

**A Reply to Bryan Levman's  
*The Language the Buddha Spoke***

*Stefan Karpik*

Though it is welcome that Bryan Levman's paper (2019) in this issue, *The Language the Buddha Spoke*, seeks common ground with the SOTT (Single Oral Transmission Theory) proposed in my recent paper (Karpik 2019), I hope it does not seem churlish to reject his position almost completely. Examples of the SOTT in my view are the arguments of Gombrich (2018: 84-5) that Pali was the Buddha's idiolect, of Wynne (Gombrich 2018: 82-3) that it was his dialect and mine that Pali was a sociolect.<sup>1</sup> I had not anticipated Levman's innovation of a single transmission having an earlier stratum, a *lingua franca* influenced by the north-western dialect (p.71-2), coeval with the Buddha (p.66), but changing into a later mutually unintelligible stratum represented by Pali (p.74) some 200 years later (p.75). He wishes to call both strata *Pali*, but to avoid confusion, I will call Levman's alleged stratum *pre-Pali* and the attested language *Pali*.

Although I agree with Levman's claim of an underlying language to Pali, I do not think this has any significance; there is an underlying layer to any language. What I do deny is that the available evidence is able to date an underlying layer to the Buddha. The reconstructions suggested by Levman could, if accurate, belong to any time in the millennium contemporary with the Vedas and antedate the Buddha by centuries. Uncovering an earlier layer, even if accurate and even if somehow dateable to the Buddha, does not prove a *lingua franca*. The need

---

<sup>1</sup>Thus, Levman (p.64) misrepresents my view as being the same as Gombrich's.

for communication across language groups could also be met by bilingualism or by adopting a standard language, much as standard English is often used today as an international language. Epigraphical Prakrit provides direct evidence of such a standard language across India, while Levman admits there is no direct evidence of a lingua franca/ koine (p.73).

In common with MOTT (Multiple Oral Transmission Theory) advocates, Levman gives no account of why the underlying layer was discarded and lost, despite repeated injunctions in the suttas to memorise them to the letter (Karpik 2019: 14-15); difficulty in understanding archaic language did not deter Vedic reciters or Catholics worldwide using the Latin liturgy in the last century. Nor does Levman engage with my argument that the alleged composite character of Pali is a feature of natural languages (Karpik 2019: 67-69); he merely reiterates the MOTT position that it 'proves' Pali is an artificial language (p.76).

Moving on to more detailed discussion:

1. Levman (p.80-81, n.13) claims *mantā* is an example of *-ttā* absolutive. It does not look like a *-ttā* absolutive and is more likely to be an instrumental or nominative of *mantā* or *mantar*. The commentary, *mantā mantā ca vācaṃ bhāsati ti ettha mantā ti vuccati paññā, mantāya paññāya. puna mantā ti upaparikkhitvā* (D-a III 892 on D III 106), can be seen as referring to nouns, not verbs. I have yet to see an unambiguous example of the *-ttā* absolutive in Pali; Geiger §210A gives only tentative examples and Oberlies §119.1 admits they are sporadic and mostly unrecognisable. The obvious candidates, such as *kattā* and *chettā*, can also be seen as agent nouns and other supposed examples of the absolutive in *-ttā*, e.g. *sammasitā, āharitā, paccuggatā, parivajjayitā*. do not even look like this absolutive. Wynne (2013:151-156) argues that *puchitā, āpajjitā, apassayitā, upapajjitā, chinditā, nahāyitā, nisīditā, passitā* and *bhuñjitā* may not be agent nouns, but absolutives in *-ittā* which follow the early orthographic convention of not marking geminates. On the other hand, there are over 13,000 *-tvā* absolutives in the Tipiṭaka and, if the *-ttā* absolutive can one day be proven from the above handful of cases, it is surely more likely that these are a few accidental Prakritisations of pre-Aśokan *-tvā* instead of 13,000+ Sanskritisations of *-ttā*.

