Ṛgvedic *śáktīvant-*: Accentuation and Statistical Modeling of Allomorph Selection in Vedic *-mant*/*vant*-stems

RYAN SANDELL

University of California, Los Angeles

1. *śáktīvant-*: some formal peculiarities

The adjectival secondary derivative *śáktīvant-* 'able, powerful' has two tokens in the Ṛgveda:[1] it appears once as a vocative *śáktīvaḥ* in the midst of the Indra cycle in Book V, at 5.31.6c, and a second time as a nominative plural *śáktīvantaḥ* in the final *Anhangslied* of Book VI, at 6.75.9b. At first glance, the morphology of this stem is completely transparent: it is a derivative in *-vant-*, of which the Ṛgveda alone attests more than 200 distinct types,[2] apparently built to the nominal stem *śaktí-*/*śákti-* 'ability, skill', which is in turn a feminine abstract noun derived with the primary suffix *-ti-* built to the root √*śak* 'be able'. Although the segmental components and compositional semantics of *śáktīvant-* entirely accord with this derivational analysis, *śáktīvant-* is noteworthy in several respects:

1. Although its base *śaktí-*/*śákti-* is attested in the RV with both initial-syllable and second-syllable accent, the instances of *śáktīvant-* only show the high tone on the initial syllable.

2. The vowel immediately preceding *-vant-* is unexpectedly long, *ī*, where the base *śaktí-*/*śákti-* only attests short *i*.

3. Secondary adjectival derivatives to primary *ti*-stems are usually built with the suffix *-mant-*, (e.g., *ṛṣṭimánt-* 'having spears').

śáktīvant- is not merely interesting for this constellation of formal peculiarities, but also because it potentially furnishes valuable information about the

1 All Vedic forms cited in this paper are from the Ṛgveda (abbreviated RV) unless explicitly identified otherwise.

2 I count 208 stems with *-vant-* as the outermost derivational suffix in the reverse index of Grassmann 1876 [1976]:1728–30, to be precise. A considerable number of these *vant*-stems are *hapax legomena* and this derivational suffix can therefore be assessed as highly productive, following the corpus-based measurements of morphological productivity described in Baayen 1989 and validated for the RV'ic corpus in Sandell 2015a:Chap. 5.

David M. Goldstein, Stephanie W. Jamison, and Brent Vine (eds.). 2016.
Proceedings of the 27th Annual UCLA Indo-European Conference. Bremen: Hempen. ###–###.

grammar of Vedic accentuation and how the grammar and lexicon of Vedic may undergo changes with respect to accentuation. Specifically, I will show in Section 2 that the initial-syllable accent of *śáktīvant-* is compatible only with derivation from an inherently accented base /śákti-/. This fact crucially indicates that the initially accented *śákti-* itself must have been learned as possessing an inherent accent (i.e., /śákti-/), and thus demonstrates not only a surface distinction between *śaktí-* and *śákti-*, but an abstract phonological distinction as well.

Precisely because *śáktīvant-* exhibits the additional peculiarities listed above, due diligence requires that we should try to understand as clearly as possible how the form came to be, to ensure that the interpretation assigned to its accent in the preceding paragraph is not a consequence of any other process. I will examine more closely the selectional restrictions on the suffixes *-mant-* and *-vant-* in particular. In short, the questions to be addressed are the following three, in parallel to the three observations above:

1. Is the accent of nom.pl. *śáktīvantaḥ* explicable only through a derivation from /śákti-vant-/, or does its accent have another more likely source?

2. Why does the base *śaktí-*/*śákti-* select for the derivational suffix *-vant-*, rather than *-mant-*?

3. Why has the vowel preceding the suffix, *ī*, been transmitted as long, rather than as an expected short *i*?

Section 2 outlines how the accentuation of *śáktīvant-* can be understood in a model of Vedic accentuation generally. Sections 3 and 4 then address questions 2. and 3., while Section 4 argues further that independent and productive derivation is the most likely source of the nom.pl. *śáktīvantaḥ*. The Appendix presents the details of the statistical models.

2. The Accentuation of Vedic *-mant-* and *-vant-*

A cursory glance at possessive adjectives built with the parallel suffixes *-mant-* and *-vant-* might initially suggest inconsistency or irregularity in placement of the high tone: apart from disallowing the high tone to fall on the syllable corresponding to an inflectional suffix (gen.sg. *-as*, loc.sg. *-i*, etc.), *mant-* and *vant*-stems are like words of Vedic generally, in apparently permitting the high tone to surface

on any syllable.[3] In fact, the description of accentuation for these two derivational suffixes reduces to a simple set of possibilities: either the high tone remains on the same syllable on which it appears in the base of derivation, or the high tone appears on the suffix *-mánt-*/*-vánt-*. See the sets of examples in (1) and (2).

(1) Persistence of Base High Tone

a. *ávi-* 'sheep' : *ávimant-* 'having sheep'
b. *devá-* 'god' : *devávant-* 'having gods'
c. *pitú-* 'food' : *pitúmant-* 'having food'
d. *barhíṣ-* 'ritual straw' : *barhíṣmant-* 'having ritual straw'
e. *román-* 'hair' : *románvant-* 'having hair'
f. *tápas-* 'heat' : *tápasmant-* 'having heat'
g. *gó-* 'cow' : *gómant-* 'having cows'

(2) Shift of High Tone to -mant-/-vant-

a. *ṛṣṭí-* 'spear' : *ṛṣṭimánt-* 'having spears'
b. *sūnú-* 'son' : *sūnumánt-* 'having sons'
c. *ketú-* 'sign' : *ketumánt* 'having signs'
d. *asthán-* 'bone' : *asthanvánt-* 'having bones'
e. *pád-* 'foot' : *padvánt-* 'having feet'

For root nouns, whether *-mant-*/*-vant-* will host the high tone depends upon whether the high tone is mobile internal to the paradigm of that noun. For instance, *gó-* 'cow' is accentually immobile (dat.sg. *gáve*), whereas *pad-* 'foot' is accentually mobile (dat.sg. *padé*), and this difference in accentual behavior of the base nouns is reflected also in difference of accentuation of the possessive adjective: *gómant-* (1g), but *padvánt-* (2e) and *nṛvánt-* 'having men'.

