

WORD BORDERS IN OLD INDO-ARYAN¹

According to Erhart (1990: § 3.11), word borders are realised by three (phonological) ways: a/ by the diereme (and as Erhart supposes by delimitation word accent too), b/ by realisation of phonemes and prosodemes and c/ unrealised (by the lack of delimitation).

To these three phonological means we must of course add one morphological mean d/ syntagmatical word pattern, which is constructed (maximally) by the roots, derivational endings and inflectional endings², so they logically are one word and every other root, derivational and inflectional endings are another word³.

We can also separate another morphonological mean e/ syntagmatical word-syllabic pattern. In OIA⁴ the word's initial consonantal cluster can be only ST^R_N , word's final coda can be only monoconsonantal -C (if we consider consonantal segment of diphthongs *e*, *o*, *ai*, *au* a vocal, at least syntagmatical)⁵.

All these means can obviously combine during the process of creation of word border.

The term **word border** belongs therefore to both phonology and morphology.

Word border (as morphological unit constituted by syllabic and morphological pattern) can be realised in languages really differently and for OIA is typically this lack of delimitation in so called *sandhi*, which is individual *terminus technicus* of old Indian grammarians (*sam̐dhi-*), which is used among modern linguists of the whole world⁶.

While working with vedic text we must therefore distinguish single words from each other.

The number of words we can preliminary postulate by mechanical counting of word accents in a concrete verse. The number of accents which are carefully marked must be at least equal to the number of words.

¹ I would like to thanks my friend M. Fišmeister for his help on English version of this article.

² So characterised is syntagmatical word pattern by Hjelmslev (1953: 27).

³ Compounds are marked case of unmarked (non-compound) words.

⁴ In this article we speak only about Vedic language of R̥gved.

⁵ Šefčík 2000: § 1.2-1.4

⁶ Old grammars of Vedic Sanskrit (i. e. MacDonell 1910, MacDonell 1916 etc.) distinguish between external (word to word) and internal (within one word) sandhi. We'll use the first meaning only (as usually understood among modern linguists).

I. e. the vedic verse *mitró dādhāra pṛthivīm utá dyām* 'Mitra holds Earth and Heaven' (RV III, 59, 1) we can clearly see four word accents and therefore it's likely that this verse will include at least four words, three of which are inflected (i.e. with inflectional endings). Then we will find five roots and therefore we can presume the existence of another unaccented word. Morphological analyse tells us that this unaccented word is a verb (finite verb) a inside the verse (the main sentence) the verb is always unaccented if it doesn't stand at the beginning of the verse, which is a rule (MacDonell 1910: § 85b2β).

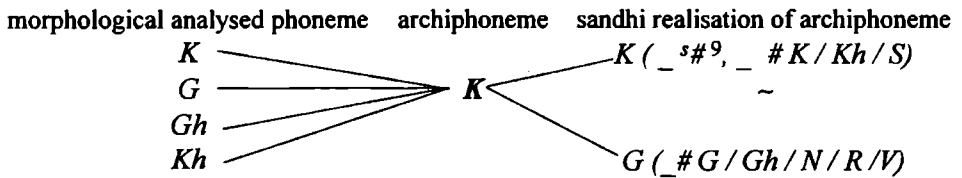
From the rule form construction of OIA word we know, that everything after the inflectional ending must already be a part of another word. Because there are clear rules for the construction of composites, it's obvious that the conjunctive particle *utá* is a separate word with zero paradigm.

OIA accent (unlike Czech, French, Polish etc.) can't help us in telling word borders, because it's free and movable within the word. It doesn't have delimitation function; we can only suppose that at least one word to word border will necessarily be between two word accents.

The sandhi principle means that the last word phonological variant is linked up in *sandhi variation* with the variants of the first following word if these words are part of one sentence (verse). The sandhi variation is clearly and regularly postulated.

The last phoneme of the last sentence's word is always neutralised to its archiphoneme⁷ before sentence's pause.

