Modeling Pāṇinian Grammar

Peter M. Scharf
Brown University
10 December 2007

It is possible to achieve the implementation of generative grammars and parsers of Sanskrit using various methodologies which have varying degrees of affinity to those of Pāṇinian grammar.[2][1] The current paper compares obvious methods to implement a few aspects of Sanskrit grammar computationally, comments upon the degree to which they approach or depart from Pāṇinian methodology and exemplifies methods that would achieve a closer model. Two questions essential to determining a basic framework in which to implement Pāṇinian grammar computaionally are dealt with in some detail: the question of levels and the role of semantics.

I. Differences among the Sanskrit grammarians and even among Pāṇinians.

In attempting to create a computational model of Pāṇinian grammar, the first problem is to determine which Pāṇinian grammar.[3] The *Aṣṭādhyāyī* itself (late 5th c. B.C.E.), consisting of nearly 4,000 rules, is known to have undergone modifications. Kātyāyana's approximately 4,300 vārtikas (4th-3rd c. B.C.E.) suggest modifications to 1,245 of Pāṇini's rules, usually in the form of additions (*upasaṅkhyāna*). Patañjali's *Mahābhāṣya* (mid-2nd c. B.C.E.) rejects many additions suggested by Kātyāyana, suggests other desiderata (*iṣṭi*), and articulates principles presupposed in the grammar. Many of the modifications Kātyāyana and Patañjali suggest are found adopted in the form in which the rules are found in Jayāditya and Vāmana's *Kāśikā*, the oldest extant complete running commentary on the *Aṣṭādhyāyī* (7th c. C.E.). Does one wish to model the *Aṣṭādhyāyī* alone? The *Aṣṭādhyāyī* and Kātyāyana's vārtikas? The grammar as known and approved by Patañjali in the *Mahābhāṣya*? Or the grammar as found in the *Kāśikā*?

II. Ambiguities in early articulations explicated differently by subsequent Indian linguists.

Articulations of Pāṇinian grammar, especially sūtras and vārtikas isolated from commentary, are subject to ambiguities. These ambiguities are resolved in different ways by different commentators. Commentaries on Patañjali's *Mahābhāṣya* disagree with each other; commentaries on the *Kāśikā* disagree with each other; and Bhaṭṭojidīkṣita's *Siddhāntakaumudī* (17th c. C.E.) differs in its interpretation of rules and procedures from Jayāditya and Vāmana's *Kāśikā*. Moreover, subcommentaries differ in their interpretations. One must determine the manner in which these ambiguities are to be resolved. Are they to be resolved using some particular commentator as the authority? Haphazardly? Or is one going to come to an independent judgment of the correct interpretation after a critical evaluation of the various interpretations?

Moreover, the supplements to the grammar,[4] particularly the lists referred to in various rules (*gaṇa*s), most prominently the list of roots, *Dhātupāṭha*, have undergone variation. Three complete commentaries composed in Sanskrit are extant on the Pāṇinian *Dhātupāṭha*,[6] which is known only through these commentaries: the *Kṣīrataraṅgiṇī* of Kṣīrasvāmin (early twelfth century C.E. Kashmir), the *Dhātupradīpa* of Maitreyarakṣita (mid-twelfth century C.E. Bengal), and the *Mādhavīyadhātuvṛtti* of Sāyaṇa (fourteenth century C.E. Vijayanagara, Karnataka). Will one use one of these? A unified critical edition of them? Or will one attempt to reconstruct the *Dhātupāṭha* as known to Patañjali? Other lists (*gaṇa*) are specified only in

---

[1] Numbers in brackets refer to slide images in the accompanying PDF document.

commentaries, and many of these are called paradigmatic rather than exhaustive. Will one rely on lexical lists external to the grammar, such as nighaṇṭus and kośas, to complete these lists?[5]

Before embarking on a computational implementation of Pāṇinian grammar, such decisions ought to be made. It may prove very interesting to compare computational implementations based upon different rule sets, different interpretations, and different sets of supplementary lists with each other and with different sets of linguistic data. As I have argued in two papers, with respect to the derivation of subjunctives (2005) and of the present stems of class eight roots (forthcoming), systematic comparison of linguistic descriptions resulting from computational implementations with each other and with various collections of extant Sanskrit texts may throw important light upon interpretational and historical questions.

III. *Utilization of contemporary linguistic models, in particular those derived from Pāṇinian methodology, to articulate Pāṇinian methodology.*

Indian grammatical commentaries composed in Sanskrit over the last two and half millennia are not the only sources of Pāṇinian interpretation. Recent work in theoretical and computational linguistics has influenced the interpretation of Pāṇinian grammar.

A. Influence of Pāṇinian methodology on contemporary linguistics generally.

Although often not explicitly acknowledged by the influential linguists indebted to it nor recognized by historians of linguistics, Pāṇinian grammar has had a profound influence on modern linguistics. Apart from the influence of ancient Indian phonology on modern phonetic feature analysis, and the emulation of ancient Indian synchronic sound change laws by diachronic laws of phonological change in modern historical and comparative linguistics, Pāṇinian

grammar supplied the basic archetype at the foundation of modern generative grammar. From Chomsky's first work on transformational grammar in 1957 to the Pāṇinian grammars of modern Indian languages such as described for Hindi in Bharati et al 1995, modern linguistic science is heavily indebted to the concepts and procedures of ancient Indian linguistics.

B. Influence of contemporary linguistic models on the interpretation of Pāṇinian methodology.

Concepts originally inspired by ancient Indian linguistics have taken their own shape in contemporary linguistics. They have responded to different concerns and been adapted to different questions. These new concepts have been applied by contemporary scholars to the interpretation of Pāṇinian grammar. One of the most prominent of these is the idea that grammar consists of modules in a generative hierarchy, or levels.

IV. *Levels*

A. Kiparsky's architecture

Clearly influenced by Chomskian generative grammar, Kiparsky and Staal (1969)[7] proposed that Pāṇinian grammar contains rules in a hierarchy of four levels of representation: semantics, deep structure, surface structure, and phonology. More recently Kiparsky (2002)[8] restates this scheme referring to the four levels as follows: (1) semantic, (2) morphosyntactic, (3) abstract morphological, and (4) phonological. Three classes of rules map prior levels onto subsequent levels: (1)[9] rules that assign kārakas and abstract tense, (2)[10] morphological spellout rules, and (3)[11] rules of allomorphy and phonology. Rules incorporate conditions at both the levels from which and to which they map, as well as at prior levels in a unidirectional derivation beginning with semantics and ending with phonology.

As an example of how derivation is understood to work in the four-level hierarchy, one may take the derivation of the sentence *Devadatta odanaṁ pacati*.

At the semantic level,[12] the speaker intends to express that Devadatta, called here John Doe, undertakes the action of cooking in present time for the purpose of making boiled rice. Pāṇinian semantics classifies John Doe as the independent agent in the action, and boiled rice as that which is desired to be obtained. Three rules[13] apply to map the semantic level onto the morphosyntactic level. 1.4.49 and 1.4.54 assign kārakas, and 3.2.123 assigns abstract tense by introducing the *l*-affix *laṭ* on the condition that present time is to be denoted.

Several "spellout" rules then apply to map the morphosyntactic level onto the abstract morphological level.[14] 3.4.78 provides that basic verbal terminations replace the *l* of the affix *laṭ* that occurs after the verbal root *pac*. Restrictive rules 1.3.78, 1.4.108 and 1.4.22, read in conjunction with 3.4.78, select the third person singular active (3sa) affix *tip* on condition that a single agent that is neither the speaker nor the adressee is to be denoted. Before the affix *tip* (termed *sārvadhātuka* by 3.4.113 *tiṅśitsārvadhātukam*), 3.1.68 provides the default verbal stem-forming affix *śap* to cosignify the agent. Then 4.1.2 provides nominal terminations. Restrictive rules 2.3.2, 2.3.46, and 1.4.22, read in conjuction with 4.1.2 select the appropriate nominal termination. 2.3.2 selects a second triplet nominal termination (*dvitīyā*) after the stem *odana* on condition that the kāraka *karman*, which has not yet been denoted (*anabhihite* 2.3.1), is to be denoted. 2.3.46 selects a first triplet nominal termination (*prathamā*) after the stem *devadatta* on condition that just the stem meaning, gender, and number are to be denoted. (The kāraka *kartṛ* has already been denoted by the verbal termination thus preventing 2.3.18 *kartṛkaraṇayos tṛtīyā* from applying.) 1.4.22 selects the singular terminations *am* (2s) and *su* (1s), respectively in each triplet.[2]

Finally, several rules of allomorphy (of which there are none in the present

example) and phonology apply to map the abstract morphological level onto the phonological level.[15][3] The state of the example sentence at each of the four levels is summarized in [16].

