supa ātmanaḥ kyac*[29] etc. need to be taken into account.[30]

---

[23] यथासंख्यमनुदेशः समानाम् ॥१.३.१०॥ ► respective assignment for equal number of elements.

[24] इको यणचि ॥६.१.७७॥ ► i*k* is replaced by ya*ṇ* if it is followed by a*c*.

[25] The rules for it markers are: उपदेशेऽजनुनासिक इत् ॥१.३.२॥ ► in grammatical instruction, an a*c* which is anunāsika is it. हलन्त्यम् ॥१.३.३॥ ► ha*l* coming at the end. न विभक्तौ तुस्माः ॥१.३.४॥ ► but not in vibhakti, the t*u*, s or m. आदिर्ञिटुडवः ॥१.३.५॥ ► the initial ñi, ṭu and ḍu (of dhātu). षः प्रत्ययस्य ॥१.३.६॥ ► ṣ of pratyaya. चुटू ॥१.३.७॥ ► c*u* or ṭ*u*. लशक्वतद्धिते ॥१.३.८॥ ► l, ś and k*u* of non taddhita.

[26] आदिरन्त्येन सहेता ॥१.१.७१॥ ► an initial element together with the final it sound includes intervening elements.

[27] कर्तुः क्वङ् सलोपश्च ॥३.१.११॥ ► after kartṛ (k)ya(ṅ) is introduced and the final s is elided (lopa).

[28] धातोः कर्मणः समानकर्तृकादिच्छायां वा ॥३.१.७॥ ► after dhātu optionally in case its action is karman of some dhātu expressing desire and both dhātu have same kartṛ.

[29] सुप आत्मनः क्यच् ॥३.१.८॥ ► after s*up* to express desire for oneself, (k)ya(c) is introduced.

[30] For details see (R. N. Sharma 1987 p. 46).

Here it is important to note that it is not that the grammar does not specify the conditions, but only that the specification through the *sūtra*-style is such that these are distributed at different locations and need to be gathered at the time of application. This process of bringing all the requisite information together in one place is carried out by a human user on the basis of her or his knowledge of the Aṣṭādhyāyī, especially the knowledge of its meta-linguistic conventions.

In an article published in 1975, Rama Nath Sharma (1975 p. 31-39) discusses this aspect. He suggests that there is an implicit device which works via reference to retrieve information necessary for the proper application of rules. He refers to the employment of domains and recurrences (*adhikāra* and *anuvṛtti*) as well as the use of technical terms, which necessitates the construction of what he terms as *referential indices*, so that the required information can be gathered for the application of rules. In the first volume of his six volume edition of the Aṣṭādhyāyī he elaborates this idea further providing ample examples (R. N. Sharma 1987 p. 60-73). It suffices here to mention that his suggestions are largely an attempt to note down the (mental) process through which relevant information is gathered in order to apply the rules of grammar. It is clear that in a formal representation, which a computer should be able to understand and apply, this process needs to be made explicit. Although Sharma rightly points out its need and explains how it can be done following applicational procedure (*prakriyā*) of Pāṇini, he does not provide a practical framework in which it may be realized.

In contrast to the suggestion put forward by R. N. Sharma (1975 p. 31), I propose that it is necessary to recast the content and processes of grammar in a new `formal framework`. The framework which I intend to introduce differs from the way in which the Aṣṭādhyāyī is formulated by Pāṇini. The prime focus of the new formal framework is to integrate the representational and applicational aspects of grammar. The rules, once recast in the new framework, can be interpreted and applied in an algorithmic manner and the derivational process can be carried out with the aid of computer systems. Another important aspect of the new framework is its non-oral or non-verbal character in a formal and unambiguous manner. This enables on the one hand the object and the meta languages to be clearly distinguished, and on the other hand facilitates computer implementation.

The proposal I make to recast the Aṣṭādhyāyī in terms of a new formal framework is significant with far reaching consequences for the organization and presentation of the grammatical corpus. An immediate question arises here: why should one attempt to represent the Aṣṭādhyāyī in a new frame- work? Why not keep the Pāṇinian formulation that evinces several formal features? After all, it follows a systematic and advanced meta-language which, to a large extent, is employed in an unambiguous and consistent

manner. Further, looking at the large amount of literature claiming the Aṣṭādhyāyī to be the oldest and best example of formal representation, it seems an uninvited exercise.

Despite several advantages—like safeguards against human frailty—achieved through concise, well knit, verbal aphorisms in which the Aṣṭādhyāyī is formulated, its complex meta-language requires an elaborate and comprehensive apparatus for interpretation and application of the grammatical corpus. The framework in which these are formulated is suitable for oral transmission and human application of the grammar, but not suitable for its formal representation and algorithmic application. In the following I will first discuss the problems of formalization of the Aṣṭādhyāyī, and then introduce a new formal framework to represent it.

3.2 Formalization of the Pāṇinian corpus: challenges and possibilities

There are certain practical challenges if one decides to keep to the verbal framework in which Aṣṭādhyāyī is formulated. First of all, there is no critical edition of the Aṣṭādhyāyī of Pāṇini or other important earlier texts like the Mahābhāṣya of Patañjali—in which the Vārttika of Kātyāyana are also embedded—and later texts like the Kāśikāvṛtti of Jayāditya and Vāmana. All the pioneering editions in the last two centuries are vulgate editions without stemma. The issue was recently raised by Michael Witzel in an animated post to one of the electronic discussion groups, where he bemoaned the lack of efforts in preparing critical editions of basic Pāṇinian texts.[31] Responding to this scholars like Johannes Bronkhorst downplayed the expectations that such critical editions might fulfil, for example, to shed some extra light on the exact date of Pāṇini or issues regarding the later additions of certain portions of Aṣṭādhyāyī.[32] Madhav M. Deshpande, notes that "[T]his debate concerns a state of the Aṣṭādhyāyī that predates Kātyāyana and Patañjali, and no manuscript material takes us back to that period". Therefore, in his opinion "crucial questions of historicity of various sections within the Aṣṭādhyāyī cannot be resolved with critical editions based on very late manuscripts."[33]

---

[31] See the discussion group: Indo-Eurasian_research (`http://groups.yahoo.com/group/Indo-Eurasian_research/`) message 6303 posted on March 19, 2007 by witzel_michael (accessed on 12.01.2013).

[32] Bronkhorst (2008a p. 475-484) supplies a detailed reply to Witzel on this issue.

[33] See the discussion group: Indo-Eurasian_research (`http://groups.yahoo.com/group/Indo-Eurasian_research/`) message 6311 posted on March 20, 2007 by deshpandem (accessed on 12.01.2013).