For example the morphologically justified *t* (*marút-* 'Marut'), *d* (*bhid-* f. 'destroyer'), *káprth-* m. 'penis'), *dh* (*kṣúdh-* f. 'hunger') share the archiphoneme $-T^8$ (*marút*, *bhít*, *káprt*, *kṣút*). Sentence border is a special example of the word border because there is no assimilation of archiphoneme realisation. Within the sentence the unvoiced regularly become voiced before pause if followed by voiced occlusive, nasal, semivowel or vocal (cf. our scheme).



⁷ By an archiphoneme we understand the product of syntagmatical neutralisation of phonological feature. Resulted neutralised archiphoneme is formally or functionally agreeing with the least marked member of the phonological opposition (cf. Trubetzkoy 1939: 71, Lass 1990: § 3.1-3.4).

⁸ We use these symbols: *V* for vocal, *T* for occlusive (*K* for unaspirated unvoiced *T*, *Kh* for aspirated unvoiced *T*, *G* for unaspirated voiced *T*, *Gh* for aspirated voiced *T*), *S* for sibilant, *N* for nasal and *R* for semivowel.

⁹ ie. se Sentence pause.

Similarly final *-s/ś* (final *-ś* has final archiphoneme *-k*, so as an velar occlusive!) which in the sandhi before *V, R, N, G, Gh* (see the following table) has an individual alternation replacing voiced sibilant, in particular zero (*0*) after *ā* or diphthongs *o, ar* after high vowels.

Let us compare the sandhi realisations of final phoneme of one word. As an example we'll use the number *eka-* 'one' (neutralised form of Nom. sg. is *ekah*) in the famous 'frogs hymn':

gómāyur éko ajāmāyur ékah pśsnir éko hárīta éka eṣām (RV VII, 103, 6)

'One (frog) is sounding like a cow, one is sounding like a goat; one is speckled, one of them is yellow.'

But the phonemes engage themselves also within the word on the morph borders, following the allophonic rules and necessities of syllabic pattern.

But are there some differences in the juncture of sandhi (word to word) and word-internal junction that would enable to find word borders in syntagmatic chain of allophones?

Sandhi processes were searched by Allen (1962) but he curiously doesn't solve the question of different engagement of phonemes within a word and on its borders.

Let us compare schematic patterns of the junction of phonemes within a word and on its borders. We emphasise again that all occlusives are at the end of the word neutralised on its least marked form of unaspirated unvoiced consonant. Each prove of internal junction is followed by an example from verbal flexion (engagement on borders of morphs¹⁰) without reference to the RV verse; word to word sandhi is introduced by the word's morphological forms (without sandhi changes) and concrete sandhi realisation with reference to the RV verse (transliterated from *devanāgarī* including writings of pause).

	internal junction of phonemes ¹¹		word to word sandhi junction	
T + T ¹²	K + K(h)	→ KK(h) ^(b)	K + K(h)	→ KK(h) ^(b)
	<i>vī-vac- + ti</i>	<i>vīvakti</i>	<i>(vr̥ṣā) + ajanayat +</i>	<i>vṛṣājanayattāsu</i> ¹³
	<i>tap + thās</i>	<i>tapthās</i>	<i>tāsu</i>	
	K + G(h)	→ GG(h) ^(b)	K + G(h)	→ GG(h)
	<i>vac- + dham</i>	<i>vagdham</i>	<i>r̥ṣvāt + br̥hatāḥ</i>	<i>r̥ṣvādb̥r̥hatāḥ</i> ¹⁴
	<i>cit- + dhī</i>	<i>ciddhī</i>		
	G + K(h)	→ KK(h)		

¹⁰ A root is represented in following table by its dictionary form.

¹¹ Evidences are taken from MacDonell 1916: Appendix I.

¹² Unvoiced aspirate (*Kh*) never takes a part with the consonant features in opposition of the voice and the aspiration (**KhT* or **KhS*).