I agree with Levman (p.80-81, n.13) that the *-ttā* absolutes of the Patna Dhammapada are later than Pali, but that merely strengthens my case that *-tvā* is the original form. Levman does not engage with my point that the alleged Sanskritisation applied only to the alleged *-ttā* absolute and mysteriously avoided the other seven Pali absolute forms (Karpik 2019: 56 fn.71). Pace Levman, the *(-)tv-* conjunct is indeed a feature of Pali as there are 2,000+ examples of *tvam* in the Tipiṭaka, compared to under 300 of *tuvam*, and 400 instances of the sandhi *-tve-*, for which Levman offers no explanation. The 1,900+ *tvāna* forms in the Tipiṭaka not only challenge the Sanskritisation narrative in that they would actually represent a Vedicisation if the Tipiṭaka had been altered, but they are also copious evidence for *-tv-* in Pali. 'Sanskritisation' raises more questions than it appears to solve: why should only *(-)tv-* and *br-* be Sanskritised; do we also have to take *ty-*, *dr-*, *dv-*, *vy-* and *sv-* as 'Sanskritisations' to conform to the dictum of Oberlies §16.1 that only single consonants are allowed in word initial position; what then of medial positions for these clusters; why isn't there *dvitiya* instead of *dutiya*, *iha* instead of *idha*, *prati-* instead of *paṭi-*, etc.; if *tatra* is Sanskritised *tattha*, why was 'Sanskritisation' incomplete; alternatively, why is the 'Sanskritisation' of *-tvā* complete but not of *tuvam*; why does Pali show the same degree of 'Sanskritisation' throughout, unlike BHS? The *-tv-* cluster is indeed original and archaic, along with many other features of Pali (Karpik 2019: 53-8), and supports the claim that Pali is pre-Aśokan.

2. Levman (p.71-72) does not engage with my argument that *brāhmaṇa* is a loan word. He instead claims it is a re-Sanskritisation (implying 3.500+ corrections in the Tipiṭaka) and he reconstructs pre-Pali *\*bāhaṇa* although several Prakrit inscriptions have *br-* (Shāhbāzgarhī and Mānsehrā *bramaṇa*, Girnar *brahmaṇa*, Bharhut *bram(h)ana*). His argument is confused: he does not acknowledge the Girnar and Bharhut forms; the similarity of the 're-Sanskritisation' with the north-western forms of Shāhbāzgarhī and Mānsehrā is a coincidence;

in contradiction to the last point (p.72) “an underlying, earlier layer to Pāli was strongly influenced by the north-western or most prestigious dialect”. The different treatment of *r* in the north-western dialect from Pali is emblematic of why they have not influenced each other. Pali has *r* dropping (*dhamma* for *dharma*); the north-western dialect not only retains *r*, it sometimes undergoes metathesis (*dhrama* for *dharma*).

3. Levman (p.77 n.10) claims the directionality of *-bb-* > *-vv-* is moot because a pun in Sn 537 could be read as either form. This doesn't address my argument that by the time of Epigraphic Prakrit and the literary Prakrits the change *bb* > *vv* was complete and Pali is earlier. The pun in Sutta Nipāta shows the transition taking place before attested written forms.
4. Levman does not discuss orthographic errors anywhere in his paper. Had he done so, he may have avoided the following error in his discussion (p.78) of Dh 335: “There are four variants to *abhivaṭṭhaṃ* in the different Pali recensions: *abhivaṭṭaṃ* (PTS = Sinhalese), *abhivaḍḍhaṃ* or *abhivutṭhaṃ* (Thai), *abhivuḍḍhaṃ* (Cambodian).” (I would add that this word is also at Th 400 as *abhivaḍḍhaṃ* (PTS), *abhivaṭṭhaṃ* (Burmese), *abhivutṭhaṃ* (Thai), *abhivaḍḍhaṃ* (Cambodian).) Levman considers lenition of *-ṭṭh-* > *-ḍḍh-* to be one explanation of the divergences, but this is the error: geminates do not undergo voiceless to voiced lenition. Kirchner (2000: 510) found this in 272 languages and neither Geiger §38 nor Oberlies §15.2 offer any geminates for this lenition. More likely there have been orthographic errors: in the Pallava script<sup>2</sup> *t*, *ṭh*, *ḍ* and *ḍh* all look similar and likely to be readily confused; in Sinhalese script<sup>3</sup> *ṭh* to my eye could be confused with *t* or, on the other hand, with *ḍ* or *ḍh*.