A joint project, however, is being carried out by École pratique des hautes études, Paris; Facolta di Studi Orientali, Universita La Sapienza, Rome; and Bhandarkar Oriental Research Institute, Pune to bring out a critical edition of the Kāśikāvṛtti.[34] In this context, a notable publication is a critical edition together with translation of a section of Kāśikāvṛtti on *pratyāhāra*s, which is an edited volume by Pascale Haag and Vincenzo Vergiani (2009). Despite these ongoing efforts, it would take some years till we have any fully fledged critical editions of some important early texts on Pāṇinian grammar. And even then, the question whether it may be helpful in resolving any question of significance, remains open.[35]

The exact number of *sūtra*s in the Aṣṭādhyāyī is not fixed. An approximate count is close to 4000. Moreover, as Hartmut Scharfe (2009 p. 33) notes: "[W]e have no independent assurance that the division of *sūtra*s in our traditional text is always the one intended by Pāṇini". Further, changes in the boundary of rules may lead to differences in the number and divisions of the *sūtra*s. Sometimes a *sūtra*, which is traditionally given as one single rule, is divided into two for explaining the formation of certain words which otherwise are likely to be stamped as ungrammatical formations. This technique is called *yogavibhāga* and *sūtrabheda* (Abhyankar 1974 p. 318,432). Joshi and Roodbergen (1991 p. 20-23), for instance, split the rule *saṃbuddhau śākalyasyetāvanārṣe*[36] into two parts: *saṃbuddhau śākalyasya* and *itau anārṣe* and combine the two rules *uñaḥ*[37] and *ūḹ*[38] into one.

There is, thus, no clarity about whether the verbal formulation of the grammatical system is without corruption. Even if one decides to begin with the grammatical corpus as laid down in modern editions of Aṣṭādhyāyī, and agrees on a *standard* version,[39] several major problems still persist which

---

[34] For more information, see: `http://bori.ac.in/manuscript\_department.html` (accessed on 12.01.2015).

[35] Johannes Bronkhorst (2008a p. 482) remarks: "My expectation is that, even if all Pāṇinīyas were to mend their ways and spend their time making critical editions, and even if Witzel were to live to see the result, he might not find in (or through) these editions the answers he is looking for. To find these answers, other ways may have to be explored."

[36] संबुद्धौ शाकल्यस्येतावनार्षे ॥१.१.१६॥ ► according to Śākalya o(t) is termed pragṛhya if it is sambuddhi and when iti that is not Vedic (*ārṣa*) follows.

[37] उञः ॥१.१.१७॥ ► according to Śākalya u(ñ) is termed pragṛhya when iti that is not Vedic (*ārṣa*) follows.

[38] ऊँ ॥१.१.१८॥ ► according to Śākalya anunāsika ū replaces u(ñ) when iti that is not Vedic (*ārṣa*) follows and it is termed pragṛhya.

[39] There are different readings in modern editions as well and the task to prepare a *standard* version would be a time intensive exercise. Some efforts in this direction are made by Wiebke Petersen and Norbert Endres under the project "*Pratyāhāra*s or features? A qualitative analysis of phonological descriptive techniques—a comparison of Pāṇini's *pratyāhāra*s and phonological features" (Personal communication). Project website: `http://panini.phil.hhu.de/?section=home` (accessed: 22.02.2013).

have more to do with the nature of the Aṣṭādhyāyī then the non-availability of its critical edition.

The rules of Aṣṭādhyāyī are not always stated in an explicit manner and require ample interpretation for their comprehension and application. The grammatical corpus needs to be supplimented with missing expressions, examples, counter examples etc. It is only then that it can be put to use. Mere recitation of the grammatical corpus (or in our case, inputting it in the computer) does not suffice. Not even resolving the *sandhi*s and identifying the individual words. A user must exercise logical interpretations based on her or his knowledge of the grammatical system and the meta-linguistic conventions employed in the formulation of this system, in order to be able to apply it. Patañjali points out this in the following lines:

> It is not that one derives linguistic expressions only through rules of grammar. Then how are the linguistic expressions derived? Through rules of grammar together with reasoned explanations (*vyākhyāna*). Now, if it is to be said that reasoned explanation is nothing but rules separated into constituent parts, then it is not correct, because it is not just dissected words such as *vṛddhiḥ āt aic* (of a rule like *vṛddhirādaic*[40]). What, then, is reasoned explanation? It consists of examples, counter examples, completion of statements by filling the missing words. Reasoned explanation is all this combined together.[41]

It follows that one needs to provide mechanisms for incorporating the reasoned explanation (*vyākhyāna*) as well, if the formalization is to be put to application. This, however, is not exhaustively mentioned in the rules of grammar. In other words, the grammatical corpus of Aṣṭādhyāyī does not explicitly incorporate the understanding of its application. Therefore, for the purpose of application of grammar, mere rules of Aṣṭādhyāyī are not enough.

Another problem which makes the task of formalization of the Pāṇinian grammar difficult is the lack of consistent application of its meta-linguistic conventions. This point can be best presented through the detailed study of the principles of *anuvṛtti* (carrying over of earlier components to subsequent rules) by S. D. Joshi and Saroja Bhate. It shows that although the conventions for *anuvṛtti* hold good for a number of rules, there are several counter examples. To quote Joshi and Bhate (1984 p. 252):

> The conventions of *anuvṛtti*, which are called rules in a loose sense, are valid in the majority of cases. We do not claim that they are valid in each and every case. [...] There are counterexamples which go against the conventions of *anuvṛtti*, yet the most salient fact about the assumption of these conventions is that they can not be given up, even in the face of apparent counterexamples. These contradictory examples do

---

[40] वृद्धिरादैच् ॥१.१.१॥ ► vṛddhi stands for ā*t* and ai*c*.

[41] न हि सूत्रतः एव शब्दान्प्रतिपद्यन्ते। किं तर्हि। व्याख्यानतः च। ननु च तदेव सूत्रं विगृहीतं व्याख्यानं भवति। न केवलानि चर्चापदानि व्याख्यानं वृद्धिः आत् ऐच् इति। किं तर्हि। उदाहरणं प्रत्युदाहरणं वाक्याध्याहारः इति एतत्समुदितं व्याख्यानं भवति। (P. 122-124).

not disprove our conventions. Therefore we claim that the apparent counterexamples are clearly manifestations of later insertions.

The reason why Joshi and Bhate do not want to give up the conventions is that otherwise the functioning of the grammatical system would be impossible. At the same time the presence of counter-examples shows the impediments to the process of formalization. On the basis of the inconsistencies in the systematic use of the principles of *anuvṛtti*, they propose later insertions of rules in the Aṣṭādhyāyī. In particular, they point out that insertion of the Vedic rules and *nipātana* rules, the rules dealing with samāsa and taddhita formations leads to the irregularities in the application of the conventions of *anuvṛtti*.