¹³ RV II, 35, 13

¹⁴ RV VII, 61, 3

	internal junction of phonemes ¹¹	word to word sandhi junction
	<i>ad- + tu</i> <i>átu</i> G + G(h) → GG(h) <i>ad- + dhi</i> <i>addhí</i> Gh + K(h) → GG(h) <i>dambh- + tum</i> <i>dábdhum</i> Gh + G(h) → GGh <i>budh- + dhi</i> <i>bódhi</i>	
T + S	G + S → KS ^(b) <i>mṛc- + ṣiṣ</i> <i>mṛkṣīṣṭá</i> G + S → KS <i>chand- + si</i> <i>chántsi</i> Gh + S → KS <i>dí-dhā- + sa</i> <i>dhītsati</i>	K + S → KS <i>ariṇāt + sapta</i> <i>ariṇátsapta</i> ¹⁵
T + N	K + N ^(b) → KN <i>āp- + no</i> <i>āpnóti</i> <i>ac- + na-</i> <i>akná- (B.)</i> Kh + N → KhN <i>math- + na</i> <i>mathnántu</i> G + N → GN ^(b) <i>ad- + mi</i> <i>ádmi</i> <i>ad- + na</i> <i>ánna-</i> Gh + N → GhN <i>rdh- + no</i> <i>rdhnóti</i>	K + N → GN ^{a (+ d)} <i>havanaśrut + naḥ</i> <i>havanaśrúnno</i> ¹⁶
T + R	K + R → KR <i>kṛt- + ya</i> <i>kṛtyáte</i> Kh + R → KhR <i>manth- + ya</i> <i>mathyáte</i> G + R → GR <i>da-drā- + ur</i> <i>dadrúr</i> Gh + R → GhR <i>da-dhr- + ur</i> <i>dadhrúr</i>	K + R → GR ^a <i>srānāt + yavavantu</i> <i>srānādyavavantu</i> ¹⁷
T + V	K + V → KV <i>pat- + a</i> <i>patati</i> Kh + V → KhV <i>pruth- + a</i> <i>próthati</i> G + V → GV <i>khid- + a</i> <i>khidáti</i> Gh + V → GhV <i>jambh- + iṣ</i> <i>jámhīṣat</i>	K + V → GV ^a <i>kṛnavat + atrāthiḥ</i> <i>kṛnavadárāthiḥ</i> ¹⁸

¹⁵ RV II, 12, 3

¹⁶ RV II, 33, 15

¹⁷ RV VIII, 48, 5

¹⁸ RV VIII, 48, 3

	internal junction of phonemes ¹¹		word to word sandhi junction	
S + S	S + S <i>śiṣ-</i> + <i>sya</i> <i>śās-</i> + <i>si</i>	→ KS/(S)S <i>śekṣyāti</i> <i>śāssi</i>	S + S <i>antaḥ</i> + <i>samudre</i>	→ SS <i>antāḥ samudré</i> ¹⁹
S + T	S + K(h) <i>śās-</i> + <i>te</i> S + G(h) <i>vas-</i> + <i>dham</i> <i>si-sad-</i>	→ SK(h) <i>śāsté</i> → (G)G(h) ^a <i>saddham</i> <i>sīdate</i>	S + K(h) <i>vasubhiḥ carāmi</i> S + G(h) <i>taptāḥ + gharmāḥ</i> <i>jugupuh</i> + <i>dvaśāsya</i>	SK(h) ^b <i>vāsubhiścārāmy</i> ²⁰ → 0G(h)/rG(h) ^a <i>taptā gharmā</i> ²¹ <i>jugupurdvadaśasya</i> ²²
S + N	S + N <i>vaś-</i> + <i>mi</i>	→ SN <i>vāsmi</i>	S + N <i>naraḥ</i> + <i>na</i>	→ 0N/rN ^a <i>nāro na</i> ²³
S + R	S + R <i>aś-</i> + <i>ya</i>	→ SR <i>aśyáte</i>	S + R <i>enaḥ</i> + <i>varuṇa</i> <i>adhvasmabhiḥ</i> + <i>viśvahā</i>	→ 0R/rR ^a <i>eno varuṇa</i> ²⁴ <i>adhvasmābhirviś-</i> <i>vāhā</i> ²⁵
S + V	S + V <i>as-</i> + <i>a</i>	→ SV <i>āsati</i>	S + V <i>saḥ</i> + <i>it</i> <i>ādityaiḥ</i> + <i>uta</i>	→ 0V/rV ^a <i>sá it</i> ²⁶ <i>ādityāirutá</i> ²⁷
N + N	N + N <i>ja-gam-</i> + <i>ma</i>	→ NN <i>jagánma</i>	N + N <i>manyum</i> + <i>martyeṣu</i>	→ NN <i>manyúrṁ márt-</i> <i>tyeṣv</i> ²⁸
N + T	N + K(h) <i>tan-</i> + <i>tha</i> N + G(h) <i>yuj-</i> + <i>na</i>	→ NK(h) ^b <i>tatántha</i> → NG(h) ^b <i>yuñjáte</i>	N + K(h) <i>mahāntam</i> + <i>kośam</i> N + G(h) <i>aham</i> + <i>dadhāmi</i>	→ NK(h) ^(b) <i>mahāntam kóśam</i> ²⁹ → NG(h) ^(b) <i>aham dadhāmi</i> ³⁰
N + S	N + S <i>ram-</i> + <i>s</i>	→ NS ^b <i>āramsta</i>	N + S <i>ṛtim</i> + <i>su</i> <i>avindan</i> + <i>śiśiyāṇam</i>	→ NS ^b <i>dṛtim sú</i> ³¹ <i>avindañchiśiyā-</i> <i>nam</i> ³²