---

<sup>2</sup> <https://www.omniglot.com/writing/pallava.htm>

<sup>3</sup> [http://buddhism.lib.ntu.edu.tw/BDLM/en/lesson/pali/lesson\\_pali1.htm](http://buddhism.lib.ntu.edu.tw/BDLM/en/lesson/pali/lesson_pali1.htm)

5. Levman frequently gets directionality wrong:
- a. In the PTS editions D I 223 has *sabbato-pahaṃ* whereas M I 239, Vv114 and J VI 46 have *sabbato-pabhaṃ* 'shining everywhere'. (Be has only *sabbato-pabhaṃ* in all four cases.) Levman (p.66) considers *paha* to be pre-Pali and is thus alleging a rare fortition. The literature Levman references does not refer to Vv114 or J VI 46, but on the principle of 'majority wins' (Campbell 2004:131), I consider *pabha* the original form and *paha* a later accidental lenition during dictation to a scribe, based on the analogy of Pali *nabha(s)* > Māhārāṣṭrī *naha* 'sky' (Bubenik 1966:56; see also Pischel §188). This directionality is confirmed by Oberlies §15.15 (b) who gives OIA *prabhū-* > Pali *pahu-* 'able' and OIA *prabhūta* > *pahūta* > Pali *bahūta* 'much'. (There is also an alternative explanation of *pahaṃ* as a copyist's error based on a confusion between *bh* and *h*, which are difficult to distinguish in Sinhalese characters.)
  - b. Levman (p.69-70) gives an example from Lévi (1912: 502-3) of Pali *opapātika* and BHS *aupapādika* 'spontaneously reborn'. They claim that both forms are derived from Prakrit \**uvavāya* (AMg *uvavāiya*). Geiger (1916: 6) found Lévi's derivations unconvincing and so do I; Geiger §38 attributes the voicing of unvoiced intervocalic consonants, a common lenition, to dialect influence and so do I. My scepticism is based first on the fact that both *utpatti* and *utpāda* are nouns found in Sanskrit with the meaning 'birth' and neither Lévi nor Levman refer to either. They may well be correct in alleging a confusion over whether the root was *pat* or *pad* in the formation of adjectives, but *utpatti* is found in the *Suśruta-saṃhitā*, which may be a late Vedic text, and the Pali *opapātika* appears to have an early provenance based on this form. I doubt that anyone can date the AMg reflex with any certainty as the Jain scriptures were not agreed till the Council of Patiliputra in the 4<sup>th</sup> century BCE and not written down till the council of Valabhi in the 5<sup>th</sup> century CE with much scope for inadvertent sound change. My second ground of

scepticism is that unusual fortitions,  $-y- > -t/d-$  from pre-Pali to Pali are implicitly being claimed without giving analogies as evidence. I doubt that such analogies exist.

