

# The Buddha Spoke Pāli

by Stefan Karpik

## The Problem

Pāli means “text” and the Theravādin commentarial tradition tells us that the actual language of the texts is Māgadhī, which is what the Buddha spoke.<sup>1</sup> However, this is dismissed by most Western scholars and Theravādin Buddhist writers have deferred to their opinion. Even among Thai monastics, the opinion that the Buddha spoke Māgadhī, and not Pāli, is common. I have been impressed by the confidence of the claims of Western scholars and rather surprised. For implied in these claims are (a) that the Buddha’s words were translated into Pāli and (b) that, once translated, the original words were lost. This process is fairly typical of a written tradition, such as the Christian New Testament and Classical authors, but it would be, so far as I know, unique, if it were true, in an oral tradition. This is therefore a large claim to suggest that without the benefit of writing materials, a translation equivalent to thousands of pages of scripture was made. It is the more remarkable because, in the same culture and geographical region, other canons, such as the Vedas and the Jain texts, have been transmitted orally over millennia.

Prominent in claims that the Buddha spoke something other than Pāli are K.R. Norman<sup>2</sup> and Richard Gombrich<sup>3</sup>. Both are former Presidents of the Pāli Text Society (PTS), but I trust that neither would wish me to accept their opinions without examining some of the evidence. Their argument runs as follows:

- a) Pāli contains features which appear to come from several dialects and includes incorrect backformations from Sanskrit. It therefore bears the hallmarks of an invented or literary language translated from other authentic dialects or languages.
- b) Furthermore, the Buddha specifically allowed translation of his words.
- c) Māgadhī, the language of Māgadhā (now Bihar) was clearly different from Pali.

However, now I have examined the evidence, I am far from convinced.

## An Out of Date View

Before arguing against Norman and Gombrich, I will explain what my position is. Essentially, I follow Geiger<sup>4</sup>, whose introduction to the reissue of his grammar for the PTS was replaced by Norman in favour of a new one by Gombrich:

- Pali forms are derived from Vedic, the language of the Vedas, not from Sanskrit, the language of the Upanishads, and stand beside Sanskrit forms as later formations from Vedic. Numerous double forms indicate that Pali is a mixed dialect, but it had its origin in a particular dialect.<sup>5</sup> The Theravāda tradition states that Pāli is Māgadhī; however, distinguishing features of Māgadhī, such as no r or

---

<sup>1</sup> Vibhanga Commentary 388 *sammāsambuddho pi teṭṭakam buddhavacanam tantim āropento Māgadhībhāsāya eva āropesi.*

<sup>2</sup> K.R. Norman *Pali Literature* Otto Harrassowitz: Weisbaden 1983

K.R. Norman *The Value of the Pali Tradition* Collected papers Vol 3 PTS Oxford 1992

K.R. Norman *A Philological Approach to Buddhism* SOAS 1997

<sup>3</sup> R. F. Gombrich *What is Pāli? A Pāli Grammar* W. Geiger, tr.B. Ghosh, Ed K.R. Norman PTS 1994

<sup>4</sup> Wilhelm Geiger. *Pāli Literature & Language*. Oriental Books Reprint Corporation, New Delhi 1978

<sup>5</sup> Much, I suppose, as BBC English has its origins in the South East of England, but slips into Americanisms, such as “prehaps, nucelar, did you do that yet, like he said”.

s and ending nominative masculine and neuter -a stems in -e are absent in Pāli. Even so, Geiger regards Pāli as a form of Māgadhī spoken by the Buddha:

- Ārsa, the language of the Jain suttas, is called Ardha Māgadhī, “half-Māgadhī”, but the special features of Māgadhī are also absent. Therefore, Pāli may be considered as a kind of Ardha Māgadhī.
- Pāli is a language of the higher and cultured classes which had been brought into being in pre-Buddhist times through the needs of intercommunication in India (following Rhys Davids, *Buddhist India* p140ff)
- Pāli is a *lingua franca* free of the most obtrusive dialectical characteristics<sup>6</sup>, but flavoured by the dialect of the speaker. The Buddha, though not a Māgadhan, was active in Māgadha and this could have influenced his language.<sup>7</sup>
- The Pāli canon was not translated. Its features can be explained by a gradual integration of elements from various parts of India, a long oral tradition and the texts being written down in a different country.
- Cullavagga V.33.1 of the Vināya (*anujānāmi bhikkhave sakāya niruttiyā buddhavacanam pariyāpuṇitum*) is mistranslated as “I allow you, oh brethren, to learn the words of the Buddha, each in his own dialect” by Rhys Davids and Oldenberg. Two monks were complaining that monks were spoiling the Buddha’s words with their own dialects and suggested putting them into Sanskrit verse. Geiger follows Buddhaghosa, the most prominent commentator, in concluding the Buddha refused any translation both negatively by refusing verse and positively by requiring the Buddha’s words to be learnt in the Buddha’s language. The correct translation should be “I ordain the words of the Buddha to be learnt in his own language (in Māgadhī, the language used by the Buddha himself)”. This is more in keeping with Indian tradition and the context in which the monks were complaining that the Buddha’s words were being corrupted.