The hypothesis that there must be successive additions is further corroborated by the presence of "conflicting and incompatible elements in different parts of the text". In this regard, the following remark of Joshi and Bhate (1984 p. 253) is significant:

> For instance, the same suffix is referred to differently. The instrumental suffix (ṭ)ā (4.1.2) is reffered to as ā(ṅ) in 7.3.105 and 7.3.120. The accusative dual suffix au(ṭ) is referred to as au(ṅ) in 7.1.18. The Aṣṭādhyāyī (for instance, 7.2.9) refers to certain elements which are not prescribed in its basic layer. Instead of the genetive case, the nominative is used in the samāsa-section to indicate that A becomes B (i.e. *sthānyādeśabhāva*).

Even if one is ready to ignore or amend the irregularities in consistent application of the meta-level conventions, it does not suffice for a direct formalization of the corpus of Aṣṭādhyāyī. The reason for this lies in the very nature of some of these conventions. As an example, one may quote a few of the principles specified by Joshi and Bhate in case of *anuvṛtti*:

> Sometimes to have a proper interpretation of the rule the locative case is to be changed into the nominative. (Convention no. 18c). Topics or sections need not necessarily be introduced by the section-heading rules. They can be understood on the basis of Pāṇini's style of phrasing the rules. (Convention no. 16l). If the same section heading (*adhikāra*) occurs in different sections, it must have some reason to do so. (Convention no. 18k).[42]

An examination of the above conventions makes it evident that although many of them are conducive for recognizing certain patterns in the formulation of the rules, they are not sufficient for a formal representation which a computer can understand and implement.

One further handicap is the language of the Aṣṭādhyāyī itself. The grammatical corpus is in Sanskrit. This means, the rules of the grammar are applied to the language of the grammar as well. Yet they are applied selectively and not in a consequent manner. Not all the rules that are applied for formation

---

[42] For the list of these conventions, see (Joshi and Bhate 1984 p. 271-279).

of linguistic expressions are applied in the formulation of grammatical expressions. In the words of S. D. Joshi and J. A. F. Roodbergen (1991 p. 2):

> Pāṇini's *sūtra*-language differs from ordinary, literary Sanskrit in this that the rules followed in this type of Sanskrit are not necessarily applied to the *sūtra*-language also. In literary Sanskrit a *samāhāradvandva* ending in /-c/ takes the *samāsānta*-suffix (ṭ)a(c) (5.4.106).[43] For instance, *vāktvācam* "the aggregate of voice and skin". But this rule is not applied to *ādaic*. An important reason for not applying Pāṇini's own rules to his own *sūtra*-language is the concern to keep the expression clear and unambiguous.

In other words, the rules of Aṣṭādhyāyī are applied to the language of Aṣṭādhyāyī only as long as clarity of the grammatical corpus is not undermined.

Another problem is the loss of meta-linguistic information, especially the accentuation and nasalization of the grammatical elements. According to the convention *svaritenādhikāraḥ*[44] a circumflex accent (svarita) indicates the beginning of a domain (*adhikāra*). But as P. S. Subrahmanyam (1999 p. 182) notes:

> The technical use of the circumflex accent was lost in later times along with the general loss of accent in the language. The author of Kāśikā (7th century A.D.) acknowledges this loss when he remarks: *pratijñāsvaritāḥ pāṇinīyāḥ* which amounts to saying "the circumflex accent is understood to be present only there where the Pāṇinian scholars think that it should be present".

Moreover, the question as to which vowel of an *adhikāra-sūtra* is accented is also not clear. According to Subrahmanyam (1999 p. 182):

> Patañjali did not touch this problem. But the author of the Kāśikā and his commentators (i.e. the authors of the Padamañjarī and Nyāsa) think that all the vowels of an *adhikāra-sūtra* have the circumflex accent.

The same problem is faced once one wants to decide which sounds in a component are it-markers. The meta-rule *upadeśe'janunāsika it*[45] says that the nasal vowels of a component should be an it-marker. But the nasalization of vowels in the original instruction was also lost and Kāśikāvṛtti notes that "the Pāṇinīyas decide about the nasalization on the basis of established agreement on this".[46] Similarly, accents (udātta, anudātta and svarita) were also used to mark the verbal roots. In this case as well, the accentuation was lost and later grammarians had to make separate lists of roots carrying these accent markers.[47] Meta-rules also do not cover all the eventualities. The set of meta-rules specifying the sounds that are it-markers in the original

---

[43] द्वन्द्वाच्चुदषहान्तात्समाहारे ॥५.४.१०६॥ ► the suffix (ṭ)a(c) is introduced after a dvandva compound ending in c*u*, d, ṣ or h provided it expresses collection.

[44] स्वरितेनाधिकारः ॥१.३.११॥ ► through svarita a domain is marked.

[45] उपदेशेऽजनुनासिक इत् ॥१.३.२॥ ► in grammatical instruction, an a*c* which is anunāsika is it.

[46] प्रतिज्ञाऽनुनासिक्याः पाणिनीयाः । (KV on 1.3.2).

[47] In the Dhātupāṭha the listing is sub-categorized under *udāttāḥ, udātta-itaḥ, anudātta-itaḥ, svarita-itaḥ* etc. See, for example, (Katre 1989 p. 1173-1200).

instruction do not include all the cases.[48] Kātyāyana notes this fact and suggests that the group *ir* added at the end of verbal bases in the Dhātupāṭha should be called it.[49] Patañjali gives rudh(ir) as an example, and clarifies that it is not enough to designate i and r one after another as markers, but the entire group *ir* should be deleted together.[50] It follows that one can not depend upon the meta-linguistic rules to decide the it-markers. An explicit annotation is necessary.

The arguments noted above confirm that the strong version of the formalization hypothesis is not tenable. In other words, it would not be possible to input the text of Aṣṭādhyāyī and get as output a formal representation which a computer program can understand or implement. This does not negate the fact that Aṣṭādhyāyī is itself an attempt to present the grammatical system in a formal manner. Still, it was meant for oral transmission and human application.[51] It would be anachronistic to expect that it should also fit the requirements of machines developed some two-and-a-half millenia later. Moreover, any such claim would ignore the efforts of both later Pāṇinīyas and the works of modern researchers to amend, explain and bring consistency in the corpus through several conventions.

3.2.1 Possibility of formalization of the grammatical system

The problems faced while attempting to provide a formal representation of the corpus of Aṣṭādhyāyī in a direct manner lead one to look for other ways. Here the following differentiations need to be recognized. Firstly, it is necessary to differentiate between the system and the framework within which this system is articulated. Secondly, it is important to distinguish between the oral/verbal framework in which Aṣṭādhyāyī is formulated and a formal framework which a computer program expects. Finally, it is essential to recognize the features of an organizational framework in contradistinction to the characteristics of an applicational framework. As these distinctions

---

[48] The following rules specify it-markers: उपदेशेऽजनुनासिक इत् ॥१.३.२॥ ► in grammatical instruction, an a*c* which is anunāsika is it. हलन्त्यम् ॥१.३.३॥ ► ha*l* coming at the end. न विभक्तौ तुस्माः ॥१.३.४॥ ► but not in vibhakti, the t*u*, s or m. आदिर्ञिटुडवः ॥१.३.५॥ ► the initial ñi, ṭu and ḍu (of dhātu). षः प्रत्ययस्य ॥१.३.६॥ ► ṣ of pratyaya. चुटू ॥१.३.७॥ ► c*u* or ṭ*u*. लशक्वतद्धिते ॥१.३.८॥ ► l, ś and k*u* of non taddhita.