Immobility of accentuation internal to an inflectional paradigm, however, does not entail persistence of the high tone in a *mant-*/*vant-*stem: for instance, *śruṣṭí-* 'willingness' maintains the high tone on the second syllable from the left edge in all of its inflected forms (acc.sg. *śruṣṭím*, inst.sg. *śruṣṭī́*, dat.sg. *śruṣṭáye*), yet cedes the high tone to *-mant-*, which yields *śruṣṭimánt-*. The accentuation of *mant-*/*vant-*stems is therefore *not* predictable from the surface accent of their

3 However, in nouns and adjectives the high tone does not appear farther from the left edge of the word than the syllable corresponding to the outermost derivational suffix, with few exceptions (e.g., genitive plurals to *i-* and *u*-stems like *matīnā́m* 'of the minds'). A very similar restriction seems to be operative in Attic-Ionic Greek of the 1st millennium BCE (cf. Steriade 2013): in polysyllabic stems high tones largely do not occur on an inflectional suffix.

bases alone: there is not a principle of surface-level transderivational accentual faithfulness at work. Instead, it is necessary to refer to underlying prosodic properties of morphemes, following Kiparsky (1984, 2010). The distinct behaviors of *pad-* and *gó-* in inflection and derivation can be handled by treating /pad-/ and /gó-/ as *underlyingly unaccented* and *underlyingly accented* morphemes, respectively, where accented morphemes attract the accent. This paper will not treat the precise mechanism by which surface high tones are determined in Vedic (see again Kiparsky 1984 and 2010, as well as Sandell 2015a:Chap. 7), but two general observations concerning accentual defaults (encoded in the accounts of both Kiparsky and Sandell) should be noted:

1. If no underlyingly accented morphemes are present, and the stem is polymorphemic, the high tone surfaces on the rightmost syllable of the stem.[4] Thus: acc.sg. /pad-vant-am/ → *padvántam*.

2. If no underlyingly accented morphemes are present, and the stem is monomorphemic, the high tone surfaces on the leftmost syllable of the stem.[5] Thus: acc.sg. /pad-am/ → *pádam*.

Note that in 1. directly above that the suffix /-vant-/ is given as inherently unaccented. However, the argument that *-mant-* and *-vant-* are themselves unaccented (contrary to Kiparksy [forthcoming:15], who assumes accented /-mánt-/ and /-vánt-/) is in fact not easily demonstrated from surface distributional evidence in derivation alone (unlike the case of /-ti-/, the unaccentedness of which can be readily inferred from the fact that it cedes the high tone to suffixes to its right). The principal suffixes added to *-mant-* or *-vant-* are the secondary comparative and superlative, *-tara-* and *-tama-*, and feminine *-ī-*. All of these suffixes are properly part of the paradigm of a *mant-*/*vant*-stem, and hence no movement of high tone from the base to those suffixes *-tara-*, *-tama-*, or *-ī-* is expected (there are, however, some exceptional cases, such as *śáśvant-* 'continual' → superlative *śaśvattamá-*). Instead, the larger behaviors of the Vedic accentual system as discussed in Sandell 2015a:Chap. 7 require the assumption of an unaccented /-mant-/ and /-vant-/. Likewise, Kiparsky's assumption of accented /-mánt-/ and /-vánt-/ is motivated by his overall analysis, rather than by decisive surface accentual distributions. The crucial point for present purposes is, given that forms

4 This "default" behavior may reflect attraction of the accent to the head of morphologically complex words, which is usually the outermost (rightmost) derivational suffix in Vedic.

5 This default behavior reflects an attraction of the accent to the left edge of the prosodic word.

such as *ṛṣṭimánt-* and *śruṣṭimánt-* show high tone on *-mánt-*, the derivational suffix /-ti-/ must be "unaccented"—otherwise, if /-tí-/ were "accented," like /gó-/, the forms ˣ*ṛṣṭímant-* and ˣ*śruṣtímant-* would be expected.[6]

With respect to *śáktivant-*, then, this adjective must be derived from the noun *śákti-*, with persistent high tone on the initial syllable (acc.sg. *śáktim*, inst.sg. *śáktī*, inst.pl. *śáktibhiḥ*), rather than the noun *śaktí-* (nom.sg. *śaktíḥ*, acc.sg. *śaktím*, inst.sg. *śaktī́*, acc.pl. *śaktī́n*). If *śáktīvant-* took /śak-ti-/ as its input for derivation, the expected output would instead have been (in terms of accentuation) **śaktīvánt-*.[7] The question of why the distinctly accented forms *śákti-* and *śaktí-* exist alongside one another in the RV cannot be addressed here (for one account specific to *ti*-stems, see Lundquist 2015), but this doublet must be an early instance of the tendency in the history of Vedic for nouns with non-initial high tone to develop parallel forms with high tone on the first syllable. The distribution of *śaktí-* and *śákti-* within the RV is chronologically coherent: *śaktí-* occurs in the Family Books (6×, Books II, III, IV, VII) and Book I (1.83, 1.109), while *śákti-* occurs in Book I (1.31) and Book X (3×).

At first glance, the attestation of *śáktīvant-*, however, does not appear to accord neatly with the chronology of attestation for *śaktí-* and *śákti-*: *śáktīvant-* occurs already in Books V and VI, where, given that only *śaktí-* appears in the Family Books, one might assume that the lexeme available was /śak-ti-/, and thus

6 Kiparsky (forthcoming:15) lists *-ti-* as an underlyingly accented dominant suffix. However, to assume that the suffix is underlyingly accented generates incorrect predictions for the accentuation of *mant*-stems built to possible *ti*-stem bases, given Kiparsky's assumption that the suffixes *-mant-* and *-vant-* are also inherently accented. Since the accentuation of *ti*-stem simplexes and their further derivatives can, within the larger context of Kiparsky's analysis of Vedic accentuation, be made to work given the assumption of an unaccented suffix /-ti-/, an unaccented suffix /-ti-/ should be posited. The RV attests eight possible *ti*-stems to which a derivative in *-mant-* is built, all of which show the high tone on *-mant-*: *abhiṣṭimánt-* 'offering help', *praṣṭimánt-* 'having a kind of horse', *puṣṭimánt-* 'rich in food', *śruṣṭimánt-* 'attentive', ṛṣṭimánt- 'having spears', bhṛṣṭimánt- 'having points', vṛṣṭimánt- 'containing rain', and *svastimánt-* 'having well-being'. Besides *śáktivant-*, two forms possibly containing *ti*-stem bases, *sáptīvant-* 'having teams' and *svádhitīvant-* 'having axes', also have high tone on the base and show an unexpected long *ī*.