### Against the Modern View

I will now attempt to argue against the Norman and Gombrich view. Please bear in mind that I am in no position to contradict their learning, but I am questioning their deductions:

#### a) Pāli is not an invented language.

A mixture of dialects is a feature of most living languages. Regularity is the exception rather than the rule and belongs to artificial languages such as Esperanto and classical Sanskrit. The fact that a language shows mixed features does not imply that it has been translated. It would be absurd for future 45<sup>th</sup> century professors of Middle Americanoid languages to suppose that a transcript from a current BBC programme had been translated from an earlier language because it shows ancient features e.g. “oxen” instead of “oxes”, “mice” instead of “mices” or includes Americanisms or has regional variations like “chimdey” for “chimney”.

Actually, the mixing of dialects was not great. Even Norman (1997:44-45) states: “We can see that in the canon as a whole there are very few non-Pāli characteristics and most

<sup>6</sup> Similar, I imagine, to received pronunciation/ Queen’s English/ BBC English. M iii 235 states that a monk should not insist on local language and should not override normal usage: *janapadaniruttim nābhinvaseseyya samaññaṃ nātidhāveyyāti*.

<sup>7</sup> Much, I suppose as a northerner might change “bath” to “bāth” when moving to the South of England.

of these are due to a consistent introduction, at a later date, of Sanskritisms, which are restricted in number, the most obvious being the absolutive in *tvā*.”

The evidence of artificiality from supposed Sanskritisation can be looked at differently: Geiger believed that Pali developed from Vedic in parallel with Sanskrit. Thus Pāli could have developed its own forms without any backformation. Backformations are in any case a normal feature of languages of living languages; notice the debate on referenda or referendums, phenomena or phenomenon; I once met a professor of a Canadian university who argued that the plural of pizza was *pizae*.

**b) The Buddha specifically disallowed translation of his words.**

As noted above, Geiger and Buddhaghosa disagree with this interpretation of the Pāli<sup>8</sup>. I would like to add that I don't see how translation could be practicable in an oral tradition equivalent to over 5,000 pages of Vinaya and Suttas when written down in the 1<sup>st</sup> century B.C.E.<sup>9</sup> The labour involved in producing a standard consistent translation capable of recitation among a group of monks would have been enormous and without parallel in world history so far as I know. Not only that, but those disagreeing with Buddhaghosa are implying this heroic effort was repeated several times over in different regions of India.

In fact, the word “translation” is misleading, though arguably academically correct. Normally we think of translation from one different language to another, but the boundaries between language and dialect are not always clear. A speaker of Norwegian can chat with a speaker of Swedish and understand the other, though speaking different a language. A speaker of Mandarin Chinese and a speaker of Cantonese Chinese cannot do this, though speaking different dialects. Some linguists therefore claim that the difference between a language and dialect is political.<sup>10</sup> With regard to the different states of Northern India in the Buddha's time, it would therefore be academically correct, though misleading to laypeople, to talk of several languages being spoken. In common parlance, however, the languages of North India could be said to represent different dialects of the same language. Ashokan inscriptions, dating 150 years after the Buddha's death in c. 400 B.C.E., are available in up to six dialects, which look very alike and whose differences would probably not cause much difficulty to native speakers. My point is that there was no need for “translation”, which is why the Buddha should not be interpreted as allowing it.

Gombrich (1994: xxvii) implies that I am on the right lines: “Before the texts were ever written down, it is not likely that their dialect was ever completely fixed, or even that the differences between the dialects were clearly conceptualised; it must have been a matter of reciting in what appeared like “regional accents.”

If the Buddha did not allow translation and translation was unnecessary, what was the nature of the corruption of the teachings that was complained of? Perhaps the monks were using such different accents that they could not chant together harmoniously. The

<sup>8</sup> Cullavagga V.33.1 *anujānāmi bhikkhave sakāya niruttīyā buddhavacanam pariyāpunitum*.