¹⁹ RV X, 125, 7

²⁰ RV X, 125, 1

²¹ RV VIII, 103, 9

²² RV VIII, 103, 9

²³ RV VIII, 103, 9

²⁴ RV VII, 86, 3

²⁵ RV II, 35, 14

²⁶ RV IV, 50, 8

²⁷ RV X, 125, 1

²⁸ RV VII, 61, 1

²⁹ RV V, 83, 8

³⁰ RV X, 125 2

³¹ RV V, 83, 7

	internal junction of phonemes ¹¹		word to word sandhi junction	
N + R	N + R <i>am- + ya</i>	→ NR <i>amyáte</i>	N + R <i>idam + vadāmi</i>	→ NR ^b <i>idam vadāmi</i> ³³
N + V	N + V <i>an- + a</i>	→ NV <i>ánati</i>	N + V <i>poṣam + eva</i> <i>jujuṣan + imāni</i>	→ (N)NV <i>póṣamevá</i> ³⁴ <i>jujuṣannimāni</i> ³⁵
R + R	R + R <i>cāy- + ya</i>	→ RR <i>cāyyáte</i>	R + R <i>viśve + rathāsaḥ</i>	→ RR <i>viśve rathāsaḥ</i> ³⁶
R + T	R + K(h) <i>ju-hu- + ti</i>	→ RK(h) <i>juhóti</i>	R + K(h) <i>asmai + tisraḥ</i>	→ RK(h) <i>asmái tistró</i> ³⁷
	R + G(h) <i>ja-bhr- + tas</i>	→ RG(h) <i>jabhrtás</i>	R + G(h) <i>asmai + dhatta</i>	→ RG(h) <i>asmái dhatta</i> ³⁸
R + S	R + S <i>bhr- + s</i>	→ RS <i>ábhārṣam</i>	R + S <i>viṣṇave + śúṣam</i>	→ RS <i>viṣṇave śúśám</i> ³⁹
R + N	R + N <i>ju-hu- + mi</i>	→ RN <i>juhómi</i>	R + N <i>śardhate + na</i>	→ RN <i>śardhate na-</i> ⁴⁰
R + V	R + V <i>jinv- + a</i> <i>i + anti</i>	→ RV ^(c) <i>jínva</i> <i>yánti</i>	R + V <i>kṛṇomi + aham</i>	→ RV ^c <i>kṛṇomyahám</i> ⁴¹
V + T	V + K(h) <i>ji- + ta</i>	→ VK(h) <i>jítá-</i>	V + K(h) <i>dr̥ṣpāya + kítavam</i>	→ VK(h) <i>dr̥ṣtvāya kítavám</i> ⁴²
	V + G(h) <i>gaccha + dhvam</i>	→ VG(h) <i>gácchadhvam</i>	V + G(h) <i>trī + dhanva</i>	→ VG(h) <i>trí dhánva</i> ⁴³
V + S	V + S <i>a + sad-</i>	→ VS <i>ásīdat</i>	V + S <i>vi + suparnaḥ</i>	→ VS <i>ví suparnó</i> ⁴⁴
V + N	V + N <i>da-dā- + mahe</i>	→ VN ^c <i>dádāmahe</i>	V + N <i>pari + naḥ</i>	→ VN ^c <i>pári no</i> ⁴⁵
V + R	V + R	→ VR ^c	V + R <i>pra + yat</i>	→ VR ^c