[49] V 4 on 1.3.7.

[50] See (Joshi and Roodbergen 1994 p. 14). Pāṇini refers to entities with *ir* as it-markers in इरितो वा ॥३.१.५७॥ ► After irit dhātu replace (c)l[i] through a(ṅ) optionally before parasmaipada substitutes of l(u)(ṅ) expressing kartṛ.

[51] For arguments that the Aṣṭādhyāyī was developed keeping only oral representational apparatus at hand see also (P. 155-163) in (Joshi and Roodbergen 1986 p. 40).

are central to my approach for providing a formal representation of the Aṣṭādhyāyī, I examine them briefly.

The first distinction—namely between the system and the framework—corresponds to the difference between the questions as to *what* is being told and *how* it is being expressed. In the present case the questions would be: what is the grammatical system and how is it presented in the grammatical corpus. The conventions through which a framework is regulated correspond to the meta-linguistic conventions of the Aṣṭādhyāyī.

As an example, consider the use of it-markers. In the corpus of Aṣṭādhyāyī Pāṇini invests specific grammatical characteristics to the constituent components by attaching marker sounds to them. The Pāṇinian term for these indicators is "it" (meaning: that which goes away or disappears) and the commentators use the term *anubandha* as well.[52] The it-markers are not a part of the actual form of the components and are deleted unconditionally.[53] Their temporary character is noted by Patañjali who compares them with a crow perched on a house:

> If someone asks now, "from among the two houses, which one is the house of De-vadatta", one can say "that on which the crow is perching". And later, even if the identifying mark of that house is no more, once the crow flies away, the questioner recognizes the house.[54]

The above remark of Patañjali points out that even after the it-markers are deleted, their identifying character is *retained* or *saved* by the user in her or his memory. Why must they then be deleted at all? Several problems may otherwise arise.

Firstly, marker sounds increase the number of phonemes in a component. This poses a difficulty where, for example, components with only one phoneme are allowed. Therefore, a meta-rule is formulated which says that while counting the number of phonemes, the markers should not be counted.[55]

Secondly, markers may influence the final phoneme of a component. For example, the root dai(p) in fact ends with ai and not p. It is important to delete the marker sound p otherwise the long vowel substitute ā can

---

[52] This term "was chosen for mute significatory letters by ancient grammarians probably on account of the analogy of *anubandha paśu*, tied down at sacrifices to the post and subsequently slaughtered" (Abhyankar 1986 p. 25). Cf., e.g., MB on 1.1.26.

[53] Their unconditional elision is provided by the rule: तस्य लोपः ॥१.३.९॥ ▶ its elision (takes place).

[54] कतरत् देवदत्तस्य गृहम् । अदः यत्र असौ काकः इति । उत्पतिते काके यदि अपि नष्टं तत्गृहं भवति अन्ततः तं उद्देशं जानाति ।(MB on 1.1.26). Translation with minor variations from (Subrahmanyam 1999 p. 92).

[55] नानुबन्धकृतम् अनेकाल्त्वम् (PŚ. 6).

not replace the final sound ai of the verbal root as prescribed by the rule *ādeca upadeśe'śiti*[56]. An extra meta-rule is stated to clarify this. It posits that while looking at components ending with e*c* the markers should be ignored.[57]

Thirdly, markers also change the form of the components and may cause problem while deciding the similarity in form of two components. For example, the suffixes (k)a and a(ṇ) have the same form only if one ignores the markers. Therefore, an extra meta-rule is mentioned which says that while deciding the (dis)similarity of forms, markers should be ignored.[58]

As is evident, these problems are because of the fact that the addition of markers changes the *form* of the components. Grammatical conditions, however, take the form of a component without markers into consideration and therefore, they must be unconditionally deleted immediately after their introduction in the derivational process. Their effect or function, however, must be *remembered*.

This takes us to the second distinction, namely between the oral and formal frameworks. The critical question is how to record or represent the grammatical information in a formal framework. The text of Aṣṭādhyāyī was recited and then later put in writing in more than one script. The step which is needed for representing oral text in a written corpus needs to be taken further to represent it in a formal framework. Continuing with the example, the marker sounds which in roman transliterations of the corpus are noted by capital letters[59], or bold face letters[60], or put into brackets[61], can be recorded as an attribute attached with the corresponding component.

In the new framework, it-markers are represented as attributes attached to the form of corresponding components. Instead of mentioning it at the form level and thus mixing it with the form of the constituent components, they are placed at the content level together with other attributes.[62] For example, the markers ṇ and k attached with the phonetic form of the suffixes a(ṇ) and (k)a are not stored at the form level but at the content level as attributes ṇit and kit.

Separating the two levels and maintaining this distinction in the repre-

---

[56] आदेच उपदेशेऽशिति ॥६.१.४५॥ ▶ *āt* replaces e*c* occuring at the end of verbal roots enunciated in the original recitation, except when a suffix with ś as it-marker follows.

[57] नानुबन्ध्यकृतम् अनेजन्तत्वम् (PŚ. 7).

[58] नानुबन्ध्यकृतम् असारूप्यम् (PŚ. 8). This decision is important for application of the suffixe (k)a as an exception to the suffix a(ṇ). See (Subrahmanyam 1999 p. 187).

[59] Sumitra M. Katre's edition of the Aṣṭādhyāyī (1989) uses this convention as also the volumes by S. D. Joshi and J. A. F. Roodbergen (1975) etc.

[60] George Cardona (1997) follows this in his edition.

[61] This is used by P. S. Subrahmanyam (1999).

[62] See section 2.2.1 for the form and content levels.

sentation of grammar has several advantages. The meta-rules mentioned in (PŚ. 6-8) are now redundant. The Pāṇinian rule *tasya lopaḥ*[63] becomes redundant as well. This is because the marker attribute now being at the content level does not change the form of the component. Moreover, there are instances when a component with a marker looses its marker and is used without it. In other words, sometimes a particular marker is *removed* although originally it constitutes a part of the component. Thus by the rule *na ktvā seṭ*[64] the it-marker k is removed. This is easily executed by removing the corresponding attribute kit from the component.