- c. Levman (p.82-83) examines *sobhissāmi* at D I 105 and its variant readings, *sodhissāmi*, *sodissāmi*, *sodhāssāmi* and *sovassāmi*. He provides evidence for the loss of aspiration *sodhissāmi*  $>$  *sodissāmi* and for the lenition *sodhissāmi*  $>$  *sovissāmi*. I can add an orthographic element to the discussion: in the Pallava script *bh* and *dh* are easily confused and that *dh* and *v* are virtually indistinguishable, so I am not convinced these lenitions did actually occur, even though they are indeed plausible. So far Levman and I are not so far apart, but then he wishes to claim pre-Pali *\*sohissāmi* as the common ancestor of *sodhissāmi* and *sobhissāmi* without giving any evidence. The change  $h > dh/bh$  reverses the directionality and conflicts with Levman's own evidence, Pischel §213 and Brough §43,49; if correct, it would represent not one but two rare fortitions. Of course, some changes move in both directions, but there is no evidence that this change does so. Levman's unattested fortitions are both implausible and pointless, adding nothing to the understanding of the attested forms.
- d. Levman (p.83-84) claims the underlying form of *vegħa*, *vedħa*, *vekħa*, *vekkħa* and *veṭħa* is pre-Pali *veħa*. In the Pallava script, *gha* and *ha* are readily confused, so this appears to support Levman's claim of  $ha > għa$ . However, his claim of, not one, but five separate unattested fortitions is not credible. I doubt there exists any attested example of five fortitions in any language generated by *h* or any other sound. But according to Levman, there are eight separate rare fortitions of *h*: we have already seen (p.82-83) the claim of  $h > dh/ bh$ , here we have  $h > gh/ dh/ kh/ kkh$  and  $ṭh$ , and, for *ganṭhita* (p.85-86), we also have  $h > th$ , thus producing eight separate rare fortitions of *h*! <sup>4</sup> Levman does not discuss

---

<sup>4</sup> Fortitions are rare in Indo-Aryan. Pischel §267 gives examples of fortitions of nasal

this utterly implausible situation. If he argues that these are hypercorrections and not natural language changes, then without evidence or discussion Levman is free to conjure any pre-Pali form he fancies and call it a hypercorrection in Pali. I don't see how this is helpful to scholarship and it would undermine Levman's claim of a degree of natural evolution of Pali (p.76) as most of his reconstructions follow the pattern of a lenition from OIA to pre-Pali followed implicitly by a fortition from pre-Pali to Pali.

6. Levman claims an astonishingly fast pace of change from pre-Pali into mutually unintelligible Pali in a mere 200 years from 380 BCE to 180 BCE. Yet in the following 400 years inscriptions from Bharhut (2<sup>nd</sup> century BCE) to Nasik in the time of Vāsiṣṭhīputra Puṣumāvi (2<sup>nd</sup> century CE) hardly show any change in Epigraphic Prakrit, thus making Levman's thesis improbable. To make a convincing claim of mutual unintelligibility between pre-Pali and Pali, he would need to show different syntax and different lexis. He shows neither; *\*n(ṅ)iv(v)aṅa* for *nibbāna* and *\*bāhana* for *brāhmaṇa* are not, as claimed (p.65), lexical differences but phonological, i.e. differences of accent. Significantly, he does *not produce any* syntactical differences. He claims instead that unintelligibility is entirely subjective and states that he can read Pali, but not Gāndhārī and Ardhamāgadhī. Despite his obvious diligence, he cannot hope to replicate the experience of a native speaker of Indo-Aryan and I have never claimed mutual intelligibility without effort for native speakers. There are in fact objective tests: we can objectively infer that Shakespeare is intelligible to modern native English speakers because films of his plays are not shown with modern English subtitles - in Britain at least. Similarly, we can infer that the varieties of Indo-

---

+ consonant clusters becoming aspirated with the caveat that some may be older forms than classical Sanskrit. Pischel §190,191 gives,  $g > k$ ,  $gh > kh$ ,  $j > c$ ,  $jh > ch$ ,  $d > t$ ,  $dh > th$ ,  $d > t$ ,  $dh > th$ ,  $b > p$ ,  $bh > ph$  in Paisācī; Konow (1910) explained these features as either archaic or as the influence of (bilingual) Dravidian speakers. Geiger §39.1 lists sporadic fortitions explaining them as dialectical variation and Oberlies §15.4 lists many of the same under the tendentious heading hyper-translations; I claim they demonstrate hyper-corrections and we should not assume that Sanskrit pre-dates Pali (Karpik 2019: 72).