³² RV V, 11, 6

³³ RV X, 125, 5

³⁴ RV I, 1, 3

³⁵ RV VII, 61, 6

³⁶ RV II,, 12, 7

³⁷ RV II, 55, 5

³⁸ RV II, 12, 6

³⁹ RV I, 154, 3

⁴⁰ RV II, 12, 9

⁴¹ RV X, 125, 6

⁴² RV X, 9, 11

⁴³ RV I, 35, 8

⁴⁴ RV I, 35, 8

⁴⁵ RV II, 33, 13

	internal junction of phonemes ¹¹		word to word sandhi junction	
	<i>hū-</i> + <i>ya</i>	<i>hūyáte</i>		<i>prá yád</i> ⁴⁶
V + V	V + V <i>sthā-</i> + <i>a</i> <i>ju-hu-</i> + <i>āna</i> (<i>pat-</i>) <i>a</i> + <i>ī</i> + <i>t</i>	→ \tilde{V} /VR/RV <i>sthāti</i> <i>jūhvāna</i> <i>pátet</i>	V + V <i>yathā</i> + <i>abhavat</i> <i>manmāni</i> + <i>rcase</i> <i>uta</i> + <i>īm</i>	→ \tilde{V} /VR/RV <i>yáthābhavad</i> ⁴⁷ <i>mánmānyrcase</i> ⁴⁸ <i>utém</i> ⁴⁹

Table explanation: high indexes mean: ^a = regressive assimilation of the voice, ^b = assimilation of location, ^c = creation of 'diphthongs' *e, o, ai, au, ar, ār, an, ān, am, ām*, ^d = articulation assimilation. round brackets mean limitation of the given syntagmatical changes on some (unmentioned) examples.

From the table is clear that linked up phonemes within word are engaged in other ways than cross word border phonemes. For example within one word neither semivowels nor nasals nor vocals cause a change of the voice of occlusives and sibilants.

The differences of engagements within one word (for example between root and endings) and sandhi can be evidently be put down to the existence of segmental unit that influences phonological realisation of juncted phonemes – **diereme** (#). Erhart (1990: § 3.2) states that diereme is **segmental prosodeme** realised by pause or glottal stop⁵⁰.

Word to word diereme has an influence to allophonic realisation of phonemes (F). Its existence is marked. We can express this markedness by the following pattern:

$$F + F \quad \neq \quad F + \# + F$$

It is obvious that word to word diereme has its language function. Its existence is marked to its non-existence.

The existence of diereme in OIA doesn't necessarily have to be indicated just by the diereme's pause but also by a special allophonisation of engaged phonemes. Only when syntagmatical connected with remarkable zero of diereme the semivowels, vocals and nasals have the ability to influence the assimilation of voice. The ability to cause an assimilation change of voice of

⁴⁶ RV I, 85, 5

⁴⁷ RV X, 135, 6

⁴⁸ RV VII, 61, 6444

⁴⁹ RV II, 12, 6

⁵⁰ The whole Erhart's system of segmental and suprasegmental units of phonological plan looks as follows (Erhart 1990: § 3.4.2):

segmental phonemes	segmental prosodeme
supragmental phonemes	suprasegmental prosodeme

final occlusive or sibilant is a result of an obviously existing phoneme (i.e. *D, Dh, R, N, V*).