Further, the new representation also makes it possible to take care of situations where the characteristic of having a marker is carried over to a component, although the marker sound is not expressedly attached to it. For example, the rule *sārvadhātukamapit*[65] says that a sārvadhātuka suffix which is not pit (or not marked with p) is as if ṅit (or marked with ṅ). This amounts to attaching the attribute ṅit to the appropriate component.[66]

At this point one may ask whether I am suggesting a new jargon to note down the Aṣṭādhyāyī? This is not the case for the following reasons: firstly, it is not that I am only providing new transliteration conventions or merely inputting the corpus into the computer. Such efforts provide for editions which only human readers can read, interpret and apply. On the other hand, a formal representation renders the grammar in such a manner that an algorithm can read, interpret and apply it. Secondly, in order to achieve this, the applicational aspects of the grammar need to be incorporated with the organizational concerns. Aṣṭādhyāyī is formulated with the aim of optimal organization of the grammatical content and processes. Its application is left to the user. This is natural as the prime concern was to record the information and since the system is learnt by a student, its application is not explicitly stated.

However, it does not mean that Aṣṭādhyāyī is oblivious to application. The rules are specified in an operational set-up, where the derivational process is carried out. The rules of Tripādī, or last three sections of the final chapter of Aṣṭādhyāyī, have a clear procedural thrust as well.

It is much later in the 14th and 15th centuries CE that the task of reorganization of the Aṣṭādhyāyī with procedural application (*prakriyā*) as its main focus was taken up. The two most important texts which attempt this are the Prakriyākaumudī of Rāmacandra (late 14th-15th cent.) followed by

---

[63] तस्य लोपः ॥१.३.९॥ ▶ its elision (takes place).

[64] न क्त्वा सेट् ॥१.२.१८॥ ▶ (k)tvā together with i(ṭ) looses kit.

[65] सार्वधातुकमपित् ॥१.२.४॥ ▶ sārvadhātuka pratyaya, if not pit, are assigned ṅit.

[66] In the Aṣṭādhyāyī the rules 1.2.1 to 1.2.26 provide for the assignment of it-attributes.

the Siddhāntakaumudī of Bhaṭṭoji Dīkṣita (late 16th-17th cent.). Both of them rearrange the *sūtra*s of the Aṣṭādhyāyī for the purpose of derivation of forms.[67]

The approaches of grammarians like Rāmacandra and Bhaṭṭoji Dīkṣita however are not without problems. Pāṇinian *sūtra*s make sense only in their original context. A *sūtra* is not an independent unit and for the sake of application several inter-connected *sūtra*s need to be taken into account. A reorganization disturbs the original contextual location of the *sūtra*s. This takes us to the third distinction, namely, how some information is organized, i.e. the organizational framework and how it can be put to use, i.e. the applicational framework. A formal representation takes care of both aspects and records them in an explicit and non-ambiguous manner. This facilitates an implementation through a computer program. Keeping the above distinctions in mind, I propose the following hypothesis.

Weak version of the formalization hypothesis: The Aṣṭādhyāyī in its current formulation cannot be formalized directly but the grammatical structure which it encapsulates can be reformulated in a formal framework. The formalized reformulation can then serve as an input to a computer program that can interpret and implement it.

3.3 Specification of the formal framework

The discussions in the previous sections show that a formal framework needs to be developed in order to represent the content and processes of Aṣṭādhyāyī if one wants logical systems to interpret and apply it. An important feature of such a framework has to be a clear distinction and separation between the content of the grammar and the manner in which it is formulated. In other words, the framework should be independent of the content of grammar. It should only facilitate the representation of the structures and the system, without in any way influencing them.

There are two basic concepts upon which the present framework is based. The first concept is that any given *whole* can be analyzed into *components*. The second is the idea that any given entity can be described through a number of characterizing *attributes*.

While proposing the manner in which the grammatical elements—namely the components, attributes and meaning-expressions—should be represented

---

[67] Compared to the earlier attempt of Rāmacandra, Bhaṭṭoji Dīkṣita covers all rules of the Aṣṭādhyāyī and follows them consequently. See (Houben 2008 p. 563-574).

in the new framework, the way in which they are employed during the process of derivation is also taken into account. Similarly, the simulation of the derivational process takes into consideration the manner in which the grammatical elements are represented. The focus, therefore, is to integrate the organizational and applicational aspects of grammar.

As discussed in section 2.3 the derivational process demands a mechanism through which the changing forms and contents of the grammatical elements can be tracked in a convenient manner. Equally important is to keep track of the process of derivation, i.e. the previous stages attained during the process, and record them in a manner so that they can be consulted when deciding the future steps. Moreover, central to the Pāṇinian process is the interaction with the user. While specifying the formal framework, these general guidelines need to be taken into consideration.

The new framework which I put forward is designed to represent the categories and processes outlined in the previous chapter about the grammatical system. Three kinds of basic elements are proposed:

1. The constituent `components` p_i of the language.

2. The diverse `attributes` a_k which are either explicitly mentioned or sometimes implicitly used in the grammar.

3. The copious `meaning-expressions` x_j^m that encode a multitude of information which is external to the grammatical system.

While attributes and meaning-expressions work at the *content-level*, the *form* of a component is of utmost relevance for the derivation of the linguistic expressions. This is depicted in the following diagram.
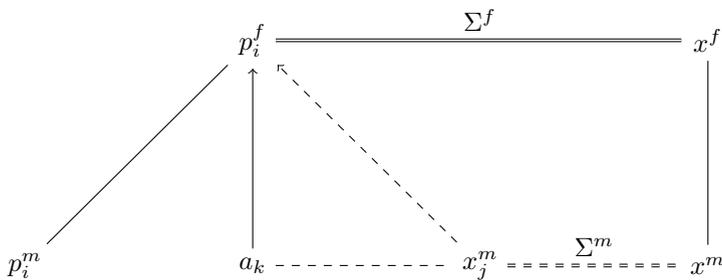
There are two fundamental *processes* in this formulation.

1. Specification of the components. A component is characterized by associating the information from the content level. Thus, some meaning can be established through usage p_i^m, or some grammatical attribute a_k may be attached to it, or some meaning-expression x_j^m may be associated with

it. Further, some specific form—e.g. presence of some phonetic feature—characterizes a component as well.

2. Combination of the components. This involves introduction of new components in the derivational process.

3.3.1 Representing a component

The first task is to specify how the constituent components can be represented. For this purpose it is necessary to look into how these are formulated within the corpus of Aṣṭādhyāyī and whether the new representation adequately incorporates all the aspects of Pāṇinian formulations.

In section 2.2.1 we have seen that the constituent components p_i of a linguistic expression x consist of two levels, namely the form and the content levels. At the form level these are clearly made up of a sequence of sound units. An obvious manner to represent them would be a sequence or list. For example, the component bhū can be seen as a sequence of two sounds bh and ū which is represented as

$$[\text{bh, ū}]$$

Aṣṭādhyāyī contains "the first enunciation" (*upadeśa*) of the constituent components of Sanskrit.[68] Looking from the perspective of the form level one can speak of two distinct parts in the Aṣṭādhyāyī according to the two fundamental types of the constituent components—the basic sounds and secondly their sequences.