B. Houben 1999

Houben (1999)[17] aptly criticized earlier articulations of this four-level hierarchy because they did not explicitly include pragmatics and intentionality in the semantic level and did not permit semantic factors (including pragmatics and intentionality) to serve as conditions in phonological rules directly. In addition, he criticized the portrayal of Pāṇini's grammar as a complete automaton that produces utterances from meanings. He pointed out that there are no rules that introduce verbal roots and nominal stems based upon semantic conditions and that the fundamental morphophonemic elements appear in Pāṇinian derivations from the start. It is therefore improper, he argued, to characterize the grammar as originating in semantics and culminating in phonological form. Rather, he (1999: 48) stated, it originates in meaning mixed with form and culminates in a perfected form.

C. The purpose of the science of language

Houben is correct to reemphasize that it is not the function of the *Aṣṭādhyāyī* to teach semantics. The science of grammar does not teach the communication of meaning that is already known from ordinary usage; rather, it teaches correct usage in the conveyance of the desired meaning. In his very first vārtika,[18] commented upon at length by Patañjali in the *Paspaśāhnika*, Kātyāyana places the function of grammar in the context of what is already known from ordinary behavior. There is an established relation between words and their objects, which is known from ordinary usage, such that certain words are used to denote certain objects. The purpose of using speech forms is to convey knowledge of objects by following the conventions of ordinary

---

[2] Rules 1.4.99-108 that designate verbal and nominal terminations in the lists 3.4.78 and 4.1.2 by terms that allow selection according to person, number, and voice are not shown.

[3] The rule that deletes markers, 1.3.9, is shown here though its application is simultaneous with the introduction of affixes.

usage. Since this is the case, the purpose served by the science of grammar is to make known which speech forms among those in use are correct and hence lead to dharma. Kātyāyana states:

> Siddhe śabdārthasambandhe lokato 'rthaprayukte śabdaprayoge śāstreṇa dharmaniyamaḥ, yathā laukikavaidikeṣu.[4]
> Since speech, its object, and the relation between the two are established (and are known) from ordinary usage, and since one uses speech prompted by meanings in accordance with ordinary usage, the science (of grammar) restricts (usage to correct speech forms) for the sake of dharma just as (other disciplines restrict behavior) in ordinary and Vedic affairs.[5]

D.   Semantics

While it is obviously correct that the *Aṣṭādhyāyī* does not include any rules that are concerned with semantics to the exclusion of syntax, morphology, and phonology, the system of rules clearly presupposes that semantics drive the derivation. Meaning is the reason for speech. Under 1.1.44, Patañjali describes that the purpose of speech is to convey understanding:[19]

> The use of words is for the purpose of the comprehension of the objects they denote. With the intention, "I will give the understanding of an object" a word is used.[6]

Modeling the fact that a speaker selects speech forms to use on the basis of the meaning he wishes to convey, the *Aṣṭādhyāyī* is composed in a manner that selects certain speech forms for use on the basis of certain semantic conditions. Specific semantic factors pervasively serve as conditions for the classification of lexical items, and for the introduction of kāraka terms, cover symbols, and speech forms.

1.   Lexical organization

The use of words in rules to refer to classes of words rather than just to their own speech form is discussed in the *Mahābhāṣya* under 1.1.68 *svaṁ rūpaṁ śabdasyāśabdasañjñā*.[20] The word *vṛkṣa* 'tree', etc. in 2.4.12 *vibhāṣā*

vṛkṣamṛga... refers to terms for species of trees.[7] The word *sva* 'property', etc. in 3.4.40 *sve puṣaḥ* refers to itself as well as to its synonyms,[8] while the word *rājan* in 2.4.23 *sabhā rājāmanuṣyapūrvā* refers to its synonyms but not to itself. Finally, the word *matsya* in 4.4.35 *pakṣimatsyamṛgān hanti* refers to itself as well as to terms for species of fish. The use of words in the grammar to refer to classes of words rather than to the speech forms themselves succeeds through the intermediary of the words' meaning, against the norm in the grammar for words to refer just to their own form. By referring to their meaning, in the way words are ordinarily used, the meaning of the word can serve as the condition to class groups of words of related meaning.

There are 735 words used in the locative to state semantic conditions in rules (including repetitions and excluding individual compound elements).[21] Conditions that serve to classify lexical items include place (*deśa*),[9] district (*janapada*),[10] river (*nadī*),[11] mountain (*parvata*),[12] measure (*parimāṇa*),[13] genus,[14] species,[15] or ethnicity (*jāti*),[16] age (*vayas*),[17] fish (*matsya*), and conscious being (*cittavat*),[18] among others.

2.   Semantic conditions for kārakas, cover symbols, and phonetics

It is well known[22] that the terms *dhruva* 'fixed point', etc. in rules 1.4.24-55 *dhruvamapāye 'pādānam*, etc. serve as

---

[7] *sittadviśeṣāṇāṁ vṛkṣādyartham* vt. 5, K1.176.25. The scheme of distinguishing the ways in which words are used to refer to various classes of words or to themselves proposed in vārtikas 5-8 is not adopted in the *Aṣṭādhyāyī*. It nevertheless illustrates these various usages in the grammatical treatise.

[8] *pitparyāyavacanasya ca svādyartham.*

[9] *deśa* 3.3.78, 4.2.52, 4.2.67, 4.2.119, 5.2.105, 5.2.135, 6.3.98, 8.4.9; *adeśa* 8.4.24.

[10] 4.2.81, 4.2.124.

[11] 4.2.85.

[12] 4.3.91.

[13] 4.3.153, 5.2.39.

[14] *jāti* 4.1.161, 5.2.133; *ajāti* 5.4.37, 6.4.171.

[15] 6.3.103.

[16] 6.2.10.

[17] *vayas* 3.2.10, 4.1.20, 5.1.81, 5.2.130, 5.4.141, 6.2.95; *avayas* 5.1.84.

[18] 5.1.89.

---

[4] K1.6.8.

[5] Scharf 1995.

[6] *Arthagatyarthaḥ śabdaprayogaḥ. Arthaṁ sampratyāyayiṣyāmīti śabdaḥ prayujyate.* K1.105.2.

semantic conditions for the introduction of kāraka terms, and[23] that terms such as *bhūta* 'past', *vartamāna* 'present', and *bhaviṣyat* 'future', used in the locative in 3.2.84 *bhūte*, 3.2.123 *vartamāne laṭ*, and 3.3.3 *bhaviṣyati gamyādayaḥ*, serve to introduce *l*-affixes. Houben (1999: 46) has illustrated the direct use of semantic and pragmatic factors as conditions for phonetic modifications to strings in the section of rules 8.2.82-108.[24] Such factors conjoin with the syntactic condition, specified in the heading to the section, 8.2.82 *vākyasya ṭeḥ pluta udāttaḥ*, that the string be a sentence (*vākya*).

3.   x-vacana

A number of rules explicitly use the term *vacana* 'denoting' to designate the semantic conditions that serve as the criteria to class together words that denote entities in major categories.[25] Hence semantic conditions serve to form a class of words that denote entities other than substances (*asattvavacana*),[19] a class of words that denote qualities (*guṇavacana*),[20] a class of words that denote common properties (*sāmānyavacana*),[21] or distinguishing properties (*viśeṣavacana*),[22] and a class of words that denote the essence (*bhāva*) of what is denoted by the stem after which certain affixes forming such words occur (*bhāvavacana*).[23]

Similarly, other rules explicitly use the term *vacana* to designate the semantic conditions that serve as the criteria to form narrower classes of lexemes subject to common operations. Hence in one rule semantic conditions serve to form classes of indeclinables that denote proximity (*samīpa*), flourishing (*samṛddhi*), lack of prosperity (*vyṛddhi*), absence of an object (*arthābhāva*), going beyond (*atyaya*), unsuitability for the moment (*asaṁprati*), the appearance of a sound or word (*śabdaprādurbhāva*), posteriority (*paścāt*), a meaning of *yathā*, sequence (*ānupūrvya*), simultaneity (*yaugapadya*), similarity (*sādṛśya*), success (*sampatti*), completeness (*sākalya*), end (*anta*), and senses denoted by nominal terminations and other affixes provided by rules 5.3.1-26 (*vibhakti*).[24] In other rules the term *vacana* designates classes of words that denote remembrance (*abhijñā*),[25] stages of bodily growth (*vayas*),[26] haste (*kṣipra*),[27] wish (*āśaṁsā*),[28] boundary (*maryādā*),[29] imagination or supposition (*sambhāvana*),[30] fitness (*paryāpti*),[31] and half (*sāmi*).[32] In commenting upon several of these rules, the *Kāśikā* notes that the term *vacana* is used to include synonyms of the word that precedes it in compound.[33]