7 I mark **śaktīvánt-* here with an asterisk, rather than ˣ because it is an entirely possible RV'ic form, and it would show the grammatically correct position of the high tone, given the inputs /śak-ti-vant-/. Since **śaktīvánt-* could be a grammatical form given the correct underlying inputs, explaining the accent of *śáktīvant-* will precisely require the assumption of a different input.

would generate **śaktīvánt-*.[8] The attestation at 5.31.6c, voc.sg. *śáktīvaḥ*, is, however, useless for purposes of determining the accentuation of a lexeme: as a vocative, the high tone on the initial syllable of *śáktīvaḥ* is the only option. It is possible that this vocative belongs to the unattested **śaktīvánt-*, the accentuation of which would be expected were a *vant*-stem derived from *śaktí-*. The fact that nom.pl. *śáktīvantaḥ* occurs in an *Anhangslied* 6.75 (dedicated to "Weapons") is compatible with the later attestation of *śákti-*: given that 6.75 is a later addition to Book VI, the hymn may be assumed to belong to the later chronological stratum in which only *śákti-* appears.[9] The two forms voc.sg. *śáktīvaḥ* and nom.pl. *śáktīvantaḥ* might then reflect independent derivations: the former from the unaccented /śak-ti-/ that gives the *śaktí-* of the Family Books, the latter from the accented /śákti-/ that gives the *śákti-* of Books I and X.[10] The underlying structure of nom.pl. *śáktīvantaḥ* may thus be represented as /śákti-vant-as/, where the underlying accent of the base /śákti-/ determines the position of the high tone in surface *śáktīvantaḥ*.

3. Predicting selection of *-mant-* and *-vant-*

To find a Vedic adjective in *-vant-* built to a *ti*-stem (or base terminating in the sequence /-ti-/) is somewhat surprising: at the level of raw frequency, the RV contains eight adjectives in *-mant-* built to *ti*-stems, versus only three stems in *-vant-* built to *ti*-stems (*śáktīvant-*, *sáptīvant-*, and *svádhitīvant-*); cf. n.6 above. This small group of three forms exhibiting leftmost accent and an unexpectedly long *ī* might disguise some previously unnoticed similarity. The first issue is to understand what factors might condition the selection of the allomorph *-mant-* or *-vant-* in the productive derivation of Vedic possessive adjectives to nouns.

---

8 Incidentally, neither *śaktí-* nor *śákti-* occurs in either Book V or VI.

9 In fact, the compositional history of this hymn, verses 8–11 in particular, is somewhat uncertain. Some or all of verses 8–11—our form is in verse 9—might be yet later interpolations into the hymn. See the discussion with references in Oldenberg 1909:415.

10 I write /śákti-/, without morpheme boundaries, because I assume that the learning of a different underlying prosodic property is intimately connected with a failure to learn the morphological structure underlying a surface form *śaktí-*. This change is precisely the same sort of "demorphologization" proposed by Probert (2006:291–2) to account for type variation in the accentuation of several thematic derivational suffixes in 1st-millennium Greek. On arguments for "demorphologization" in Greek and Vedic, see Sandell 2015a:192–214 and 2015b, and in Hittite, see Yates 2015.

I refer to *-mant-* and *-vant-* as allomorphs because they are semantically indistinguishable and phonologically very similar. Although they do not occur in a perfect phonologically complementary distribution (and indeed, there are a few bases such as *rayí-* 'wealth' to which both *-mant-* and *-vant-* attach), we will see that there are highly robust, regular, and predictable patterns to their distribution. In rough terms, I think that the situation of Vedic *-mant-*/*-vant-* is rather like roots subject to alternations of diphthong and monophthong in Italian, e.g., [sjɛd-]/[sed-] 'sit' in Italian (for a recent analysis, cf. Booij and van der Veer 2015): both allomorphs are stored in the lexicon, but the selection of allomorphs is phonologically motivated. To take the stem *ṛ́ṣṭimánt-* 'having spears' as an example, one can think of the input as the base stem, plus a set containing both suffix allomorphs, thus /ṛ́ṣṭ-i-{mant, vant}-/. The choice of surface allomorph and accentuation of the form (as well as any other word-level phonology) is determined simultaneously, when an optimal candidate is selected.

This investigation proceeds from the assumption that the selection of *-mant-* or *-vant-* is conditioned by properties of the base to which the derivational suffix attaches; in principle, such factors could be semantic, syntactic, morphological, or phonological. These potential predictive factors will, however, be highly collinear: the syntax and semantics (e.g., abstract noun, action noun) of the base noun are usually determined by the outermost derivational suffix of the base, which is itself a morphological factor, and the phonological form of the right edge of the stem, which will be directly local to *-mant-* or *-vant-*, is likewise a consequence of the outermost derivational suffix. Furthermore, since the exact morphological constituency of a given base is difficult to judge—for instance, although I argued above that *śákti-* is best understood as a morphological simplex, to say with great confidence whether the typical Vedic speaker understood it as a simplex *śákti-* or complex *śák-ti-* is difficult—it might be counterproductive to try to investigate the selection of *-mant-*/*-vant-* on the grounds of the bases' morphological makeup. For these reasons, the most straightforward basis on which to consider the selection of *-mant-*/*-vant-* is the phonological properties of the base, especially the segments at the base's right edge. Even assuming that the genuine conditioning factors are instead semantic, syntactic, or morphological, the phonological properties of the stem will be an effective proxy, because all of these potential predictors will be largely predictive of one another.

Indeed, no sophisticated analysis is needed to see that the rightmost segment of the base appears overwhelmingly decisive in conditioning the selection of *-mant-* or *-vant-*. See Tables 1 and 2 below, which report the type frequency of bases terminating in given segments for *-mant-*/*-vant-* stems in the RV, based on

Grassmann 1872 [1976]:1728–30, where 283 types are given. The exceptionless conditions are given in Table 1; Table 2 reports the non-exceptionless distributions of types having a given rightmost segment in a base. Even among the exceptions, however, the distributions are largely one-sided. For example, bases whose rightmost segment is *a* or *ā* exhibit a near-total (110/112 types) preference for *-vant-*, while stems in *u* have a decided tendency to prefer *-mant-* (28/29 types).

Table 1. Exceptionless relationships between final segment of base and selection of *-mant-*/*-vant-*

Rightmost segment of base	Possessive adjective suffix (# types)
-ā-	*-vant-* (59)
-o-	*-mant-* (1)
-r-	*-vant-* (4)
-ṇ-	*-vant-* (3)
-h-	*-vant-* (1)
-ś-	*-vant-* (1)

Table 2. Non-exceptionless relationships between final segment of base and selection of *-mant-*/*-vant-*[a]

Rightmost segment of base	# *-mant-* types	# *-vant-* types
-a-	2	51
-i-	19	10
-ī-	4	19
-u-	28	1
-ū-	1	1
-ṛ-	2	1
-k-	1	3
-t-	3	6
-d-	1	3
-n-	1	9
-ṣ-	8	2
-s-	1	33

a These numbers reflect the number of cases in the Saṃhitā text in which the segment immediately preceding the suffix has a given segment; these numbers are complicated by the fact, discussed below, that some of the transmitted long vowels at the right edge of the base may in actuality have been short vowels.