<sup>9</sup> This is when palm leaves and ink became available to replace rock and clay tablets.

<sup>10</sup> This argument is aired in *The Story of Language* by Mario Pei, Allen & Unwin 1952.

PTS dictionary gives for *nirutti*, the word for what was spoiling the Buddha's speech, "pronunciation, dialect, way of speaking, expression". Norman (1980) thinks it was the glosses of the Buddha being altered and the passage requires monks to use the Buddha's own glosses. I could similarly guess that "term" is a suitable translation for what was altered, i.e. some monks were imprecisely saying, for example, *citta* for *nāma*, *kāya* for *rūpa*; however, I admit I don't know, for the Pāli is too open to interpretation.

The reader may wonder how Norman can stretch the dictionary meaning and how I can presume to guess it; the reason is that academic work in this area is often speculative. Here is Norman (1997:59): "We very commonly find in books and articles about early Buddhism such statements as: "The Buddha preached in the Prakrits, the language of the common people, and resisted the suggestions of some of his ex-Brahman followers to translate his sermons into Sanskrit". There is frequently no hint that these statements are anything other than accepted fact, but readers need to be very wary, because such statements are frequently not fact, and are anything but accepted by all scholars working in the field. To tell you the truth, there is a great deal of the Bellman principle in the academic world. You all know about the Bellman in Lewis Carroll's *The Hunting of the Snark*, who maintained that: "what I tell you three times is true". I am as guilty in this respect as anyone else, I fear. I may have an idea about something, and so I incorporate my idea, as a suggestion, in an article I am writing, and wait for someone to reject or disprove it. No one does, and I repeat the idea, still as a suggestion, in another article. again, no one rejects it. I do this a third time, and if there is still no reaction, it becomes fact in my mind – I have said it three times, so it must be true, and I consequently refer to it in future publications as an established fact. The thought that no one ever reads my articles and so no one has ever seen my suggestion, and so no one had had any desire to reject it, or the alternative explanation, that those who read my first article thought that the idea was so preposterous that it was not worth wasting paper and ink refuting it, so that the second and third repetitions were dismissed as: "I see that Norman is still pushing that stupid idea of his", does not enter my head."

### **c: Pāli and Māgadhī are not different**

There is no clear reason to doubt the Theravadin tradition:

1. Norman and Gombrich object to the notion that the Buddha spoke Pāli, which is not the language of the area called Māgadhā. The earliest examples of anything that could be called Māgadhī are Ashokan inscriptions. Gombrich<sup>11</sup> gives the gap between the Buddha's death and the ascension of Ashoka to the throne as 136 years; the inscriptions are later. There was therefore time for the dialects of Māgadhā to change from something similar to Pāli to something similar to the Māgadhī of the Ashokan inscriptions.
2. Alternatively Pāli was a Western dialect, used across other parts of India, which went out of fashion in Eastern India by the time of Ashoka. Perhaps the prestigious dialect of Māgadhā's court became a lingua franca and the basis of the eastern dialect of the Ashokan inscriptions.

---

<sup>11</sup> R.F. Gombrich "Dating the Buddha: A Red herring Revealed" *The Dating of the Historical Buddha Part 2* ed H. Bechert, Gottingen 1992

3. In an early work Norman<sup>12</sup> comes to some conclusions which are compatible with Geiger's: the occasional use of the -e form in the nominative case in Pāli suggests that it was spoken in Eastern Māgadha; Māgadhi was considered the root language of all languages and in the usage of the time of introducing Buddhism to Ceylon, it was correct to call Pāli Māgadhi; the Buddha spoke Māgadhi and Pāli was spoken in Māgadha; it is broadly speaking correct to call the language of the Theravada canon 'Māgadhi'.