Elsewhere the term *vacana* explicitly designates the semantic condition for a particular triplet of nominal terminations, secondary affix, or finished form (*nipātana*). Such semantic conditions include master (*īśvara*),[34] virgin (*apūrva*),[35] momentary (*ādyanta*),[36] particular sort or manner (*prakāra*),[37] extolled (*prakṛta*),[38] and dependent

---

[19]   2.3.33   *karaṇe   ca stokālpakṛcchrakatipayasyāsattvavacanasya.*
[20] 2.1.30 *tṛtīyā tatkṛtārthena guṇavacanena*, 4.1.44 *voto   guṇavacanāt*,   5.1.124 *guṇavacanabrāhmaṇādibhyaḥ karmaṇi ca*, 5.3.58 *ajādī guṇavacanād eva*, 6.2.24 *vispaṣṭādīni guṇavacaneṣu*, 8.1.12 *prakāre guṇavacanasya.*
[21]   3.4.5   *samuccaye   sāmānyavacanasya*, 8.1.73 *nāmantrite samānādhikaraṇe sāmānyavacanam.*
[22] 8.1.74 *vibhāṣitaṁ viśeṣavacane bahuvacanam.*
[23] The term *bhāvavacana* occurs in three sūtras: 2.3.15 *tumarthācca bhāvavacanāt*, 2.3.54 *rujārthānāṁ bhāvavacanānāmajvareḥ*, 3.3.11 *bhāvavacanāśca*, and the term *bhāvakarmavacana* in one: 6.2.150 *ano bhāvakarmavacanaḥ.*

[24]   2.1.6   *avyayaṁ vibhaktisamīpasamṛddhivyṛddhyarthābhāvātyayā sampratiśabdaprādurbhāvapaścādyathānupūrvyaya ugapadyasādṛśyasampattisākalyāntavacaneṣu.*
[25] 3.2.112 *abhijñāvacane lṛṭ.*
[26]   3.2.129   *tācchīlyavayovacanaśaktiṣu   cānaś*, 5.1.129 *prāṇabhṛjjātivayovacanodgātrādibhyo 'ñ*, 6.3.85 *jyotirjanapadarātrinābhināmagotrarūpasthānavarṇ avayovacanabandhuṣu.*
[27] 3.3.133 *kṣipravacane lṛṭ.*
[28] 3.3.134 *āśaṁsāvacane liṅ.*
[29] 3.3.136 *bhaviṣyati maryādāvacane 'varasmin*, 8.1.15   *dvandvaṁ rahasyamaryādāvacanavyutkramaṇayajñapātrapray ogābhivyaktiṣu.*
[30] 3.3.155 *vibhāṣā dhātau sambhāvanavacane 'yadi.*
[31] 3.4.66 *paryāptivacaneṣvalamartheṣu.*
[32] 5.4.5 *na sāmivacane.*
[33] Under 3.2.112, 3.3.133, 5.4.5 the *Kāśikā* states: *vacanagrahaṇaṁ paryāyārtham.*
[34] 2.3.9 *yasmādadhikaṁ yasya ceśvaravacanaṁ tatra saptamī.*
[35] 4.2.13 *kaumārāpūrvavacane.*
[36] 5.1.114 *ākālikaḍādyantavacane.*
[37] 5.3.23 *prakāravacane thāl*, 5.3.69 *prakāravacane jātīyar*, 5.4.3 *sthūlādibhyaḥ prakāravacane kan.*
[38] 5.4.21 *tatprakṛtavacane mayaṭ.*

(*tadadhīna*),[39]   The term *vacana* also designates a broad class of semantic conditions that serve as conditions for the formation of tṛtīyā-tatpuruṣa compounds. These include additional significance such as praise or censure (*adhikārtha*).[40]

E. Ontology

In addition to various specific semantic factors that serve as conditions for the classification of lexical items, and for the introduction of kāraka terms, cover symbols, and speech forms, the *Aṣṭādhyāyī* incorporates certain ontological presuppositions. The grammar presupposes a certain structure in the semantic field in order to operate properly. Rules have been formulated with certain conceptions regarding the nature of things in mind. Numerous passages in Patañjali's *Mahābhāṣya* analyze such presuppositions, as do the works of later philosophers of language from Bhartṛhari (5th century C.E.) to Kauṇḍabhaṭṭa and Nāgeśa (seventeenth and eighteenth centuries). Patañjali, for instance, has his interlocutors asks questions concerning the nature of action, time, and change in the course of their arguments about the formulation and scope of rules. They ask:[26][27][28]

> What do you consider action to be when you say, "The term *dhātu* doesn't apply to the roots *as* (class 2), *bhū* (class 1), and *vid* (class 4)."?
> *kāṁ punaḥ kriyāṁ bhavān matvāhāstibhavativid yatīnāṁ dhātusaṁjñā na prāpnotīti.* (1.3.1, vt. 5. K1.258.8-9)

> What do you consider time to be when you say, "The rule doesn't make sense because the object denoted by the word with which the word for time is compounded is not what gets measured."
> *kaṁ punaḥ kālaṁ matvā bhavān āha kālasya*

> *yena samāsas tasyāparimāṇitvād anirdeśa iti* (2.2.5, vt. 1. K1.409.21-22)

> What do you consider change to be when you say, "It doesn't work (the taddhita suffix doesn't apply) in the case of *bali* and *ṛṣabha*."?
> *kam punar bhavān vikāraṁ matvāha balyṛṣabhayor na sidhyati.*
> 5.1.13 K2.342.16)

Examination of the *Aṣṭādhyāyī* itself reveals that it presupposes a certain ontology. Substances (*dravya*), qualities (*guṇa*), and actions (*kriyā*) are distinguished as are time (*kāla*), the divisions of time past (*bhūta*), present (*vartamāna*), and future (*bhaviṣyat*), and the degrees of proximity in time near (*āsanna*), today (*adyatana*), and not today (*anadyatana*).   Number (*saṁkhyā*) is recognized.   Common properties (*sāmānya*) are recognized, as is also essence (*bhāva*). Much of this ontology subsequently appears as categories in the Vaiśeṣeṣika system of philosophy.

The various ontological categories refered to in the *Aṣṭādhyāyī* serve as the conditions that characterize sets of speech forms.   Speech forms are subject to various operations on the condition that they do or do not denote a certain entity in a certain ontological category. The semantic condition is frequently placed in the locative.   For example, 5.4.11 *kimettiṅavyayaghād āmv adravyaprakarṣe* provides a suffix *ām* (*āmu*) to a stem ending in a comparative and superlative affix *tara* or *tama* on the condition that the excellence to be denoted is not located in a substance (*dravya*).   Similarly, the speech forms in the list beginning with *ca* are termed *nipāta* if they do not denote a substance (*sattva*).[41]   They are subsequently termed indeclinable (*avyaya*).[42]   Other ontological categories that serve as semantic conditions in the locative include time (*kāla*),[43] and essence (*bhāva*).[44]

---

[39] 5.4.54 *tadadhīnavacane.*
[40] 2.1.33 *kṛtyairadhikārthavacane.* The *Kāśikā* comments, "The expression of additional meaning is the expression of the superimposed meaning connected with praise or censure." (*stuti-nindā-prayuktam adhyāropitārtha-vacanam adhikārtha-vacanam*).   In   2.3.46 *prātipadikārthaliṅgaparimāṇavacanamātre prathamā*, the term *vacana* is taken by commentators to denote number rather than to refer to reference explicitly, i.e. it is not the case that the rule provides as a condition for the occurrence of a first-triplet nominal termination merely the denotation (*vacana*) of measure (*parimāṇa*), gender (*liṅga*), and the meaning of the stem (*prātipadikārtha*).