To explore the importance of different phonological factors in determining the selection of *-mant-* and *-vant-* in greater detail, I constructed models, the aim

of which was to predict whether a given base should take *-mant-* or *-vant-*. In investigating the impact of these phonological properties of the base on the selection of *-mant-* or *-vant-*, I used two types of predictive models (technically, supervised machine learning techniques applied to a classification problem): a Bayesian implementation of a generalized linear model (which for a binary response variable is equivalent to logistic regression) as developed by Gelman et al. (2015), and a conditional inference regression tree as developed by Hothorn et al. (2015). Further details of the models are presented in the Appendix.

Both the logistic regression model and the conditional inference tree agree that the features FS [+ROUND], 2S [+HIGH, –CONS.], FS [+HIGH], and BASE ACCENT are very significant predictors of suffix selection, and both models obtain a high degree of accuracy in prediction on the data set (above 90%). In short, a base with a [+round] final segment or a base with a [+high, –cons.] final segment (i.e., *-i-* or *-u-*) prefers *-mant-*, while bases with a final segment that is [–round] and [–high] prefer *-vant-*; even if the final segment of a base is [+high], if the accent falls on the base rather than the suffix, there is nevertheless a strong preference for *-vant-* (see Nodes 6 and 7 at the bottom of the Figure in the Appendix).

That the position of the high tone on the base strongly predicts selection of *-vant-* is directly relevant for present purposes: it validates the accent of *śáktīvant-* and suggests that it is precisely because the base is accented that we find *śáktīvant-* rather than **śáktīmant-*. Indeed, of the twenty-two forms on which the conditional inference tree makes an incorrect prediction, eight (*práṣṭimant-*, *aśánimant-*, *dhúnimant-*, *ávimant-*, *dhrájīmant-*, *vā́śīmant-*, *táviṣīmant-*, and *tvíṣīmant-*) are attributable to the fact that they are accented on the first syllable (though for one, *táviṣīvant-* is also attested).[11] The preference for base-accented forms to take *-vant-* may also help to account for one instance of variation in suffix selection. To the base *rayí-* 'wealth' both *rayimánt-* and *rayivánt-* are attested; three of the four instances of *rayivánt-*, however, are vocatives, which would

11 The reader should bear in mind that this analysis assumes that grammars are fundamentally probabilistic. The eight forms whose bases end in *-i* and are accented are in fact predicted by the model to have a ~78% chance of taking *-vant-*. It is also possible, of course, that some incorrect predictions result from a failure of the analyst to diagnose other relevant predictor variables, or from highly localized factors whose inclusion would result in models that overfit the data. As an instance of the latter, the surprising selection of *-mant-* in *yávamant-* (rather than **yávavant-*) may be due to the fact that, in three of its four occurrences, it is collocated with *gómant-*, which takes *-mant-* on account of the [+round] final segment of the base.

(without suppression of the high tone), have the accent on the base *ráyivaḥ**; neither of the instances of *rayimánt-* is a vocative. Similarly, *adrivant-* 'having pressing stones' and *vajrivant-* 'having a cudgel', which would be good candidates for selection of *-mant-* on account of the high vowel as the final segment of the base, occur exclusively as vocatives, but *-vant-* would be expected for base-accented vocatives *ádrivaḥ**, *vájrivaḥ**, *hárivaḥ**, *táviṣīvaḥ**, and *ṛ́ṣīvaḥ*. Most importantly, this allows us to make sense of the vocative *śáktīvaḥ* at 5.31.6c, in a chronological layer where /śak-ti-/ may still have been living, and, by the models implemented above, would have been more likely to give **śaktimánt-* (choosing *-mant-*)—it is precisely the initial syllable high tone in the vocative that elevates the likelihood of *-vant-* over *-mant-* in this form.

In sum, close statistical analysis has revealed that the quality of vowel at the right edge of the base and the accentuation of the base are the best predictors of allomorph selection among RV'ic *-mant*/*-vant*-stems. That allomorph selection in this case in Vedic exhibits such robust connections to phonological factors is also consistent with the claim that morphological selection takes place in the phonological component of the grammar (e.g., Wolf 2015).

4. Trying to get to *ī*

In the preceding section, I have argued that the occurrence of the suffix *-vant-* with the base /śákti-/ is, contrary to the list of "peculiarities" articulated at the outset, expected. This leaves the problem of why, given a base terminating in short *i*, the two instances of *śáktīvant-* in the RV Saṃhitā, appears as long *ī*.[12] There are two major possibilities to account for the *ī*: either it is genuine, or **sáktivant-* should be restored and the Saṃhitā's long *ī* should be assumed to have been created in the later transmission of the text. In contrast to the Padapāṭha's morphological parsing where the underlying *i* is recognized, many mantra repetitions of 6.75.9b all contain the reading with long *ī*.[13]

Arnold (1905:127), however, notes numerous instances in which the Saṃhitā text shows a long vowel preceding the suffix *-vant* where a light syllable would

12 The Padapāṭha text reads *śakti-vaḥ* at 5.31.6c and *śakti-vantaḥ* at 6.75.9b. For the similarly long *ī*'s in *sáptīvant-* and *svádhitīvant-*, the Padapāṭha correctly parses and gives short *i* as *sapti-vantaḥ* and *svadhiti-vān*.

13 Atharvaveda Paippalāda 15.10.9b, Vājasaneyī Saṃhitā Mādhyaṃdina recension 29.46b, Taittirīya Saṃhitā 4.6.6.3b, Maitrāyaṇī Saṃhitā 3.16.3b: 186.13, and Kāṭhaka Saṃhitā, Aśvamedhagrantha 6.1b. For commentary on the repetition in the Atharvaveda Paippalāda, see now Lelli 2015:135–6.

better fit the quantitative tendencies of the meter. Furthermore, he specifically identifies both instances of *śáktīvant-* and the occurrence of *sáptīvant* at 7.94.10c as candidates for the restoration of short *i*. In 6.75.9b, the position of the *ī* of *śáktīvant-* in the sixth syllable of a triṣṭubh line (*svāduṣaṃsádaḥ pitáro vayodhā́ḥ / kṛchreśrítaḥ* ***śáktīvanto*** *gabhīrā́ḥ*) strongly favors scansion as light and restoration of the short *i*. On the other hand, the position of the *ī* of *śáktīvant-* in 5.31.6c (*śáktīvo yád vibhárā ródasī ubhé*), in the second syllable of a jagatī line, would favor a heavy syllable, according to Arnold's own summary of metrical tendencies in the RV (1905:148). Oldenberg (1909:328, 416) considers Arnold's restoration of **śáktivant-* in both instances to be "doubtful." For other similar cases where Arnold proposes the restoration of a short vowel preceding *-vant* (e.g., for **áśvavant-* and **saptivant-* in RV 7.94; cf. Oldenberg 1912:63) Oldenberg likewise voices skepticism, but does not comment in any detail on Arnold's proposals.