### Some Evidence

My belief is that the differences between the North Indian dialects 150 years after the Buddha's death would not be significant for native speakers. Below I give the six Ashokan inscriptions<sup>13</sup> of Rock Edict I, which exemplifying local dialects of c.250 B.C.E. The first three are considered Western dialects and the last three are Eastern; Gīrnār is considered by some to be most like Pāli. Unfortunately, there are no parallels of Gīrnār inscriptions with any from Māgadha, but Māgadhan inscriptions resemble those of Dhaulī and Jaugaḍa. I include translations into English, Pāli and Sanskrit:

Location: <b>Śāhbāzgāhī</b> (Pakistan, NW Frontier)	Location: <b>Mansehrā</b> (Pakistan, near Rawalpindi)	Location: <b>Gīrnār</b> (India: Gujarāt State, Gīrnār Hills)
Ayaṃ dhramadipi devana priasa .... raṅo likhapitu hida no kici jive arabhitu prayuhotave	Ayi dhramadipi devana priyena Priyadraśina rajina likhapita hida no kici jive arabhitu prayuhotaviye	Iyaṃ dhammalipī devānaṃ priyena Priyadasinā rāṅā lekhāpitā idha na kimci jīvaṃ ārabhitpā prajūhitavyaṃ
Location: <b>Kālsi</b> (India:Uttar Pradesh, near Mussourie)	Location: <b>Dhauī</b> (India: Orissa, near Cuttack)	Location: <b>Jaugaḍa</b> (India: Orissa, near Lake Chilka)
Iyaṃ dhammalipī devānaṃ piyenā Piyadasinā lekhītā hidā nā kichi jive ālabhitu pajohitaviye	Iyaṃ ...devānaṃ piyena ... lājīnā likhāpitā ... .... kichi jīvaṃ ālabhitu pajohitaviye	Iyaṃ dhammalipī devānaṃ piyena Piyadasinā lājīnā likhāpitā hida no kichi jīvaṃ ālabhitu pajohitaviye
<b>Translation</b> by writer	<b>Pāli</b> composed by writer	<b>Sanskrit</b> by James Whelan <sup>14</sup>
This law edict was ordered to be inscribed by King Piyadasi <sup>15</sup> , Beloved of Gods. Here no living being is to be slaughtered and sacrificed.	Ayaṃ dhammalipī devānaṃ piyena Piyadasinā raṅṅā likhāpitā idha na kiñci jīvaṃ ārabhitvā pāhuneyyaṃ <sup>16</sup>	iyaṃ dharmalipir devānām priyeṇa Priyadarśinā rājṅā likhāpitā iha na kiñcij jīvaṃ ārabhya prahotavyaṃ

<sup>12</sup> K.R. Norman *The dialects in which the Buddha preached*. In "The Language of the earliest Buddhist Tradition" Ed Bechert Gottingen 1980

<sup>13</sup> A.C. Woolner *Asoka: Text & Glossary* Low Price Publications Delhi 1993

<sup>14</sup> My thanks to James Whelan of the Totnes Pāli study group for the translation into Sanskrit.

<sup>15</sup> "Piyadasi" is the name Ashoka gives himself.

<sup>16</sup> The Pāli word *pāhuneyyaṃ* has exactly the same root as the other examples. However, Buddhism changed its meaning from "should be sacrificed" to "should be offered gifts". In an early example of "spin-doctoring", the meaning of "sacrifice" was changed to "gift to monk". Similarly the meaning of "Brahmin" was changed from "one of good birth" to "one of good conduct".

## Conclusion:

There is no clear evidence to suggest that the language in which the Buddha taught was different from Pāli. Pāli shows the features of a normal, non-literary dialect and has a resemblance to real dialects of Northern India 150 years later. Pāli does not represent a translation, as that would be neither practicable nor necessary. Pāli is acknowledged to contain features of the Eastern Ashokan dialect. Therefore it could well have been spoken in Māgadha 150 years before the Ashokan inscriptions and there is no reason to distrust the tradition that it was indeed spoken there.

This is not to claim that the Pāli canon is a pristine representation of the Buddha's words. There is certainly fabrication<sup>17</sup> and possibly omissions. One would expect any real language to change subtly despite the checks imposed by communal recitation; for example, *bhikkhavo* (monks!) is less common than the *bhikkhave*, which is more in keeping with the Ashokan Eastern dialect. Possibly, there was a shift towards the Ashokan Eastern dialect which was not much noticed by monks who regarded the forms as equivalent.

With these caveats, I have come to the opinion that, if today the Buddha heard Sri Lankan monks (who have the best pronunciation) reciting Pāli, he would recognise at least some of his own words in a slightly different accent. That the Buddha's actual words can resound to us after 2,400 years is perhaps a romantic notion, but that does not prevent it being the truth.

Stefan Karpik  
Totnes, England  
stefankarpik@hotmail.com  
December 2004

Suggestions to improve this article are welcome.

---

<sup>17</sup> There is an anachronism in Miii 253 where Mahāpajāpatī Gotamī, the first Buddhist nun, is still a laywoman and is nonetheless referring to existing Buddhist nuns.