Aryan were mutually comprehensible in the Buddha's day from Vinaya rules that treat *Ariyaka* as a single language and from arguments about the correct word for *bowl* in different localities and from the similarities of the Aśokan inscriptions in syntax, lexis and morphology (Karpik 2019: 15-17, 58-63).

7. In what I consider to be intellectual legerdemain, change and translation are held to be equivalent (p.67). This entitles Levman to claim that his position has much in common with Geiger (p.68) and von Hinüber (p.66). However, the MOTT in my paper was not restricted to Norman's views, as Levman interprets, but includes those of Geiger and von Hinüber, as they regard the single original language transmission as being intentionally different from the later artificial, literary language, Pali, thus implying at least two oral transmissions. This also entitles Levman to misrepresent me (p.85, n.18) for not understanding that native speaker hypercorrection *\*isivayana* is a change, whereas I argue it does not prove translation. (I imagine part of the confusion is that Levman believes this well researched reconstruction by other scholars is, like his own reconstructions, datable to the Buddha, whereas I believe *\*isivayana* may predate the Buddha by centuries.)<sup>5</sup> Levman further misrepresents Geiger (p.68) as claiming a *koine* underlies Pali; actually Geiger (1916: 4-5) refers to a *lingua franca* which could be described as *Māgadhī* spoken by the educated, and presumably not a simplified dialect.<sup>6</sup> He again (p.73) misrepresents Geiger as saying the underlying layer was a *Kunstsprache*; actually this is how Geiger described the alleged later literary language of the Buddha's disciples.

---

<sup>5</sup> Levman (p.69-70, n.4) also misrepresents me by saying I do not understand the term 'Middle Indic'. In fact, I don't use the term at all and refer to *Ariyaka*, which I translate as 'Indo-Aryan'. If Levman has a better translation, he should say so.

<sup>6</sup> Levman's use of *koine* is incorrect in modern linguistics, in which *koine* refers to a de-regionalised variety. The original *Koine* lost its Attic features and became a de-regionalised form of Greek. Cf. Kerswill & Williams (2005:1023): "The establishment of new towns in the twentieth century in many parts of the world is a test bed of koineization, the type of language change that takes place when speakers of different, but mutually intelligible language varieties come together, and which may lead to new dialect or koine formation." The original *Lingua franca* was a pidgin, but this is not what Geiger meant and the term appears not to have a precise definition.



Levman claims that he is following procedures established by eminent Indologists and there is some truth to this. In my view, the hunt for the underlying language of Pali has been a disaster for its scholarship. It has used up academic time and energy on a conspiracy theory of hundreds of fortitions/ hyper-corrections and thousands of Sanskritisations, on a fruitless ghost-hunt for the *-ttā* absolute and on ghastly castles in the air consisting of unfounded speculation, where opinion is treated as evidence. For example, Levman (p81-82) follows Lévi's false deduction of *avayesi* > *avadesi*, which flies in the face of the general trend of MIA towards lenition; from the evidence of *avayesi* at Bharhut, a reasonable person would assume that Pali *avadesi* was earlier; instead Lévi claimed that Pali was later and alleged a Sanskritisation to back his claim that Pali is a kind of BHS.

This is a circular argument known as 'begging the question' or *petitio principii*, where one assumes what one wishes to prove – Sanskritisation and therefore a reversal of directionality – in order to prove it. Levman expands this unfortunate inheritance by proposing on the basis of improbable reconstructions, such as the stand-alone aspirate developing eight separate consonant aspirations, a *lingua franca* spoken in the Buddha's day consisting of, *inter alia*, lenition and assimilation of OIA forms which were reversed 200 years later by fortition, hypercorrection and Sanskritisation. This is the antithesis of Occam's Razor. Levman does not refer to the reversals as such, so I wonder if he has noticed the implications of his theory. On the other hand, I propose no such reversals, but a steady development in a consistent direction. This simpler view conforms to the overall directionality of Indo-Aryan linguistics and is more elegant.