Table 3. Frequency of occurrences of *śáktīvantaḥ* and words of the shape HLHh and HLHH in possible positions in eleven-syllable lines in the RV

	1–4	2–5	3–6	4–7	5–8	6–9	7–10	8–11
śáktīvantaḥ	0	0	0	0	1	0	0	0
HLHh	55	44	0	2	174	24	1	707
HHHh	69	3	0	1	5	1	0	4

Table 4. Frequency of occurrences of *śáktīvaḥ* and words of the shape HLh and HHh in possible positions in twelve-syllable lines in the RV

	1–3	2–4	3–5	4–6	5–7	6–8	7–9	8–10	9–11	10–12
śáktīvaḥ	1	0	0	0	0	0	0	0	0	0
HLh	23	90	40	1	171	73	1	315	0	854
HHh	118	36	157	0	16	7	7	3	0	1

Here, it is useful to apply the method proposed by Gunkel (2010:Chap. 3) for assessing metrical restorations to the Saṃhitā text based on statistical agreement or disagreement between the spelling shape of a word form and the different prosodic shapes that potentially underlie that spelling.[14] In the case of nom.pl. *śáktīvantaḥ*, the form occurs just once, falling between the fifth and eighth syllables of

14 I am very grateful to Dieter Gunkel for sharing with me the frequencies of words of given metrical shapes in the RV from his joint work with Kevin Ryan. The counts in Tables 3 and 4 are from Gunkel and Ryan's work.

of triṣṭubh line, and has a spelling shape of HHHh (that is, three heavy syllables, followed by a syllable which may be either heavy or light, depending upon the initial segment of the following word). The question is: assuming that the form genuinely is of the shape HHHh, how likely would the RV'ic poets have been to place it between the fifth and eighth syllables of a triṣṭubh line? Based on Gunkel's method, I employ Fisher's Exact Test (Fisher 1945) to assess the hypothesis that *śáktīvantaḥ* has a significantly different distribution in eleven-syllable lines from words of the shape HHHh (as the text reads) or from words of the shape HLHh (as Arnold's emendation would propose).

Fisher's Exact Test comparing the distribution of *śáktīvantaḥ* and words of the shape HLHh returns a *p*-value of 0.2986 (df = 7); comparing the distribution of *śáktīvantaḥ* and words of the shape HHHh returns a *p*-value of 0.1786 (df = 7). In neither case then can we confidently reject the null hypothesis that *śaktīvantaḥ* and words of either the shape HLHh or HHHh are drawn from the same distributions. However, the lower *p*-value of 0.1786 on the frequencies of *śaktīvantaḥ* and HHHh shaped words indicates that it is more likely that *śáktīvantaḥ* has a genuinely different distribution from words of the shape HHHh in eleven-syllable lines.[15] Overall, it is more likely that Arnold's emendation to **śáktivantaḥ* is correct than that the form was in fact of the shape HHHh.

The same statistical testing carried out for the vocative *śáktīvaḥ* against words of the shape HLh and HHh in twelve-syllable lines tells a rather different and conflicting story. Here, comparison of *śáktīvaḥ* and words of the shape HLh returns a *p*-value of 0.01657 (df = 9)—which suggests that *śáktīvaḥ* belongs to a

15 To be clear: the results of the significance testing here do not permit a firm conclusion (i.e., it is well within the realm of possibility that *śáktīvantaḥ* follows either HLHh or HHHh distributions), but to find the single token in syllables 5–8 indicates that it more likely belongs to the HLHh group (where 17% of the tokens fall in that slot) than to the HHHh distribution (where only 6% of the tokens fall in that slot). Since the form cannot be both *śáktīvantaḥ* and *śáktĭvantaḥ* at the same time in the same place, we must choose, and the choice that is better supported by the statistics is to restore **śáktĭvantaḥ* (with short *i*). In effect, we use the *p*-value here to make a binary choice, and decide to reject the null hypothesis on the lower *p*-value. This reasoning is entirely independent of the occurrence of *śákti-vantaḥ* in the Padapāṭha (cf. n.12 above), but is rather based on the seemingly straightforward analysis that *śáktīvant-* contains the stem *śákti-*/*śaktí-*. Since *śáktĭvantaḥ* is the expected result of /śákti-vant-/, it is thus the *lectio facilior*, while *śáktīvantaḥ*, conversely, being the *lectio difficilior*, could arguably be maintained on that basis. Indeed, the more conservative choice is to leave *śáktīvantaḥ*, though we cannot readily establish a reason behind the form's unexpected *ī*, bearing in mind that the statistical evidence marginally favors the *facilior* reading.

different distribution than HLh-shaped words, among which Arnold's emendation to **śáktivaḥ* would place it. Conversely, Fisher's Exact Test on comparison of *śáktīvaḥ* and HHh-shaped items returns a *p*-value of 0.5449 (df = 9), indicating that *śáktīvaḥ* might well belong to the class of HHh-shaped items.

The results here are best interpreted as inconclusive: emendation to **śákti-vantaḥ* is (slightly) better supported than reading *śaktīvantaḥ*, but the long *ī* in vocative *śáktīvaḥ* is more difficult to dismiss. But based on the data as it stands, I tentatively propose the following interpretation: Arnold's emendation to **śákti-vantaḥ* in 6.75 should be accepted, while *śáktīvaḥ* in 5.31 contains a metrical lengthening, likely due to the very fact that trisyllabic words at the opening of a twelve-syllable line are more often of the shape HHh than HLh (118 out of 141 instances, or 83.7%); Gunkel and Ryan (2011:57) state that the second position of a twelve-syllable line is occupied by a heavy syllable in 87% of instances).[16] In the transmission of the RV, **śáktivantaḥ* at 6.75 was altered to **śáktīvantaḥ* due to the morphological and accentual connection to *śáktīvaḥ* in the other (likely earlier) verse. Nevertheless, given a possibly more general tendency for the vowel /i/ to be lengthened preceding the suffix *-vant-* (50% of *vant*-stems types in the RV in which the rightmost segment of the base is presumably /i/, based on independent attestation of the base, and have the high tone on the base, present such an apparent lengthening) that the two tokens of *śáktīvant-* were each independently generated a form with long *-ī-* from /śákti-vant-/ cannot be excluded either.