[41] 1.4.57 *cādayo 'sattve.*
[42] 1.1.37 *svarādinipātam avyayam.*
[43] 2.3.64, 5.3.15.
[44] 3.1.107, 3.3.18, 3.3.44, 3.3.75, 3.3.95, 3.3.98, 3.3.114, 3.4.69, 4.4.144, 6.2.25. As a Buddhist, it is natural for Jayāditya to avoid accepting essence as the meaning of the word *bhāva.* Jayāditya understands the root *bhū* to refer to generic action (*kriyāsāmānya*); hence he takes the term *bhāva* to refer to the generic action common to the meaning of any root. In the *Kāśikā* under 3.3.18 *bhāve*, he states *kriyāsāmānyavācī bhavatiḥ*, following Patañjali's   statement   *kṛbhvastayaḥ*

F. Challenges to unidirectionality

Although the *Aṣṭādhyāyī* does not provide explicit rules exclusively regarding semantics, the fact that it does incorporate extensive organization of the semantic field is significant. It is particularly significant that the organization of the semantic field is carried out in part on the basis of reference to syntactic and morphological elements. Such elements are generally introduced subsequently to and on the basis of semantic conditions. Hence, the organization of the semantic field by reference to syntactic and morphological elements challenges the assertion that a hierarchy of levels is unidirectional.[29]

1. x-arthe

In the level hierarchy articulated by Kiparsky (2002), Pāṇini employs elements at levels two and three to specify semantic criteria at level one. Twenty-five of the 735 words that specify semantic criteria employ the term *artha* 'meaning' in order to specify semantic conditions on level one on the basis of morphosyntactic elements at level two and morphological elements at level three.[45] [30] In one case, an abstract morphological element on level two is employed to specify a semantic item on level one that serves as a semantic condition for another abstract morphological element at level two.[31] 3.4.7 *liṅarthe leṭ* provides that in Vedic the abstract morphological element *leṭ*

occurs in the meaning of the abstract morphological element *liṅ*. In this case, the rule that assigns abstract tense incorporates conditions only at the levels from which and to which it maps; it thereby accords with the general restriction that rules incorporate conditions only at the levels from which and to which they map.

The remaining 25 rules containing words ending in the term *artha* that specify semantic criteria violate the enunciated condition that rules incorporate conditions only at the levels from which and to which they map, as well as at prior levels in the unidirectional hierarchy beginning with semantics and ending with phonology. They incorporate conditions at level three that specify semantic criteria at level one, two levels prior in the unidirectional hierarchy. Two examples suffice to demonstrate the problem.[32] 1.1.19 *īdūtau ca saptamyarthe* provides that the sounds *ī* and *ū* occurring in the meaning of the seventh vibhakti (*Kāśikā: saptamyarthe vartamānam*) in the *Padapāṭha* are termed *pragr̥hya* and therefore do not undergo sandhi. The rule thereby specifies a semantic element, the meaning of the seventh vibhakti, at level one on the basis of items termed the *seventh vibhakti*, namely morphological elements *i os su*, at level three. The semantic condition in turn specifies a phonological trait, the absence of sandhi, at level four. Similarly, 3.4.9 *tumarthe sesenaseasen...* specifies several affixes that occur in the same meaning as the meaning of the infinitival affix *-tum* (*Kāśikā: tumuno 'rthas tumarthaḥ*). The rule thereby employs a morphological element *-tum* at level three to characterize a set of semantic conditions at level one, which then conditions allomorphs *-se*, *-sen*, etc. at level four.

The first example supports the criticism of earlier versions of the levels theory already articulated by Houben (1999) that it did not permit semantic factors to serve as conditions in phonological rules directly. 1.1.19 provides just what was not permitted: the semantic condition consisting of the meaning of the seventh vibhakti inhibits sandhi. The present version of the levels theory accommodates this criticism by

---

kriyāsāmānyavācinaḥ (K2.144.20, K2.47.24, etc.). Since the affixes provided under the heading of 3.3.18 occur after roots, which denote action, the *bhāvavacana* words refered to in 3.3.11 would denote generic action *kriyāsāmānya* even if the term *bhāva* did refer to essence; the common property in all action is the essence of action. A long tradition of comment on the meaning of the term *bhāva* determines that it denotes non-vibratory action (*aparispandamāna-kriyā*) when it specifies the condition for nominal affixes.

[45] *saptamyarthe* 1.1.19, *caturthyarthe* 1.3.55, *tr̥tīyārthe* 1.4.85, *mātrārthe* 2.1.9, *anyapadārthe* 2.1.21, *cārthe* 2.2.29, *caturthyarthe* 2.3.62, *liṅarthe* 3.4.7, *tumarthe* 3.4.9, *kr̥tyārthe* 3.4.14, *matvarthe* 4.4.128, *dhātvarthe* 5.1.118, *vidhārthe* 5.3.42, *jīvikārthe* 5.3.99, *śakyārthe* 6.1.81, *tadarthe* 6.1.82, *nityārthe* 6.2.61, *atadarthe* 6.2.156, *atadarthe* 6.3.53, *īṣadarthe* 6.3.105, *anyadarthe* 6.4.60, *śakyārthe* 7.3.68, *upamārthe* 8.2.101, *kr̥tvo'rthe* 8.3.43, *adhyarthe* 8.3.51.

permitting rules to incorporate factors at any prior level in the hierarchy as conditions. An additional problem not previously articulated, however, plagues the present version of the levels theory: rules incorporate factors at subsequent levels of the hierarchy as conditions at prior levels.

In Kiparsky's hierachy of levels, the meaning of the seventh is at a prior level of derivation to the seventh triplet of nominal terminations (*saptamī vibhakti*). One would have to run through the hierarchy to level three to get the seventh triplet terminations in order to establish the semantic range of the meaning of the seventh triplet at level one.

It is not licit to dismiss the problem by claiming that the use of the term *artha* serves merely to state synonymy at levels two or three and does not involve mapping to the prior semantic level. As Houben (1999) has reiterated, Pāṇini does not state rules that operate exclusively on the semantic level. Yet, as I have demonstrated above, Pāṇini does incorporate organization of the semantic level in his rules. The organization of the semantic level is achieved in part by reference to syntactic and morphological criteria. Since syntactic and morphological criteria serve to express the structure of the semantic level, subsequent levels of the hierarchy, including the morphological level, which is two levels removed, serve as conditions for prior levels.

2.   x-vacana

In two cases of the use of the term *vacana*, the semantic condition that serves to characterize a set of speech forms is specified by reference to levels considered to be subsequent to the semantic level in the hierarchy of four levels proposed by Kiparsky and Staal. In 6.2.150 *ano bhāvakarmavacanaḥ*,[33] the kāraka term *karman* designates a class of items that serve as the semantic conditions that characterize a set of speech forms. In accordance with this rule, a subsequent compound element (*uttarapada*) that meets three conditions has its final vowel high-toned. The three conditions are the following: 1. it ends in an affix of the form *ana*; 2. it denotes

non-vibratory action (*bhāva*) or a direct object (*karman*); and 3. it is preceded by a compound element denoting a kāraka. The fact that a kāraka is referred to as the direct object of the root *vac* in the term *vacana* is significant. It indicates that Pāṇini considered kārakas to be denotable just as purely semantic conditions are denotable.

In 2.1.6,[34] one of the semantic conditions that serves to characterize a class of indeclinables is itself characterized by morphological criteria. The rule provides that indeclinables that occur in a number of senses combine with subsequent elements to form avyayībhāva compounds. The senses specified include those denoted by nominal terminations and other affixes provided by rules 5.3.1-26 (*vibhakti*). Hence the morphemes that constitute vibhaktis serve to characterize the semantic conditions under which certain indeclinables are used. Morphological criteria therefore serve as the grounds for the organization of semantics which was considered a prior level in the hierarchy proposed by Kiparsky and Staal. Note that the adoption of cyclicity in the formation of the compounds in question does not escape the problem of counterdirectionality in the hierarchical ordering. Regardless of whether rules that generate compounds and their accentuation occur subsequent to rules that generate their compound elements, the indeclinable that constitutes the prior element of the avyayībhāva compound must have access to the morphological level even before the question of its entering into a compound arises. Indeclinables are classed according to semantic criteria that are themselves specified by morphological units.

3.   Avoidance of circularity

Although the seventh vibhakti arises subsequently to its semantic conditions, yet it can serve as the criterion to characterize its semantic conditions without resulting in circularity much in the way circularity is avoided by invoking the fact that speech is eternal.[35]

The fact that speech is eternal is invoked by Kātyāyana and Patañjali to

solve the problem of circularity in the use of terms such as *vṛddhi*. In the process of the generation of the speech form *mārṣṭi* (3sa pre) from the root *mṛj* 'wipe', the *ṛ* of the root is replaced by *ā* in accordance with 7.2.114 *mṛjer vṛddhiḥ*, utilizing the term *vṛddhi*.[46] The sounds *ā*, *ai*, and *au* are termed *vṛddhi* in accordance with 1.1.1 *vṛddhir ād aic*. The problem is raised that 7.2.114 will be ineffective because the term *vṛddhi* can only apply to an *ā* that already exists; yet the *ā* in *mārṣṭi* doesn't exist when the rule applies. Since 1.1.1 terms sounds *ā* that already exist *vṛddhi*, and 7.2.114 creates the *ā* in *mārṣṭi* by using the term *vṛddhi*, the rules are mutually dependent, the grammar involves circularity and fails.[47]

Under 1.1.45 *ig yaṇaḥ samprasāraṇam*, Patañjali discusses a similar situation in the case of the reference of the term *samprasāraṇa* to sounds *i*, *u*, *ṛ*, and *ḷ*. There a concept is introduced that is passed over in the discussion of the term *vṛddhi* under 1.1.1: the concept of a future term (*bhāvinī saṃjñā*). The term *samprasāraṇa* could be used to refer to the samprasāraṇa sounds *i*, *u*, *ṛ*, and *ḷ* that will be brought into existence. The analogy is made to a customer who approaches a weaver, hands him some thread and asks him to weave him a saree. Since the saree doesn't exist until after the threads are woven together, and one doesn't undertake the act of weaving on an already complete saree, the weaver understands that the customer uses the term *saree* as a future term: it refers to that which will be a saree once it has been woven.[48] A more familiar contemporary example might be the use of the term *cake* in the sentence, "Bake a cake." One bakes the ingredients that will be a cake once baked; one does not put a finished cake in the oven to bake.