1. In the beginning of the Aṣṭādhyāyī the first part enumerates the fundamental sounds. These listings are called the Śivasūtras.

2. The rest of the Aṣṭādhyāyī enunciates components of the language which are a sequence of one or more basic sounds. These collections are called the Sūtrapāṭha, the Dhātupāṭha and the Gaṇapāṭha.

The listing of vowels in the Śivasūtras corresponds to a group of sounds and each one of them is further specified by additional attributes. For example, the sound a stands for several variations of this fundamental sound, namely short (hrasva) /a/, long (dīrgha) /a/ etc. If one represents a sound through a set, then the general sound /a/ would be represented by {a} and the long vowel /ā/ by {a, dīrgha}. In other words, adding more attributes to the set

---

[68] The word *upadeśa* refers to the original or first instruction of the grammatical components in the corpus before it enters and changes its form through the derivational process (*prakriyā*). For a detailed study on the concept of *upadeśa* in Sanskrit grammar see (Biswal 1996).

which represents a sound would result in further restricting and specifying a particular sound. Thus, {a, dīrgha, udātta, anunāsika} denotes a particular, more specific instance of the vowel /a/.

> A sound can be represented through a set consisting of a *fundamental-sound* and a number of other characterizing *attributes*. It should be noted that within a set representing any sound, there can be *exactly one* fundamental-sound. In the following, I will call such a set a `sound-set`.

In the above example, the fundamental sound is /a/ and the characterizing attributes are dīrgha, udātta etc. The Śivasūtras provide the set of fundamental-sounds.

{a, i, u, ṛ, ḷ, e, o, ai, au, h, y, v, r, l, ñ, m, ṅ, ṇ, n, jh, bh, gh, ḍh, dh, j, b, g, ḍ, d, kh, ph, ch, ṭh, th, c, ṭ, t, k, p, ś, ṣ, s, h}

The form of any constituent component of a linguistic expression can now be represented as a sequence of sound-sets. Consider again the example of the component bhū. It can be represented as a sequence of two sound-sets:

bhū = [{bh}, {u, dīrgha, udātta}]

This representation allows one to include any number of attributes that may characterize the sounds. The attributes that are associated with a sound include phonetic and phonological features.[69] Thus, one can include other attributes e.g. the attribute *vowel* for which the Pāṇinian term is a*c*.

bhū = [{bh}, {u, dīrgha, udātta, a*c*}]

In this manner the information which is specific to a particular sound can be incorporated in the corresponding sound-set.

Some information, however, is shared by more than one sounds. For example, the attribute dhātu (verbal root) is valid for both sounds of this component. Accordingly, this attribute can be included in both sound-sets.

bhū = [{bh, dhātu}, {u, a*c*, dīrgha, udātta, dhātu}]

The sound-sets in a sequence share those attributes that are common to them. The presence of the attribute dhātu in both of these sets indicates that both of them in this sequence are part of dhātu or a verbal root.

In the above sequence of sound-sets, thus far, there is no reference to

---

[69] For an introduction to feature systems see (Clark, Yallop and Fletcher 2007 p. 372-390). James Stanton Bare (1980) has studied the system of features implicit in the Aṣṭādhyāyī. Interpretations of the Pāṇinīyas on the question of nearness (*āntaratamya*) is discussed by Robert A. Hueckstedt (1995).

bhū, the original component which is being represented here. This information can be included by adding bhū to both the sound-sets that form a part of this component.

$$X = [\{bh, dhātu, bhū\}, \{u, ac, dīrgha, udātta, dhātu, bhū\}]$$

The linguistic component can now be summarised as X, and all the information can be gathered by the sound-sets that are contained in it.[70]

The above example leads us to specify a general representation for the components.

$$X = [\{sd_1, a_1, a_2, \ldots\}, \{sd_2, a_1, a_{17}, a_{43}, \ldots\}, \{sd_i, a_k, \ldots\}, \ldots]$$

Here, X is a sequence of sets. Each set consists of *exactly one* key from the collection of fundamental sounds. Thus, the first set has sd_1, the second sd_2 etc. Further, each sound-set may have a number of attributes. Thus the first sound-set has attributes a_1 and a_2 etc. It should be noted that the unique fundamental sounds in each sound-set represent the basic phonetic form of that unit, and the attributes comprehend the characteristics shared commonly among several sound-sets or which are individual to a particular sound-set. For example, the attribute a_1 is common to the first two sound-sets and a_2 is unique to the first sound-set.[71]

Within the new framework, each element, i.e. the constituent components, attributes or meaning-expressions are assigned a unique identifier or an ID. There is an ID

- for every sound, like /a/ (ID: a_0), /i/ (ID: i_0), /u/ (ID: u_0) etc.

- for the phonetic form of every component: a (ID: a_2), bhū (ID: bhU_a), (ś)a(p) (ID: zap_0), ti(p) (ID: tip_0) etc.

- for the attributes: vṛddhi (ID: vRddhi_0), dhātu (ID: dhAtu_0) etc.

- for meaning-expressions: *vartamāna* (ID: vartamAna_x), and

- for lexical content: *bālaka* (ID: bAlaka_x).

Assignment of an ID ensures a non-ambiguous identification of the elements of grammar. For example, the substitute suffix a (ID: a_2) is distinguished from the phoneme /a/ (ID: a_0) on the basis of their IDs. However, for the sake of readability, I will not display the underlying IDs.

---

[70] There are a number of other attributes that are needed to be included in order to have complete information about the component bhū, e.g., that it is part of the first group of verbal roots etc.

[71] Thus far, only one constraint is mentioned—namely the presence of exactly one of the fundamental sounds in each of the sound-sets. In due course, other constraints will be mentioned that are imposed by the grammar to specify particular components.

The components of any linguistic expression specified in the Pāṇinian corpus can be represented in terms of a sequence of sound-sets. Each of the sound-sets in such a sequence consists of *exactly one* ID from the set of fundamental-sounds. Further, it contains a number of other IDs corresponding to other attributes and meaning-expressions. Moreover, if this sound-set is part of some specific component, then this information is also incorporated by including the ID of that component with it. Within the new framework, such sequences of sound-sets are termed as `language-components` and can be defined as follows.

> A `language-component` corresponds to any component of grammar and is a sequence of one or more sound-sets.

One of the features of the grammatical system specified in the previous chapter was the attachment of the attribute to some component.

$$p_i^f$$

$$\uparrow$$

$$a_k$$

In the above diagram, the attribute a_k is attached to the form of the components p_i^f. The present framework implements the attachment of the attributes in the following manner. If the form of some component has the sounds sd_1, sd_2 etc., then the attributes that are common to them are attached with each of the constituting sounds and are included in the respective sound-sets.

$$p_i^f: \qquad sd_1 \qquad sd_2 \qquad sd_i$$

$$a_k$$
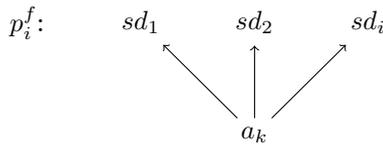
The lexical meaning p_i^m as well as the meaning-expressions x_j^m associated with some component are recorded in a similar manner. Thus, if a lexical meaning is connected with some component, then it is connected to each sound-set of that component.