16 Among the five word forms that make up 5.31.6c (*śáktīvo yád vibhárā ródasī ubhé* "when, o mighty one, you will take apart both word halves"), three words could be used to form a good HL× cadence: *śáktivaḥ* (assuming short *i*), *ródasī*, or H+*ubhé* (in which case the preceding form would likely have to be *ródasī*, because *ródasī* is never separated from *ubhé* in the RV when *ubhé* follows *rodasī*). Given that *yád* 'when' almost always occurs either clause-initially or after the first accented element, *ubhé, vibháras,* or *śáktivaḥ* would be the best candidates to occur in line-initial position, because the sequence *ródasī X ubhé* (where *X* is any word) is unattested in the RV. A heavy second syllable is strongly preferred (without exception, the second syllable of every line in 5.31 is heavy), and only *ubhé* (assuming /śaktivah/) could naturally provide a heavy second syllable; the resulting line, *ubhé yád ródasī vibharā śáktivo*, however, would have a light first syllable and a light eighth syllable, and a heavy ninth syllable —all positions where the opposite value of syllable weight is favored. In this sense, the existing line is the most optimal, in that it obtains a good rhythm, at the cost of unfaithfulness to the underlying length of just one vowel.

5. An unlikely alternative: *śáktīvaḥ* ≫ *śáktīvantaḥ*

The possibility that the vocalism of nom.pl. *śáktīvantaḥ* has been affected, either in its original composition or in later transmission, by the voc. *śaktīvo* (granting that the *ī* of the latter is not wholly secure) raises the specter of influence in terms of accentuation as well. That is, it is conceivable that the nom.pl. *śáktīvantaḥ* merely carries along the accentuation of the vocative *śáktīvaḥ*.

If such influence were exerted at the time of the original composition of 6.75, it would require that the poet of 6.75 was specifically familiar with 5.31, and that he wished to directly emulate some feature of 5.31, or that his grammar and lexicon could not productively derive *śáktīvant-*. The latter option (Pāyu Bhāradvāja could not have independently generated *śáktīvant-*) is not viable: the simplex *śákti-* was likely known to the poet (*śaktí-*/*śákti-* has a token frequency of 14 in the RV—hardly an uncommon lexeme), and the derivation of novel lexemes with *-mant-* or *-vant-* was entirely productive (such formations have a high type frequency, among which is a considerable number of hapaxes). In the absence of any evident metrical, phraseological, or thematic connection between the forms at 5.31.6c and 6.75.9b, none of which stands out to me, the more efficient hypothesis is that both occurrences of *śáktīvant-* reflect independent derivations with productive /-vant-/ to the base *śaktí-* or *śákti-*. Given the high productivity of *-vant-*, one ought to assume that a Vedic speaker could more easily apply an active morphological process (/śakti-/ + /-vant-/ → [śaktīvant-]) made available by his grammar than recall a specific instance of a specific form in order to model the creation of a new form (*śáktīvaḥ* ≫ *śáktīvantaḥ*).

On the other hand, one could assume that the original form in 6.75 was **śaktivántaḥ* (with accent on *-vant-*, assuming derivation from /śak-ti-/), and that *both* accent and vowel length were altered in later transmission under the influence of *śáktīvo*. This option is unlikely for reasons discussed in Section 3: if the underlying accentuation of the base were such that the accent would not surface on the base itself, it is more likely that the suffix *-mant-* would have been selected.[17] In effect, the expected "original" form for 6.75.9c under this scenario would have been **śaktimánt-*. In that case, the "original" **śaktimánt-* would have effectively been replaced wholesale by the *śaktīvant-* of 5.31, not only in vowel length and accent, but also in dervational suffix.

17 Specifically, the conditional inference tree trained in Section 3 predicts that an unaccented base *śakti-* would have a 21% chance of taking *-vant-*.

All in all, the easier pathway is to assume that the poet of 6.75 generated **śáktivantaḥ* with accent on the first syllable, and with the suffix *-vant-* because the lexical item with which he operated was /śákti-/, having an underlying accent on the first syllable. Possibly, later transmission of the RV altered the vowel length of the second syllable to produce the *śáktīvantaḥ* of the Saṃhitā.

6. Summary and conclusion

Although the accentuation of *śákti-* 'ability, skill' (acc.sg. *śáktim*, inst.pl. *śáktibhiḥ*), under the analyses of Vedic accentuation proposed in Kiparsky 2010 and Sandell 2015a:Chap. 7, is compatible with either an unaccented stem /śakti-/ or an accented stem /śákti-/, the accent of the adjectival derivative in a nom.pl. form *śáktīvantaḥ* proves that this form must reflect derivation from the latter, /śákti-/. Close analysis of phonological factors determining selection of the suffixes *-mant-* and *-vant-* in the RV supports a connection between accent on the base and suffix *-vant-*: because *-vant-* occurs more often on items whose bases bear the accent, while *-mant-* occurs more often on bases whose final segment is *-i-*, we may conclude that the accentuation and form of the derivational suffix in *śáktīvantaḥ* are very much intertwined. The unexpected long *ī* in *śáktīvantaḥ* may reflect a tendency for short vowels to be lengthened preceding *-vant-* in the course of the RV's transmission, as already observed by Arnold (1905:127–8). The analysis in Section 4 suggested that the long *ī* in the vocative form *śáktīvaḥ* has a claim to being original, in which case the *i* of the nom.pl. *śáktīvantaḥ* may have been altered under the influence of the vocative *śáktīvaḥ*. The ultimate conclusion here is that *śáktīvantaḥ* provides positive proof of an accentual change at the lexical level in the internal history of the RV: earlier unaccented /śak-ti-/, which surfaces with suffix accent (e.g., acc.sg. *śaktím*) to later accented /śákti-/ (e.g., inst.pl. *śáktibhiḥ*).

References

Arnold, Edward Vernon. 1905. *Vedic Metre in Its Historical Development*. Cambridge: Cambridge University Press.

Bonet, Eulàlia, Maria-Rosa Lloret, and Joan Mascaró (eds.). 2015. *Understanding Allomorphy: Perspectives from Optimality Theory*. London: Equinox.

Booij, Geert, and Bart van der Veer. 2015. Allomorphy in OT: The Italian Mobile Diphthongs. In Bonet, Lloret, and Mascaró 2015, 45–69.

Fisher, Roland Aylmer. 1945. A New Test for 2 × 2 Tables. *Nature* 156.388.

Gelman, Andrew, Yu-Sung Su, Masanao Yajima, Jennifer Hill, Maria Grazia Pittau, Jouni Kerman, Tian Zheng, and Vincent Dorie. 2015. *arm: Data Analysis Using*

Regression and Multilevel/Hierarchical Models. http://CRAN.R-project.org/web/packages/arm (9 May, 2016).

Grassmann, Hermann. 1872 [1976]. *Wörterbuch zum Rig-veda*[5]. Harrasowitz: Wiesbaden.

Gunkel, Dieter. 2010. Studies in Greek and Vedic Prosody, Morphology, and Meter. Ph.D. diss., University of California, Los Angeles.

Gunkel, Dieter, and Kevin Ryan. 2011. Hiatus Avoidance and Metrification in the Rigveda. In Stephanie W. Jamison, H. Craig Melchert, and Brent Vine (eds.), *Proceedings of the 22nd Annual UCLA Indo-European Conference*, 53–68. Bremen: Hempen.