The future term explanation, however, is superseded, in the discussion of the term *samprasāraṇa* under 1.1.45,[49] in favor of another that is spelled out in greater detail in the discussion of the term

vṛddhi under 1.1.1.[50] Under 1.1.1, Kātyāyana and Patañjali conclude that the procedure of the grammar succeeds because speech is eternal (*nitya*).[51] The speech form *mārṣṭi* already exists, and the term *vṛddhi* refers to the *ā* in it that already exists. The objection is then raised that if speech forms are eternal and forms such as *mārṣṭi* already exist, there would be no purpose served by rule 7.2.114, which formally creates such speech forms, nor would there be any purpose served by generative grammar generally. This objection is met by reiterating that the rule prevents one from understanding that *mṛj*, without vṛddhi, is correct everywhere; it instructs that the correct form is *mārj*, before affixes not marked with *k* or *ṅ*. Since speech is eternal, the grammar serves the purpose, not of generating speech, but of restricting usage to correct versus incorrect speech forms.[52]

The apparent circularity in the case of semantic conditions that are defined in terms of speech forms can be solved in a similar way. The previous section pointed out that it is circular to define a semantic condition (e.g. *saptamyartha*) in terms of speech forms (e.g. *saptamī vibhakti*) that are generated by rules that include those semantic conditions. The circularity is avoided by understanding that the relationship between speech forms and their meaning is eternal. The rules do not actually generate the speech forms in certain meanings; they instruct one that it is correct to use certain speech forms in certain meanings. The linguistic description of the relation between the seventh vibhakti and its meanings is therefore timeless and legitimately referred to at any point in a derivation.

The terms *cake* and *saree* must be understood to refer to cakes and sarees generally, and to particular instances of cakes and sarees still to be produced, in order for ordinary affairs to be conducted successfully. These terms are so understood because the relation between speech forms and their meanings is virtually constant in the linguistic

---

[46] The *ā* is then followed immediately by r in accordance with 1.1.51 *uraṇ raparaḥ*.
[47] K1.40.18-21.
[48] K1.112.9-14.
[49] K1.112.14-17.

[50] K1.40.26-1.41.4.
[51] vt. 9. *siddhaṃ tu nityaśabdatvāt.* K1.40.26.
[52] vt. 10. *kimarthaṃ śāstram iti cen nivartaktvāt siddham.* K1.41.1.

community. Similarly, for the successful procedure of the grammar, the terms *vṛddhi* and *saṃprasāraṇa* must be understood to refer to the sounds *ā, ai,* and *au,* and *i, u, ṛ,* and *ḷ,* respectively, even to particular instances of them that have not been generated by the formal procedure of the grammar. The terms are so understood because the grammar, although generative in form, is understood as instruction concerning the correct usage of a language that is virtually constant in the linguistic community. Likewise for the successful procedure of the grammar, the term *saptamī* in 1.1.19 must be understood to refer to a certain triplet of nominal terminations, even if the formal procedure of the grammar has not yet generated those nominal terminations, so that the term *saptamyartha* can be understood to refer to the semantic conditions for the occurrence of certain speech forms. The metalanguage used in the grammar must be available to the mechanics of the grammar at the time of procedural implementation of rules that use it, just as language is understood by people in the conduct of ordinary affairs.

Circularity is avoided in the use of speech forms to define semantic criteria that condition those speech forms because speech is eternal, and its relationship to meaning is established. Hence the meaning of the seventh vibhakti is known even before any particular derivational sequence is exhibited.

C.   Kārakas

As early as 1964, R. Rocher (1964: 51) criticized the characterization of kārakas as syntactic categories, instead arguing that they are semantic. Calling them syntactico-semantic, Cardona (1976: 215-224) countered that it is suitable to consider kārakas as a level between the purely semantic level and the level at which nominal terminations are introduced (the abstract morphological level in Kiparsky 2002) because the rules that introduce kāraka terms include both semantic and co-occurrence conditions.

It is certainly the case that co-occurrence conditions enter into *kāraka* classification rules, and therefore that the

kāraka classification is an intermediate stage of derivation between that of semantic conditions and that of the introduction of nominal terminations. It is possible that such an intermediate stage serves merely the purpose of procedural economy and does not imply that *kāraka* classification constitutes a level in any psychological or structural sense. Pāṇini may conceive of just two levels: semantic (*artha*) and phonetic (*śabda*). *Kārakas* are objects intended in certain relations; the level of intention is that of meaning, that is, the semantic level. One prominent seventeenth century Indian philosopher of language seems to favor the conception of kārakas as semantic categories. Kauṇḍabhaṭṭa in the "Subarthanirṇaya" of his *Vaiyākaraṇa-bhūṣaṇa-sāra* speaks of basic meanings for *kārakas.* He describes the rules that do not mention syntactic conditions as circumscribing general semantic domains for them. Yet the fact that Pāṇini formulated rules categorizing certain semantic items under certain syntactic conditions in exception to these domains may capture the conception, held by speakers of the language, of such categories as natural groups. Whether this sort of conceptualization comprises a level between the semantic and the morphological, or whether all conceptualization by virtue of being conceptual is semantic, is a moot point from the point of view of Pāṇinian procedure. In Pāṇinian procedure, kāraka classification does occupy an intermediate stage between purely semantic conditions and the introduction of speech forms. The intermediate stage is a way of achieving a complex mapping between meaning and speech.

While procedurally kāraka rules intervene between semantics and phonetics, they involve both semantics and co-occurrence conditions themselves and thereby include within them semantic and phonetic parameters. From a psychological or structural perspective, therefore, they constitute a mixture of levels rather then an intermediate level.

D.   *L*-affixes

In their description of levels, Kiparsky and Staal place *l*-affixes at the same level as kārakas. Kiparsky (2002: 3) describes

"Assignment of kārakas (Th-roles) and of abstract tense" as the function of the first set of rules mapping the semantic level to the morphosyntactic level. The treatment of *l*-affixes by Pāṇini, however, differs markedly from the treatment of kārakas. Kārakas are semantic objects classified by being designated by terms (*sañjñā*). Section 1.4 classifies semantic objects intended to be expressed by a speaker in relational categories by calling them by a kāraka term. Speech forms are subsequently introduced under the condition that an item designated by a kāraka term is to be denoted. *L*-affixes, in contrast, are introduced under semantic and syntactic conditions, just as other affixes are, and then are replaced by morphological elements; they serve therefore as abstract morphological elements themselves rather than as morphosyntactic representations.[53] Kiparsky differentiates abstract morphological representation from morphosyntactic representation. Therefore, if *l*-affixes belong to abstract morphological representation and kārakas to morphosyntactic representation, it is incorrect to assert that they occupy the same level in Pāṇinian grammar.

Part of the motivation for assigning *l*-affixes to the level of morphosyntactic representation and their replacements *tip*, *tas*, *jhi*, etc. to the level of abstract morphological representation is to place the basic set of verbal terminations and the basic set of nominal terminations at the same level in the hierarchy and thereby to achieve parallelism between them. Just as the basic nominal terminations *su*, *au*, *jas*, etc. are distributed over semantic and syntactic conditions including kāraka and number, the basic verbal terminations *tip*, *tas*, *jhi*, etc. are distributed over the same conditions kāraka and number, and similar conditions such as person (*puruṣa*). Kiparsky (2002: 3) calls the rules that achieve this distribution 'morphological spellout rules'. 3.4.78 *tiptasjhi...* introduces the basic set of verbal terminations just as 4.1.2 *svaujas...* introduces the basic set of nominal terminations. These sutras are read in conjunction with restrictive rules

(*niyama*) that achieve the proper distribution over the conditions of number (1.4.21-22),[54] person (1.4.105-108),[55] and kāraka (pāda 2.3 for nominal terminations, and 1.3.13-93 for verbal terminations).