In the following section I will provide examples to demonstrate that the components specified in the grammatical corpus of Aṣṭādhyāyī can be adequately represented in the new framework. Moreover, I will also point out how the Pāṇinian articulation differs from the new one, but at the same time all the information needed for grammatical processing is aslo included in my representation.

In the Pāṇinian formulation the fundamental sounds are represented directly. The new framework stores them through a sound-set with a unique identifier. Moreover, the phonetic attributes are stored explicitly through attribute IDs.

1. The sound /a/ in the Aṣṭādhyāyī will be equivalent to {a_0}. The long variation ā is equivalent to {a_0, dIrgha_0}. Pāṇini attaches the marker sound /t/ to restrict the time duration.[72] Thus Pāṇinian ā*t* is equivalent to {a_0, dIrgha_0} and a*t* is equivalent to {a_0, hrasva_0}. One can also record the Pāṇinian attribute for short or long /a/, namely a*t* or ā*t* by including the IDs at_1 or At_2 to the corresponding sound-set. Thus {a_0, dIrgha_0, At_2} incorporates the information that the sound is called ā*t* by Pāṇini. Inclusion of this information is redundant to some extent, but I include it to record the Pāṇinian term and to be able to correlate with the corpus of the Aṣṭādhyāyī.

2. Sounds like the nasal sound ṃ or the aspirated sound ḥ are not listed explicitly in the Śivasūtra but are referred to in the grammar by their names: anusvāra for ṃ and visarjanīya for ḥ. The new formulation includes the surface form of the sounds named through anusvāra or visarjanīya. Thus, the corresponding sound-sets are {M_1, anusvAra_0} and {H_1, visarjanIya_0}.

3. In the Pāṇinian formulation the constituent components can have additional indicatory sounds, namely the it-markers. These are attached either at the beginning or at the end of the base form of a component. For example, the suffix ti(p) consists of the sound p as marker at the end. Similarly, the infix (ś)a(p) has a marker sound ś in the beginning and p at the end. Sometimes more than one sound is added at the end or in the beginning, for example, yuj(ir) or (ḍu)krī(ñ).[73] In the new framework, the markers are included as an attribute. For example:

$$ti(p) \cong [\{t, pit\}, \{i, hrasva, pit\}]$$
$$(ś)a(p) \cong [\{a, hrasva, śit, pit\}]$$
$$yuj(ir) \cong [\{y, irit\}, \{u, hrasva, irit\}, \{j, irit\}]$$

Thus, ti(p) is a component with the attribute pit i.e. "having p as it". Similarly, (ś)a(p) is śit and pit and yuj(ir) is irit. The sign $\cong$ refers to *partial equivalence* of the Pāṇinian component in the new formal framework.

---

[72] तपरस्तत्कालस्य ॥ १.१.७० ॥ ► an a*c* followed by the marker t stands for sounds having the same time duration.

[73] Vidya Niwas Misra (1966) provides a list of these markers and the "functional load" associated with them. P. S. Subrahmanyam (1999 p. 92-108) uses the term indicators and notes them as well. All such marker sounds are represented by ( ) brackets in our writing convention.

4. Within the Pāṇinian system, a sound can disappear, i.e. it can cease to be manifest at the audible level or the form level. The rule *adarśanaṃ lopaḥ*[74] states that the result of the attribute lopa is "non-appearance" (*adarśana*) of a sound. The same is the case with luk, ślu and lup which are attached only to suffixes.[75] The presence of attributes like lopa, luk, ślu and lup would imply that the corresponding sound-set is invisible or mute at the form level.

5. It is sometimes difficult to pronounce the constituent components as they consist only of consonant clusters without a vowel. In such cases, a vowel is added for facilitating pronunciation. This is necessary for an oral corpus. For example, the augmentation n[u](m) where the vowel u is just for the sake of pronunciation. In the printed edition, these are noted within square brackets [ ]. Within the new framework, this information is redundant although it is stored for the sake of exact reproduction of the components in the original corpus.

Within the new framework, a component of any linguistic expression is represented in terms of a language-component, which is modeled as a sequence of sound-sets. These components can be seen as units of an inflected word.

For example, the word *paṭhati* has components paṭh(a), (ś)a(p), ti(p) etc. Although the derivational manuals like Prakriyākaumudī and Siddhānta-kaumudī provide for derivation of individual words, the Pāṇinian system of Sanskrit grammar and especially its process of derivation functions not at the level of individual words, but at the level of an entire sentence. In the new framework, this aspect is taken care of by introducing the concept of a `sentence` and defining it as follows.

> A `sentence` is a sequence of one or more language-components.

At this point it is important to mention that the above definition of a `sentence` refers to the collection of language-components within the new framework. It should not be confused with the linguistic definition of a sentence and should only be viewed as a sort of container in which the language-components can be stored in a sequential manner.

---

[74] अदर्शनं लोपः ॥१.१.६०॥ ► non-appearance is termed lopa.

[75] प्रत्ययस्य लुक्श्लुलुपः ॥१.१.६१॥ ► non-appearance of a suffix is termed luk, ślu and lup.

3.3.2 Comprehending the dynamics of the derivational process

Thus far, a framework for representation of the Pāṇinian components has been introduced. The grammatical corpus not only provides the constituents of a linguistic expression but also the processes to combine them. This synthesis takes place in a number of steps. Within the traditional Pāṇinian framework this process is carried out by a person who takes decisions regarding the choices of introducing the new components. As mentioned earlier, the derivational process is guided by

1. the intention (*vivakṣā*) of the speaker, and

2. the constraints of the grammatical system, which involves

   a. consistency of the derivational state

   b. its saturation, and

   c. its completion.

The new framework takes care of both the human and the grammatical aspects of the derivational processes.