Hothorn, Torsten, Kurt Hornik, Carolin Strobl, and Achim Zeileis. 2015. *party: A Laboratory for Recursive Partytitioning*. http://cran.r-project.org/web/packages/party/index.html (9 May, 2016).

Kiparsky, Paul. 1984. A Compositional Approach to Vedic Word Accent. In S. D. Joshi (ed.), *Amṛtadhārā: Professor R. N. Dandekar Felicitation Volume*, 201–10. Delhi: Ajanti Publications.

———. 2010. Compositional vs. Paradigmatic Approaches to Accent and Ablaut. In Stephanie W. Jamison, H. Craig Melchert, and Brent Vine (eds.), *Proceedings of the 21st UCLA Indo-European Conference*, 137–82. Bremen: Hempen.

———. Forthcoming. Accent and Ablaut. In Michael Weiss and Andrew Garrett (eds.), *The Oxford Handbook of Indo-European Studies*. Oxford: Oxford University Press.

Lelli, Duccio. 2015. The Paippalādasaṃitā of the Atharvaveda, Kāṇḍa 15: A New Edition with Translation and Commentary. Ph.D. diss., Universiteit Leiden.

Lundquist, Jesse. 2015. On the Accentuation of Vedic *-ti*-Abstracts: Evidence for Accentual Change. *Indo-European Linguistics* 3.42–72.

Manning, Christopher, and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Oldenberg, Hermann. 1909. *Ṛgveda: Textkritische und exegetische Noten. Erstes bis sechstes Buch*. Berlin: Weidmann.

———. 1912. *Ṛgveda: Textkritische und exegetische Noten. Siebentes bis zehntes Buch*. Berlin: Weidmann.

Probert, Philomen. 2006. *Accentuation in Ancient Greek: Synchronic Patterns, Frequency Effects, and Prehistory*. Oxford: Oxford University Press.

RStudio Development Team. 2009–2015. *RStudio Version 0.99.447*. https://www.rstudio.com/ (accessed 9 May 2016).

Sandell, Ryan. 2015a. Productivity in Historical Linguistics: Computational Studies in Ancient Greek and Sanskrit Word-Formation. Ph.D. diss., University of California, Los Angeles.

———. 2015b. Prosody, Parsing, and Productivity: Effects of Morphological Processing in Vedic Sanskrit and Ancient Greek. Paper presented 31 July at the 22nd International Conference of Historical Linguistics, Naples, Italy, July 31, 2015. https://www.academia.edu/15617245 (9 May, 2016).

Steriade, Donca. 2013. A Synchronic Analysis of Ancient Greek Accent. Paper presented 22 November at the Graduate School of Arts and Sciences Workshop: "Indo-European and Historical Linguistics," Harvard University Department of Linguistics.
Wolf, Matthew. 2015. Lexical Insertion Occurs in the Phonological Component. In Bonet, Lloret, and Mascaró 2015, 361–407.
Yates, Anthony. 2015. Default Accentuation in Anatolian and Its Diachronic Consequences. *Indo-European Linguistics* 3.1–43.

Appendix: predictive models used in Section 3

I present here the details of the models whose results were reported above in Section 3. Two types of classification models, a logistic regression model and a conditional inference tree, were employed to assess the effects of the phonological predictor variables listed in the following table:

Table 5. Predictor variables for *-mant-*/*-vant-* prediction models (FS = "final segment," 2S = "second-to-the-last segment")

Variable name	Variable type	Description
BASE SYLLABLES	Integer	Number of syllables of base
BASE ACCENT	Binary	1 = surface high tone falls on base
FS [–CONSONANTAL]	Binary	Rightmost segment of base is [–consonantal]
FS [+HIGH]	Binary	Rightmost segment of base is [+high]
FS [+LONG]	Binary	Rightmost segment of base is a long vowel
FS [+ROUND]	Binary	Rightmost segment of base is [+round]
FS [+CORONAL]	Binary	Rightmost segment of base is [+coronal]
FS [+ANTERIOR]	Binary	Rightmost segment of base is [+anterior]
FS [+BACK]	Binary	Rightmost segment of base is [+back]
FS [+CONTINUANT]	Binary	Rightmost segment of base is [+continuant]
FS [+NASAL]	Binary	Rightmost segment of base is [+nasal]
FS [+SONORANT]	Binary	Rightmost segment of base is [+sonorant]
2S [+COR. –ANT. –CONT. –SON.]	Binary	Second-to-rightmost segment of base is a palatal or retroflex stop (*c*, *j*, *ṭ*, *ḍ*)
2S [+HIGH –CONS.]	Binary	Second-to-rightmost segment of base is a high vowel (*i*, *u*)

All analyses were conducted in the RStudio interactive development environment for the R statistical computing language (RStudio Development Team 2009–2015). The data consisted of the 283 stems in *-mant-* or *-vant-* found in the RV, which were entered into a tab-separated values file; each datum was coded for a

binary response variable SUFFIX IS -*VANT*- (where *1* indicates that the suffix is -*vant*- and *0* indicates that the suffix is -*mant*-) and fourteen predictor variables based on manually chosen phonological properties of the base. Thirteen of these variables are binary (where *1* is "yes" and *0* is "no"), and one is an integer between 1 and 5. Responses for all predictor variables and the response variable were hand-coded by the author.[18]

The logistic regression model

For the Bayesian logistic regression model, I carried out a stepwise regression based on the Akaike Information Criterion (AIC) of each possible model in order to determine the model that provides the closest fit to the data with the least complexity (i.e., fewest number of parameters). The coefficients of this model are given in Table 6.

Table 6. Bayesian logistic regression model coefficient estimates and significance

Coefficient name	Estimate	Standard error	*t* value	Prob(>|*t*|)
(Intercept)	1.50808	0.12990	11.610	$< 2e^{-16}$
BASE ACCENT	0.19937	0.04618	4.318	$2.21e^{-5}$
FS [–CONSONANTAL]	–0.34843	0.15291	–2.279	0.023460
FS [+HIGH]	–0.23880	0.04355	–5.483	$9.54e^{-8}$
FS [+LONG]	0.10775	0.03815	2.824	0.005087
FS [+CORONAL]	–0.47096	0.12739	–3.697	0.000264
FS [+BACK]	–0.55659	0.05864	–9.492	$< 2e^{-16}$
FS [+CONTINUANT]	–0.27530	0.06525	–4.219	$3.34e^{-5}$
FS [+SONORANT]	–0.15827	0.07237	–2.187	0.029597
2S [+COR., –ANT., –CONT., –SON.]	–0.30236	0.06756	-4.475	$1.12e^{-5}$
2S [+HIGH, –CONS.]	–0.90093	0.07191	–12.529	$< 2e^{-16}$