However, the parallelism is incomplete. The verbal terminations introduced by 3.4.78 are not distributed over the conditions of time and mood as the nominal terminations introduced by 4.1.2 are distributed over kārakas. On the contrary, it is rather the *l*-affixes introduced by 3.2.110 *luṅ*, 3.2.111 *anadyatane laṅ*, etc. that are distributed over time and mood. Moreover, *l*-affixes are distributed over certain kāraka conditions: 3.4.69 *laḥ karmaṇi ca bhāve cākarmakebhyaḥ* accounts for the distribution of *l*-affixes over karman or bhāva depending upon whether the root after which the *l*-affix occurs is transitive (*sakarmaka*) or intransitive (*akarmaka*). Verbal terminations, including the so called basic verbal terminations, are morphophonemic replacements of the *l*-affixes. On the grounds of the parallelism between *l*-affixes and basic nominal terminations, in addition to the fact that they, like the basic nominal terminations *su*, *au*, *jas*, etc. are initially introduced items rather than replacements, *l*-affixes, rather than the so called basic verbal terminations *tip*, *tas*, *jhi*, etc., would properly be placed at the same level as basic nominal terminations in Kiparsky's fourfold hierarchy of levels.

Basic verbal terminations *tip*, *tas*, *jhi*, etc. are therefore simply morphophonemic modifications of the *l* in *l*-affixes, just as, for example, the imperative terminations *tu*, *tām*, *antu*, etc. are further morphophonemic modifications of the so-called basic verbal terminations *tip*, *tas*, *jhi*, etc. and just as *ina*, *āt*, and *sya* (introduced after *a*-final stems by 7.1.12 *ṭāṅasiṅasām ināṭsyāḥ*) are morphophonemic modifications of the basic nominal terminations *ṭā*, *ṅasi*, and *ṅas*.

---

[53] Cardona (1997: 496) calls them "abstract affixes".

[54] 1.4.21 *bahuṣu bahuvacanam.* 1.4.22 *dvyekayor dvivacanaikavacane.*

[55] 1.4.105 *yuṣmady upapade samānādhikaraṇe sthāniny api madhyamaḥ.* 1.4.106 *prahāse ca manyopapade manyater uttama ekavac ca.* 1.4.107 *asmady uttamaḥ.* 1.4.108 *śeṣe prathamaḥ.*

E. Abstract morphology versus phonology

The claim that the phonological output form resides on a different level from the abstract morphological representation is problematic. The abstract morphological representation often appears unchanged as the final phonological output, without having been subject to any additional rule. Many of the so-called basic verbal terminations, which Kiparsky placed on the level of abstract morphological representation (and which the last section argued are simply morphophonemic modifications of *l*-affixes) occur as the final phonological output of present active and imperfect middle and passive indicative verb forms in many contexts. The affix *-tas* for example, appears unchanged in *bhavatas* (3da pre of *bhū*) before *t* or *th*. In the example *devadatta odanaṁ pacati* discussed in section IVA above,[12-16] the affix *-ti* in *pacati*, remains unchanged except for the dropping of the marker *p*. Basic nominal terminations often appear unchanged in final output form in many contexts. For example, the affix *-bhis* appears unchanged in *mālābhis* (f3s of *mālā*) before *t, th*. Can a string be on a different level from itself? In what sense of 'level' is this permissible? Note that Kiparsky (2002: 49) states that his scheme of levels "makes no distinction between 'phonology', 'morpho-phonology', and 'allomorphy'."

Now one can certainly argue that the choice of the particular abstract morphological representation is arbitrary and that it is just coincidental that in some cases the final output is identical to it. It is quite possible that one could select an abstract representation that never appears as phonological output. This is precisely what the previous section argued is the situation with the *l*-affixes. *L*, with various markers, is the abstract morphological representation of all verbal terminations. At least one stage of replacement for *l* always occurs to get the final output form of a verbal termination, whereas for nominals it is not necessarily the case that any additional stage occurs. Stages of replacement vary greatly in the production of speech forms; there is no clear association between those stages and any psychological or conceptual

level. Three stages of replacement occur in the derivation of the form *bhavantu* (3pa ipv of *bhū*). (1) The *l* of *loṭ* is replaced by *jhi* by 3.4.78 *tiptasjhi...* (2) The *i* of *jhi* is replaced by *u* by 3.4.86 *er uḥ*. (3) The cover symbol *jh* is replaced by *ant* after *a*-final stems by 7.1.3 *jho 'ntaḥ*. Are we to posit three levels to correspond to these three stages of derivation? At least the use of the cover symbol *jh* achieves a valuable generalization in unifying the verbal terminations of the third person plural. Are we to posit an additional level at which such generalizations achieved by the use of cover symbols of this kind reside? The use of such cover symbols achieves an economy of rules in comparison to replacement of part or all of one basic termination that appears in phonetic output by sounds that appears in phonetic output in other contexts.[56] The use of *l*'s is essentially no different. If positing separate levels for cover symbols and their replacements is not procedurally justified, then what is the justification for positing separate levels for *l*-affixes and the basic verbal terminations that initially replace them? A twentieth century conception of syntax?

In distinction to potentially multiple stages of affixes and their replacements, it seems to me that just one level is involved once an affix has been introduced. The fact that Pāṇini uses the technique of replacement for the derivation of the final output form from an abstract morphological representation indicates that the replacement is considered to belong to the same level rather than to a different one; it belongs to the morhpophonemic level as opposed to the semanticosyntactic level.

The semantic and syntactic levels are properly coalesced in a semantico-syntactic level and the abstract morphological and the morphophonemic levels are properly coalesced in a single morphophonemic level. While Pāṇini derives forms through numerous un-correlated stages of derivation, he makes

---

[56] Cardona (1997: 330-332) discusses cover symbols and (490-492) demonstrates the economy of the inclusion of the cover symbol *jh* in the basic verbal terminations.

a clear distinction between the level of meaning and the level of speech.

The concept of levels in Pāṇinian grammar, and the hierarchy of four levels proposed by Kiparsky and Staal, was inspired by divisions that evolved in modern linguistics. It is anachronistic to read them into the *Aṣṭādhyāyī*. Kiparsky himself (2002: 2) hedges his attribution of levels to Pāṇini calling them, "what we (from a somewhat anachronistic modern perspective) could see as different levels of representation." Pāṇini's grammar certainly worked with two levels: meaning and speech. Its derivational procedure certainly included more than two stages. However, it appears forced to press the derivational stages into a conceptual hierarchy of levels between the purely semantic and the purely phonetic, particularly into a four-level hierarchy corresponding to modern linguistic divisions.[57] Attempting to isolate syntax from semantics in the field of linguistics parallels the attempt to isolate relations from terms in modern logic. Both are as indefensible as the isolation of forces from particles in classical physics has proven to be.[58]

In describing Pāṇinian procedure, one must be clear about when one is superimposing conceptions from contemporary linguistics on Pāṇini. Likewise, in modeling Pāṇinian procedure one must be clear about when one is introducing contemporary computational procedures foreign to Pāṇini. In the next section, I describe the organization of Pāṇinian grammar, purely from a Pāṇinian perspective rather than from the perspective of modern theoretical linguistics. In the remainder of this paper, I differentiate computational implementations of Pāṇinian grammar that model Pāṇinian procedure from applications of non-Pāṇinian generative computational techniques to Sanskrit.

V. Sketch of an overview of Pāṇinian architecture

The grammar is set up to derive correct speech forms from an open lexicon under certain conditions. The usual conditions are semantic, i.e. that certain meanings are to be denoted. Occasionally, conditions include pragmatics and literary context. In general, therefore, the grammar derives speech forms from meaning rather than vice versa. The grammar is not organized to determine the meaning of statements; it proceeds from the speakers point of view, not from the listeners point of view. It answers the question, "How do I say x?," not the question, "What does x mean?"

A. Introduction of basic elements on semantic conditions

In general Pāṇinian grammar introduces basic speech elements, or morphological elements, under semantic conditions. Basic speech elements include roots, nominal bases and affixes. Roots are introduced in two ways:

(1) Elements listed in the dhātupāṭha are termed roots (*dhātu*) by rule 1.3.1 *bhūvādayo dhātavaḥ*.

(2) Derived elements terminating in any of a series of affixes introduced in rules 3.1.5-31 are termed roots by rule 3.1.32 *sanādyantā dhātavaḥ*.

Nominal bases are likewise introduced in two ways:

(1) Any meaningful element other than a root (*dhātu*), affix (*pratyaya*), or an element that terminates in an affix, whether listed or not, is termed a nominal base (*prātipadika*) by 1.2.45 *arthavad adhātur apratyayaḥ prātipadikam*.

(2) Derived elements, including both those terminating in affixes termed *kṛt* or *taddhita* and compounds (*samāsa*), are termed nominal base by 1.2.46 *kṛttaddhitasamāsāś ca*.