I am not unsympathetic to Levman's concept of the development of a simplified *lingua franca* through contact with Indo-Aryan, Dravidian, Munda and Tibetan speakers. However, I regard it as helpful to a narrative of the early development of Prakrit in general, rather than Pali in particular, because reconstructions cannot be dated with any precision. The Koine developed in the eastern Mediterranean following Alexander's conquests and there is an obvious parallel with the Aryan invasion of India. However, the concept of a linguistic area in India, which necessarily involves bilingualism, is widely accepted and the effects of bilingual intrusion including retroflex consonants and absolutes (Emeneau (1980: 89ff) can be seen in the earliest Vedas. Trudgill (2010: 1-35) also explores how bilingual contact in the Dark Ages in Britain between Celtic, Latin, Old Nordic and Old English speakers led to lexical intrusion and the simplification of the inflectional system of Middle English. This also parallels the Aryan colonisation of India and the simplification of (pre-)Vedic into Prakrit.

The narratives of *lingua franca* and bilingualism are not necessarily in competition, before the formation of the Vedas, in my view, but bilingualism would predominate in the Buddha's day after centuries of contact. I would like to see Levman's penchant for historical linguistics extend his simplification narrative to Prakrit in general, thus expanding on his work on contact with indigenous peoples (Levman 2013). I will read future work by him on his concept with much interest.

## References

- Brough, John 1962 *The Gāndhārī Dharmapada*. Delhi: Motilal Banarsidass.
- Bubenik, Vit 1996 *The Structure and Development of Middle Indo-Aryan Dialects*. Delhi: Motilal Banarsidass.
- Campbell, Lyle 2004 *Historical Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Emeneau, Murray B. 1980 *Language and Linguistic Area*. Edited by Dil, Anwar S. Stanford: Stanford University Press.
- Geiger, Wilhelm 1916, (1943) *Pāli Literature & Language*. Translated by Batakrisna Ghosh. Delhi: Oriental Books Reprint Corporation. [First published as *Pāli Literatur und Sprache*, Strassburg: Karl J. Trübner]
- Gombrich, Richard 2018 *Buddhism and Pali*. Oxford: Mud Pie Books
- Karpik, Stefan 2019 "The Buddha Taught in Pali: A working hypothesis." *Journal of the Oxford Centre for Buddhist Studies*. 2019 (16): 10-86
- Kerswill, Paul & Williams, Ann 2005 "New towns and koineization: linguistic and social correlates." *Linguistics* 43-5 (2005), 1023-1048
- Kirchner, Robert 2000 "Geminate Inalterability and Lenition." *Language*, Vol. 76, No. 3 (Sep., 2000) 509-545.
- Konow, Sten 1910 "The Home of Paiśācī." *Zeitschrift der Deutschen Morgenländischen Gesellschaft*. ZDMG Bd. 64 (1910) 95-118.
- Lévi, S. 1912 "Observations sur une langue précanonique du bouddhisme." *Journal Asiatique* Novembre-Décembre 1912: 495-514.
- Levman, Bryan 2013 "Cultural Remnants of the Indigenous Peoples in the Buddhist Scriptures." *Buddhist Studies Review* BSRV 30.2 (2013) 145-180.
- 2019 "The Language the Buddha Spoke." *Journal of the Oxford Centre for Buddhist Studies*. 2019 (17): 63-105.

- Oberlies, Thomas 2019 *Pāli Grammar: The Language of the Canonical Texts of Theravāda Buddhism. Vol I Phonology and Morphology*. Bristol: Pali Text Society.
- Trudgill, Peter 2010 *Investigations in Sociohistorical Linguistics: Stories of Colonisation and Contact*. Cambridge: Cambridge University Press.
- Wynne, Alexander 2013 A Preliminary Report on the Critical Edition of the Pāli Canon being prepared at Wat Phra Dhammakāya. *Thai International Journal of Buddhist Studies* TIJBS IV, 2013, 135-170.