The estimates of the coefficients give an indication of the effects of the different factors: there is a very strong baseline preference for -*vant*-, indicated by the Intercept's estimate of 1.50808, and stems that terminate in long vowels (FS [+LONG]) or which bear the high tone (BASE ACCENT) support selection of -*vant*-, as shown by the positive estimates for base accent and FS [+LONG]; all of the oth-

18 This data set and the R code used in the subsequent analyses are available online at https://github.com/rpsandell/WeCIEC27. In this data, I coded FS [+LONG] for *sáptīvant*-, *śáktīvant*-, and *svádhitīvant*- as 0 (i.e., not [+long]), because the vowel is [–long] in the base form. Where the -*mant*-/-*vant*-stem itself is only attested in a form without a high tone, namely, as a vocative, I coded the base as accented.

er parameters pull the predicted output away from the baseline preference and closer towards selection of *-mant-*, as shown by their negative estimates. The large absolute *t*-values (greater than 4.2) for BASE ACCENT, FS [+HIGH], FS [+BACK], 2S [+COR., –ANT., –CONT., –SON.], and 2S [+HIGH, –CONS.] mark those parameters as highly significant. Overall, this model performs well: it has an accuracy of 0.932, correctly predicting 264 out of the 283 items on which the model was trained, and average precision of 0.916 and average recall of 0.905.[19] The high precision indicates that the model would generalize well to new data.

The conditional inference regression tree

The conditional inference regression tree in the Figure below results from recursively building nodes in the tree based on tests of independence between the response variable and predictor variables. Where $p < 0.05$, a binary split in the tree may be implemented, and where multiple predictor variables are significant, the one with the smallest *p* value is used to implement a split. The tree fit on the predictor variables in Table 5 is given in the Figure. The nodes at the base of the tree are the responses reached from the series of inferences, in which the bar (▬) at 0 or 1 indicates as the response predicted for items that arrive at that node (0 for *-mant-*, 1 for *-vant-*; the empty circles indicate the presence of data that contradicts the prediction).

The resulting tree makes the following predictions: first, if the final segment of the base is [+round], it is predicted to prefer *-mant-* (the split on Node 1 to the prediction in Node 9); if the final segment of the base is [–round] and the second-to-the-last segment of the base is [i] or [u], it is predicted to prefer *-mant-* (the split on Node 2 to the prediction in Node 8); then, if the second-to-the-last segment of the base is not [i] or [u], and the final segment of the base is [–high], it is predicted to prefer *-vant-* (the split on Node 3 to the prediction in Node 4); finally, if the final segment of the base is [+high] and the high tone falls on the base, it

19 I converted the continuously valued predictions of the model to binary predictions by assigning a prediction of 0 (= *-mant-*) to predictions less than 0.5 and a prediction of 1 (= *-vant-*) to predictions greater than 0.5.

Average precision and recall are the average of the separate precision and recall values for predictions of *-mant-* and predictions of *-vant-*. Precision is the ratio of correct predictions on one of the codings (the *true positives* for prediction of *-mant-* or *-vant-*, respectively) to the sum of true positives and false positives; recall is the ratio of true positives to the sum of true positives and false negatives. See Manning and Schütze 1999:267–71 on these terms and their calculation.

is predicted to prefer *-vant-* (the split on Node 5 to the prediction in Node 7); else, if the base does not bear the high tone, it is predicted to prefer *-mant-* (the split on Node 5 to the prediction in Node 6). This model also performs well: it has an accuracy of 0.922, correctly predicting 261 out of the 283 items on which the model was trained, and average precision of 0.904 and average recall of 0.888.

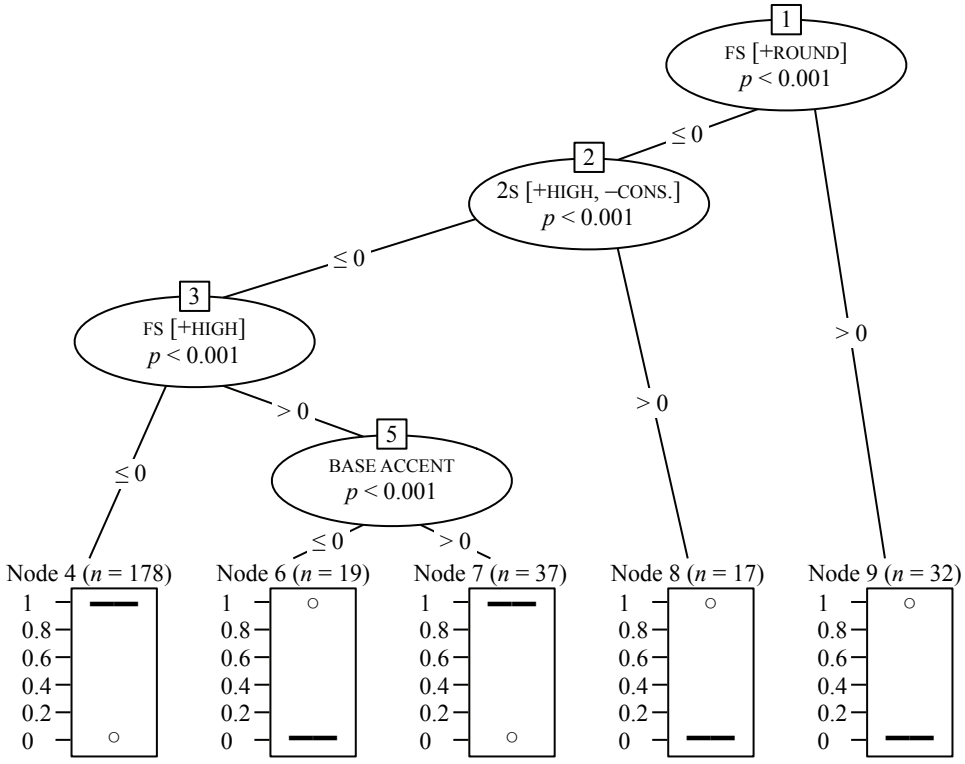

Figure. Conditional inference regression tree with four predictor variables

Overall, it is reasonable to select the conditional inference tree as a more parsimonious model, since the overall performance of the two models is nearly the same (accuracy, precision, and recall of 0.932, 0.916, and 0.905 for logistic regression vs. 0.922, 0.904, and 0.888 for the conditional inference tree), while requiring fewer predictor variables (four vs. ten). Note also that, except for the absence of FS [+ROUND] in the logistic regression model (but see above for an explanation), the two models largely agree on which predictor variables are most important: 2S [+HIGH, –CONS.], FS [+HIGH], and BASE ACCENT are used in the conditional inference tree, and (together with FS [+BACK]), have the smallest *p* values among the predictor variables in the logistic regression model.