Affixes are introduced by rules in adhyāyas 3-5 governed by the heading 3.1.1 *pratyayāḥ*. These include affixes in

---

[57] Hyman (2003: 188-89) argues that Herodian's recognition of three types of linguistic errors-- namely, barbarism, solecism, and acyrologia-- corresponds to the threefold distinction of phonology, morphosyntax, and semantics.

[58] See W. V. O. Quine, "Two Dogmas of Empiricism, 2d, 1961 "The statement, rather than the term, came with Frege to be recognized as the unit accountable to an empiricist critique." http://www.ditext.com/quine/quine.html.

the list designated by 3.3.1 *uṇādayo bahulam*.

The basic speech elements of the grammar do not constitute a fully specified set of elements. First, lists are not specified as part of the ruleset; they are specified by commentators subsequently, which leaves open to doubt which items were intended to be included by the author of the rules himself. Second, the grammar includes recursive procedures. The derivates of certain procedures serve as conditions for other procedures which in turn serve as conditions for the first procedures. The derivational procedure permits the derivation of nominal bases from roots and other nominal bases, and the derivation of words from roots and nominal bases. The derivation procedure also permits the derivation of roots from roots, roots from nominal bases, roots from nominal words, and nominal bases from words.

Aside from lists being in doubt and the presence of recursive derivation of elements, the set of basic elements is an open set since what is classed as a nominal base includes any meaningful element outside of a specified set. 1.2.45 reads, "any meaningful element other than ... is a nominal base." Moreover, commentators call many of the lists of nominal bases merely paradigmatic (*ākṛtigaṇa*) rather than complete. Finally, the fact that by 3.1.8-11 verbal roots are derived from an unspecified set of nominal words (*pada*), which are in turn derived from the open set of nominal bases, makes verbal roots an open set as well.

Now, nominal bases are explicitly stated to be meaningful, and affixes are introduced under semantic conditions. While no statement of the grammar introduces underived roots under semantic conditions, and the *Dhātupāṭha* list did not originally include semantic designations for them, they are assumed to be meaningful elements from the outset. Roots and nominal bases enter the grammar as that after which affixes are provided under specified conditions, prevalently including semantic conditions.

B.   Phonological modification

Once basic elements have been introduced in chapters 3-5 of the *Aṣṭādhyāyī*, they are subject to morphophonemic operations taught in chapters 6-8. Introduced elements are subject to augmentation, and they (1.1.55 *anekālśit sarvasya*) and their parts (1.1.52-54 *alo 'ntyasya*, etc.) are subject to replacement, and deletion. Some replacements have the status of their substituends (1.1.56 *sthānivad ādeśo 'nalvidhau*); others don't. Some types of affix-deletion (*luk, lup, ślu*) negate operations conditioned by the affix (1.1.63 *na lumatāṅgasya*); others (*lopa*) don't (1.1.62 *pratyayalope pratyayalakṣaṇam*).

Some of the operations that occur on introduced speech forms are cognizant of morpheme boundaries; others are not. Operations that are cognizant of morpheme boundaries take place on stems (*aṅga*) before affixes, on affixes after stems, or at word (*pada*), or sentence (*vākya*) boundaries, or to other entire meaningful units (*sarva*). Some take place only word-final (*padāntasya* 8.4.37, 8.4.59) or only not word-final (*apadāntasya* 8.3.24, 8.3.55). The section of rules beginning with 6.4.1 *aṅgasya* and ending at the close of the seventh adhyāya recognize stem-affix boundaries. Rules in the sections beginning with 8.1.16 and 8.3.55 through the end of the third pāda of the eighth adhyāya are cognizant of pada boundaries. Rules 8.1.1-15 apply to entire meaningful units. Operations on introduced speech forms that are not cognizant of morpheme boundaries take place in continuous speech (*saṁhitā*) with no conditions other than phonetic context. Such rules are relatively few. Augmentation with *t* (*tuk*) and general vowel sandhi rules occur in the section following 6.1.72 *saṁhitāyām*, and general consonant sandhi rules occur at the end of the last pāda of the eighth adhyāya, beginning with 8.4.40 *stoḥ ścunā ścuḥ*. The sparsity of rules that are incognizant of morpheme boundaries testifies to the great extent to which semantosyntactic conditions pervade morphophonemic operations.

VI. *Pāṇinian procedure versus non-Pāṇinian generation*

In order to illustrate the difference in approach required to create a computational model of Pāṇinian grammar as opposed to generating speech forms computationally without regard to Pāṇinian procedure, a few examples of how rules would be formulated under each approach are provided in the following sections. One example concerns the implementation of sandhi; two others concern nominal inflection and verbal inflection, respectively.

A. Sandhi

Without regard to Pāṇinian procedure, yet producing results consistent with Pāṇinian description, one could generate interword sandhi by constructing sandhi tables like the vowel sandhi table shown in [36]. Rules would then be written simply to replace the left context, shown in green in the top row, and occasionally the right context, shown in green in the right column, by the contents of the cell indexed by a cell in the top row and a cell in the right column. Items in blue show single replacements for both left and right contexts. Items in parenthesis show replacements just for the right context, and items in black italics show replacements for just the left context.

In contrast, to model Pāṇinian procedure requires creating data structures and a framework that allow one to approximate the statement of Pāṇinian rules in an executable language. Scharf (1992) wrote a Pascal program that executes sandhi between words and compound elements and presented the implementation at the 44th Annual Meeting of the Association for Asian Studies. In 2002, Scharf and Hyman designed a portable framework using modified regular expressions in an XML file to model Pāṇinian rules.[37] Each rule is written as one or more XML *rule* tags each of which contains several parameters: *source*, *target*, *lcontext*, *rcontext*, *optional*, and *c*. The optional parameters *lcontext* and *rcontext* specify the left and right contexts for the replacement of the *source* by the *target*. The optional parameter *optional* specifies that the current state is to be duplicated and subsequent parallel paths created, one in which the rule is implemented and the other in which it is not. The parameter *c* (for *comment*) contains the number of the Pāṇinian rule implemented by the *rule* tag. While most rules are implemented in a single *rule* tag, 6.1.101 requires five *rule* tags to implement. The implementation utilizes the Sanskrit Library Phonetic encoding scheme SLP1, in which Sanskrit sounds and common phonetic features such as tones and nasalization are each represented by a single character.[59]

The rule syntax utilizes a number of macros that model Pāṇinian structures. Macros are used to model Pāṇinian sound classes: varṇa, varga, guṇa, vṛddhi, saṁprasāraṇa, etc.; to create pratyāhāras: *ak*, *aṇ*, *ik*, *yaṇ*, etc.; and to group sounds with common phonetic features: aspirated sounds, unaspirated sounds, voiced sounds, unvoiced sounds, etc. For example, the macros @(f) and @(x) in 1.1.9 vt. represent the varṇas *ṛ* and *ḷ* respectively. The macros @(eN) in 6.1.109 and @(ac) in 6.1.78 represents the pratyāhāras *eṅ* (monothongs) and *ac* (vowels), respectively. Mappings are used to map sets of sounds onto corresponding sounds, such as short vowels onto long, and unvoiced stops onto voiced stops. Functions, such as lengthen, guṇate, and vṛddhiize, utilize the mappings to facilitate implementation of common operations, namely, the replacement of a vowel by its corrresponding long vowel, guṇa vowel, or vṛddhi vowel, respectively. The functions lengthen, vfddhiize, and guRate are utilized in 6.1.101, 6.1.88, and 6.1.87, respectively. Their parameter ($1) is a regular expression reference to the contents of the source parameter that appears in parenthesis. In accordance with 1.1.51 *uraṇ raparaḥ*, the later two functions include the provision of *r* after replacement of the vowel *ṛ* by its corresponding guṇa vowel *a*.

Rules are not pre-selected by hand; rather they are triggered by the data that meets the conditions for the application of the rule. Hyman wrote a Perl program that converts the XML file of regular expressions to Perl executable code. The

---

[59] http://sanskritlibrary.org/encoding/SLP1.pdf

model succeeds in encoding Pāṇinian rules in a manner that allows the rules that come into play to be tracked. Rule tracking has valuable research and pedagogical applications. Hyman (2007) describes the procedure by which the XML vocabulary to express Pāṇini's sandhi rules was developed and how a series of stages converts the rules not only into executable perl code, but also into a network, and a finite state transducer. The latter, being extremely fast, will permit realtime web use of the models.

B. Nominal inflection

Similar to the way in which an external sandhi table can be implemented without regard to Pāṇinian procedure, one could generate nominal declension without regard to Pāṇinian procedure, yet produce results consistent with Pāṇinian description. If one considers a nominal paradigm, such as that of the masculine noun *deva* in [38], it is evident that the element *dev* remains constant, while the remainder of the word varies in the paradigm. One can extract a set of endings proper to *a*-final masculine stems that consists of the varying segments.[39] One can then draft a rule by which one deletes the final *a* of the stem in any *a*-final masculine nominal and adds the *a*-final stem terminations to generate the stem's full declension.