Consider the example of derivation of the Sanskrit expression *bālakaḥ paṭhati* (a boy reads) worked out in the previous chapter. The first step is the introduction of the constituent components bālaka and paṭh(a). Their representation in the new framework would be in terms of a sentence consisting of two language-components: if X1 is the language-component for bālaka and X2 for paṭh(a) then depending upon their mutual order, the sentence S is given by a sequence of X1 and X2.

$$S = [X1, X2]$$

Further, each language-component is a sequence of sound-sets:

- X1 = [{b},{a, dīrgha}, {l}, {a, hrasva}, {k}, {a, hrasva}]
- X2 = [{p},{a, hrasva}, {ṭh}]

For better readability, I note the language-components with their respective sound-sets in the following manner:

S = [X1, X2]

X1 ss1 {b, bālaka}

    ss2 {a, dīrgha, bālaka}

    ss3 {l, bālaka}

    ss4 {a, hrasva, bālaka}

```
    ss5 {k, bālaka}

    ss6 {a, hrasva, bālaka}

X2 ss1 {p, paṭh(a), ait, udāttet}

    ss2 {a, dīrgha, paṭh(a), ait, udāttet}

    ss3 {ṭh, paṭh(a), ait, udāttet}
```

The sound-sets consist of a number of IDs. As mentioned earlier, an important constraint is that each sound-set must contain *exactly one* ID from among the set of fundamental sounds. For example, a sound-set like {a, i} is invalid. One can say that the collection of fundamental sounds forms a set of mutually exclusive IDs. Another example of a set of mutually exclusive IDs is the set {hrasva, dīrgha, pluta}. Within a particular sound-set the presence of an ID from this set excludes the possibility of inclusion of any of its other IDs. One can introduce the concept of consistency of a sound-set. One such condition of consistency would be that a sound-set should not have more than one ID from the set of mutually exclusive IDs.

There is a separation of the form and the content level. The language-component X2 represents paṭh(a). In the Pāṇinian formulation it has four sounds / p a ṭh (a)/ where the last one is only a marker, which is unconditionally deleted by the rule *tasya lopaḥ*[76]. In the new formulation, the marker sound is not represented through an extra sound-set but is included as an attribute (ait) at the content level. Thus, only as many sound-sets are formed as there are sounds that contribute to the form of the linguistic expression.

Attachment of the attribute dhātu. The rule *bhūvādayo dhātavaḥ*[77] says that the components mentioned in the list beginning with bhū—i.e. the list of verbal roots or Dhātupāṭha—be assigned the attribute dhātu. In the new framework this assignment is carried out in the following manner: the sound-sets of the language-components are checked and if an ID which is also part of the set of the verbal roots is present, then that sound-set is attached the ID corresponding to the attribute dhātu. This can be formulated as follows: given a language-component Xi of some sentence S, if it contains one of the IDs belonging to the set {bhū, …, paṭh(a), …}, then attach the attribute dhātu.

S = [X1, X2]

X1 …

X2 ss1 {p, paṭh(a), ait, udāttet, **dhātu**}

    ss2 {a, dīrgha, paṭh(a), ait, udāttet, **dhātu**}

---

[76] तस्य लोपः ॥१.३.९॥ ► its elision (takes place).

[77] भूवादयो धातवः ॥१.३.१॥ ► components in the list beginning with bhū are dhātu.

ss3 {ṭh, paṭh(a), ait, udāttet, **dhātu**}

Here it should be noted, that in the new framework there is no need to introduce extra indices for different constituents as the IDs of components, attributes and meaning-expressions provide the requisite identification. So in this case, each sound-set of the language-components will be checked for whether it has some ID (in this case paṭh(a)) belonging to the set of the IDs of dhātus, and if this is the case then the attribute dhātu is added to that sound-set.

Similarly, attachment of other attributes e.g. prātipadika, kartṛ, pratyaya, vṛddhi etc. would bring about inclusion of the attribute IDs to the *appropriate* sound-sets. The selection of the appropriate sound-sets would depend upon examination of different conditions. Thus, attachment of an attribute results in addition of the IDs to one or more sound-sets.

Adding a new component l(a)(ṭ). Addition of a new component involves extension of the sentence by adding new language-components. The rule *vartamāne laṭ*[78] introduces l(a)(ṭ) if present tense is intended. The derivational state looks as follows:

S = [X1, X2, X3]

X1 …

X2 ss1 {p, paṭh(a), ait, udāttet, sakarmaka, dhātu}

    ss2 {a, dīrgha, paṭh(a), ait, udāttet, sakarmaka, dhātu}

    ss3 {ṭh, paṭh(a), ait, udāttet, sakarmaka, dhātu}

X3 ss1 {l, l(a)(ṭ), ṭit}

Here, a new language-component X3 is added after X2. For this, the indices of the language-component with dhātu need to be identified and the new component should then be introduced after it.

Replacement of l(a)(ṭ) by ti(p). Replacement of a component by another can now be represented as a combination of addition and attribute attachment. Thus, in order to replace l(a)(ṭ) by ti(p), the language-component X4 corresponding to ti(p) is added after the language-component X3, which corresponds to l(a)(ṭ). Further, a new attribute δ (denoting "replaced") is attached to X3. The derivational state looks as follows:

S = [X1, X2, X3, X4]

X1 …

---

[78] वर्तमाने लट् ॥३.२.१२३॥ ► to express present time, introduce l(a)(ṭ).

X2 …

X3 ss1 {l, l(a)(ṭ), ṭit, abhihita-kartṛ, δ}

X4 ss1 {t, ti(p), pit, prathama, ekavacana}

   ss2 {i, hrasva, ti(p), pit, prathama, ekavacana}

It should be noted that the component l(a)(ṭ) which has been replaced is not dropped from the sentence. Instead, a new attribute δ is attached to it. In this manner the information which it carries remains available for all further steps. In the traditional framework, owing to its oral and linear nature, the substituted component needs to be removed and in its place, the new replacement is placed.

The process of derivation progresses through a number of derivational states. A particular derivational state represents the effect of some grammatical operation on the current sentence. The grammatical operations are enjoined by the operational rules of grammar. Within the new framework, the operational rules of Aṣṭādhyāyī are modeled as `statements` (see next chapter).

> A `derivational state`, therefore, stores the effect of a `statement` when applied on a sentence.

The above example indicates that from the perspective of the new framework, there are two basic types of operations which the operational rules enjoin:

1. Attachment of an attribute to some language-component or a sound-set.

2. Addition of a component to some language-component.

The first category corresponds to the process of saturation of a derivational state and the second one for its completion. A number of saturating statements are applied giving rise to a sequence of derivational states. A container is required to collect these derivational states. Such a sequence of derivational states is termed a `slice`. Thus, a slice is defined as follows.

> `Slice` is a sequence of derivational states.

Attributes are attached as long as the level of saturation is not reached. Once no more attribute can be attached to a particular derivational state, then the current slice attains saturation. Then, depending upon the conditions, a new component can be added which prompts an introduction of a new slice. As long as new components can be introduced, the process remains incomplete. Again, a container is required to collect the slices. Once no more component can be introduced it attains completion.

The sequence of slices is collected in a `process-strip`. Thus, a process-strip is defined as follows.

> `Process-strip` is a sequence of slices.

To sum up, the entire derivational process is modeled through a `process-strip` which is a sequence of `slices`. Whenever a new component is added, a new slice is introduced. Within a slice, there is a sequence of `derivational states`. Each derivational state corresponds to the attachment of some attribute. Finallly, a derivational state stores the effect of some operational `statement` on a `sentence`. Further, a sentence is a sequence of `language-components` which in turn consists of a sequence of `sound-sets`.