Diereme that stands on its own at the end of a sentence (absolute pause) works only as an element of neutralisation, which means that here is realised in the least marked form, while its realisation in the sentence is potentially more marked. Unmarked realisation of diereme at the end of the sentence is realised by a pause, inside the sentence either by the pause or an allophonic realisation of surrounding phonemes, or by combination of both.

If we wanted to characterised phonological features of diereme, we must keep in mind that at the end of the sentence the occlusives and sibilants are influenced by the neutralisation of voice and aspiration, although it can get the feature of voice within the sentence (together with the following phoneme *D, Dh, R, N, V*) secondarily and than the aforementioned phoneme from those groups of consonant's sonority that can go through this assimilation (T+ S) can assimilate by the voice too. Another phonological characters which we can consider certain for diereme is functional proximity to all consonants based on a fact that at the end of sentence of the group *a, ā + y, v* create diphthongs *e, o, ai, au* similarly as within the word before consonant (while within the word before vowel they create allophonic sequences of diphthongs *ay, av, āy, āv*).

It is obvious that when diereme is equal to zero it is something completely different than non-existence of this prosodeme. The zero of diereme is zero which is important both syntagmatically and paradigmatically.

The value variantly equal to zero is not the exception in OIA phonological system. It is also a possible variant of realisation of segmental phonemes, concretely of morphologically analysable *s* before voiced occlusive (see the table), which is realised either by gemination of occlusive (*vas-* 2. "wear": *vaddham*) or by prolongation or diphthongisation of the previous vocal: *sīdati* = *si + zd + a + ti* < *sad-*, "sit" (like *vīvakti* < *vac-* "say"); *voḍhum* = *vaz + dhum* < *vah-* "carry" + *tum*.

We could add that because the existence of diereme is demonstrable and its function is to separate words, this may be *vice versa* considered a prove that a word exists as a separate unit of *langue*.

REFERENCES

- ALLEN, W. S. 1962: *Sandhi. The Theoretical, Phonetic and Historical Bases of Word-Junction in Sanskrit*. 'S-Gravenhage: Mouton & Co.
- AUFRECHT, T. 1877²: *Die Hymnen des Rigveda I-III*. Bonn: Adolph Marcus
- ERHART, A. 1980: *Struktura indoiránských jazyků*. Brno, Univerzita J. E. Purkyně (*The Structure of the Old Indoiranian Languages*)
- ERHART, A. 1990: *Základy jazykovědy*. Praha: SPN (*Basics of Linguistics*)
- HJELMSLEV, L. 1953: *Prolegomena to a Theory of Language*. Baltimore: Indiana University

- LASS, R. 1998^a: *Phonology. An Introduction To Basic Concepts*. Cambridge: Cambridge University Press
- MACDONELL, A. A. 1910: *Vedic Grammar*. Strassburg: K. J. Trübner
- MACDONELL, A. A. 1916: *A Vedic Grammar for Students*. Oxford: Clarendon Press
- ŠEFČÍK, O. 2000a: *Příznakovost konsonantů ve staroindických trsech a jejich reflexe ve středoindických jazycích*. SPFFBU, A 48, 19–30 (*Markedness of consonants in OIA clusters and their reflexes in Middle Indo-aryan languages*)
- ŠEFČÍK, O. 2000b: *Povaha kořene ve staroindičtině coby znaku*. Jazykovědné aktuality (zvláštní číslo) XXXVII, 81–86 (*OIA verbal-root as a sign of language*)
- TRUBETZKOY, N. S. 1939: *Grunzüge der Phonologie*. Prag: TCLP 7
- WACKERNAGEL, J. 1896: *Altindische Grammatik I. (Lautlehre)*. Göttingen: Vandenhoeck und Ruprecht
- WHITNEY, W. D. 1879: *Indische Grammatik, umfassend die klassische Sprache und die älteren Dialecte*. Leipzig: Breitkopf und Härtel
- WHITNEY, W. D. 1885: *Die Wurzeln, Verbalformen und primären Stämme der Sanskrit-Sprache*. Leipzig: Breitkopf und Härtel

Ondřej Šefčík
Ústav jazykovědy FF MU
Arna Nováka 1
CZ-60200 Brno
(sefcik@phil.muni.cz)