Similarly, one can extract a set of endings proper to *jan*-final masculine stems by segmenting the string *jan*, which is constant in the paradigm of *rājan*, from the endings that vary.[40] One can then draft a rule by which one deletes the final *an* of the stem of any *jan*-final masculine nominal and adds the *jan*-final stem terminations to generate the full declension of any *jan*-stem masculine nominal. A similar procedure allows one to draft a rule for masculine stems ending in *C[vm]an*, where *C* is any consonant.[41]

While this procedure, used by Scharf and Cheifetz in 1995, achieves a computational implementation of nominal declension, it fails to capture the generalization inherent in the Pāṇinian analysis that posits a basic set of nominal terminations for all nominal declension.[42] Instead of one basic set of nominal terminations, one requires multiple sets of terminations each proper to a specific stem type.

In contrast, the XML data structures utilized in the last section to model Pāṇinian sandhi can be augmented to allow derivation of nominal stems.[43] Scharf and Hyman implemented Pāṇinian nominal derivation by introducing an additional parameter *morphid* in the XML rule tag and utilizing Scharf's (2002: 29-30) set of nominal inflection tags. In this implementation, rules are grouped in rulesets given a *name* parameter that specifies three or more stages in derivation, including changes to terminations, changes to stems, and sandhi. These rulesets are further grouped to apply to stems that match gender and phonological parameters.

While the method adopted succeeds in producing nominal paradigms utilizing rules that capture what Pāṇinian rules do, it is limited in the extent to which it models Pāṇinian procedure. Pāṇinian rules form a single cascade and are selected solely by data that meets the conditions of the rule. The XML nominal declension procedure just described, on the other hand, selects stems for sets of rules selected in advance by hand for their known application to stems that meet the ruleset's selection criteria. Therefore, although the implementation utilizes a single set of basic terminations, and hence is an advance over the procedure that relies on multiple sets of nominal terminations, it is not a close model of Pāṇinian procedure.

C. Verbal inflection

Verbal inflection can also be implemented by extracting multiple sets of terminations from various paradigms just as was done for non-Pāṇinian nominal declension. If one considers a verbal paradigm, such as the present active indicative of the root *bhū* in [44], one can segment the invariant string *bhav* from the variant strings *ati*, *atas*, etc.[45] One can extract a set of endings proper to *a*-final present stems such as *bhava*, and draft a rule for the derivation of any *a*-final present stem: delete the final *a* and add the *a*-stem terminations.

Similarly, consider the paradigm of the root *rudh*.[46] One can extract a set of active endings proper to class 7 present *dh*-final stems in which there is a preceding *r* or *ṣ*. One would have to segment only *ru* as the invariant string and infer endings *ṇaddhi, nddhas, ndhan*, etc. One could then draft the rule: delete the final sound of the root and add the endings for class 7 *dh*-final stems with preceding *r* or *ṣ*. Consider then the paradigm of *yuj*.[47] One would have to infer a separate set of endings, *nakti, ṅktas, ñjanti*, etc. for *j*-final class 7 stems. Just as in the similar approach for nominal declension, this procedure fails to capture the generalization inherent in the Pāṇinian analysis. Pāṇini posits a single basic set of verbal terminations for all verbal declension.[48] Instead of one basic set of verbal terminations, the non-Pāṇinian approach requires numerous sets of terminations each proper to one among many very specific stem types.

Scharf and Hyman successfully modeled Pāṇinian verbal conjugation by further enriching the XML structure utilized for nominal declension. They added two parameters to the rule tag: *lexid*, and *root*. The former allows reference to the class of the root in the Pāṇinian *Dhātupāṭha*. The latter allows reference to the original form of the root even when the previous rules have modified the input string.

The parameter *morphid* utilizes Scharf's (2002: 30-31) verbal inflection tags. Rules are implemented in a single cascade that applies to all strings. Rule selection is solely on the basis of the data meeting the conditions of the rule, just like Pāṇinian rules. This implementation of verbal conjugation succeeds in achieving the Pāṇinian generalization of utilizing a single set of basic verbal terminations for all verbal stems, as shown in [49]. The implementation of verbal conjugation also surpasses the implementation of Pāṇinian nominal declension in that the verbal conjugation succeeds in adequately modeling Pāṇinian procedure.

The current implementation of verbal conjugation does have some limitations, however. It relies on intermediate stems extracted from Whitney's *Roots* for all but perfect (*liṭ*) and aorist optative (*āśīr-liṅ*) verb forms, implementing only stem-conjugation for the bulk of tenses and moods. For the latter two tenses and moods, however, the computational implementation approximates Pāṇinian procedure fairly closely for the derivation of final forms directly from Pāṇinian basic elements alone. As in the implementation of Pāṇinian nominal inflection, the implementation of Pāṇinian verbal inflection includes rule tracking so that a derivational history of the form can be provided.

We look forward to utilizing the enriched framework in a revised, more faithful model of Pāṇinian declension. We are currently enriching the XML tagset further to allow derivation of participle stems and hope to go on to implement derivational morphology generally.

D.   Concluding remarks

Modeling Pāṇinian derivational procedure not only provides useful research and pedagogical tools such as derivational rule histories for derived forms. More importantly, attempting to work out details of a computational implementation of Pāṇinian generative procedure illuminates the understanding of Pāṇini's method. Understanding Pāṇini's method better contributes to the improvement of linguistic methodology generally. Working out models of Sanskrit generative grammar also has direct benefits for Indological studies by bringing computational methods to assist philological work and other humanistic pursuits related to India, and by bringing Indology into the field of digital humanities.

Bibliography

Bharati, Akshar, Vineet Chaitanya, and Rajeev Sangal. 1995. *Natural Language Processing: A Paninian Perspective.* New Delhi: Prentice-Hall of India.

Cardona, George. 1976. *Pāṇini: A Survey of Research.* The Hague: Mouton.

Chomsky, Noam. 1957. *Syntactic Structures.* The Hague: Mouton.

Houben, Jan. 1999 [2001]. "'Meaning statements' in Panini's grammar: on the purpose and context of the Astadhyayi." *Studien zur Indologie und Iranistik* 22: 23-54.

Hyman, Malcolm. 2003. "One word solecisms and the limits of syntax." *Syntax in Antiquity*, ed. Pierre Swiggers and Alfons Wouters, 179-192. Orbis Supplementa 23. Monographs published by the International Center of General Dialectology, Louvain. Leuven; Paris; Dudley, Mass.: Peeters.

Hyman, Malcolm. 2007. "From Pāṇinian Sandhi to Finite State Calculus." Proceedings of the First International Sanskrit Computational Linguistics Symposium, 29-31 October 2007, Rocquencourt, France, pp. 13-21.

K *The Vyākaraṇa-mahābhāṣya of Patañjali.* Ed. Lorenz Franz Kielhorn. 3 vols. Third edition revised and furnished with additional readings references and select critical notes by K. V. Abhyankar. Pune: BORI, 1962, 1965, 1972. Reprint: 1985.

Kiparsky, Paul. 2002. "On the Architecture of Pāṇini's Grammar." Paul Kiparsky's Home Page. http://www.stanford.edu/~kiparsky/

Kiparsky, Paul and J. F. Staal. 1969. "Syntactic and semantic relations in Pāṇini." *FL* 5: 83-117.

Rocher, Rosane.1964. "'Agent' et 'objet' chez Pāṇini." *JOAS* 84: 44-54.

Scharf, Peter. 1992. "Sanskrit Sandhi from Pāṇini to Pascal." Paper presented at the 44th Annual Meeting of the Association for Asian Studies, 2-5 April 1992, Washington, D.C.

Scharf, Peter. 1995. "Early Indian Grammarians on a speaker's intention," *Journal of the American Oriental Society* 115.1: 66-76.

Scharf, Peter. 2002. *Rāmopākhyāna--the Story of Rāma in the Mahābhārata: An Independent-study Reader in Sanskrit.* London: RoutledgeCurzon.

Scharf, Peter. 2005. "Pāṇinian accounts of the Vedic subjunctive: leṭ *kṛṇvaíte.*" *Indo-Iranian Journal*, 48.1: 71-96. (Paper presented at the 214th Meeting of the American Oriental Society, 12-15 March 2004, San Diego, California.) [The publication is marred by publisher errors and omissions; correct version: http://www.language.brown.edu/Sanskrit/ScharfSubjunctive.pdf]

Scharf, Peter. forthcoming. "Pāṇinian accounts of the class eight presents." Paper presented at the 13th World Sanskrit Conference, 10-14 July 2006, Edinburgh.