

THE BUDDHIST FORUM, VOLUME V

**A PHILOLOGICAL APPROACH
TO BUDDHISM**

The Bukkyō Dendō Kyōkai Lectures 1994

K.R. Norman

THE INSTITUTE OF BUDDHIST STUDIES, TRING, UK
THE INSTITUTE OF BUDDHIST STUDIES, BERKELEY, USA
2012

First published by the School of Oriental and African Studies (University of London), 1997

© Online copyright 2012 belongs to:
The Institute of Buddhist Studies, Tring, UK &
The Institute of Buddhist Studies, Berkeley, USA

British Library Cataloguing in Publication Data

ISBN 0-7286-0276-8
ISSN 0959-0595

CONTENTS

The online pagination 2012 corresponds to the hard copy pagination 1997

Foreword.....	vii
Abbreviations.....	ix
Bibliography.....	xi
I Buddhism and Philology.....	1
II Buddhism and its Origins.....	21
III Buddhism and Oral Tradition.....	41
IV Buddhism and Regional Dialects.....	59
V Buddhism and Writing.....	77
VI Buddhism and Sanskritisation.....	95
VII Buddhism and Aśoka.....	113
VIII Buddhism and Canonicity.....	131
IX Buddhism and the Commentarial Tradition.....	149
X Buddhism and Philology.....	167
Word Index.....	185

ABBREVIATIONS

AMg	Ardha-Māgadhī
AO	Acta Orientalia
B.C.E.	Before Common Era
BHS	Buddhist Hybrid Sanskrit
BHSD	Buddhist Hybrid Sanskrit Dictionary
BHSG	Buddhist Hybrid Sanskrit Grammar
BSO(A)S	Bulletin of the School of Oriental (and African) Studies
BSR	Buddhist Studies Review
BUp	Brhad-Āraṇyaka Upaniṣad
C.E.	Common Era
CP I–V	K.R. Norman. Collected Papers, Volumes I–IV (1990–94)
CPD	A Critical Pāli Dictionary, Copenhagen
CUP	Cambridge University Press
DEDR	T. Burrow & M.B. Emeneau, <i>Dravidian Etymological Dictionary</i> (Revised edition, 1984)
DPPN	G.P. Malalasekera, <i>Dictionary of Pāli Proper Names</i> (1937–38)
EV I, II	K.R. Norman, <i>Elders' Verses</i> I (1969), II (1971)
GD I, II	K.R. Norman, <i>The Group of Discourses</i> I (1984), II (1992)
IJ	Indo-Iranian Journal
IT	Indologica Taurinensia
JA	Journal Asiatique
JBuRS	Journal of the Burma Research Society
JOI(B)	Journal of the Oriental Institute (Baroda)
JPTS	Journal of the Pali Texts Society
JRAS	Journal of the Royal Asiatic Society
m.c.	metri causa
MIA	Middle Indo-Aryan
MRI	<i>Minor Readings and Illustrator</i>

MW	M. Monier-Williams, <i>Sanskrit-English Dictionary</i> (1899)
OUP	Oxford University Press
PDhp	Patna Dharmapada
PE	Pillar Edict
PED	The PTS's Pali-English Dictionary
Pkt	Prakrit
PTS	Pali Text Society
RE	Rock Edict
SBE	Sacred Books of the East
Skt	Sanskrit
SOAS	School of Oriental and African Studies (London)
SWTF	Goerg von Simson (ed.), <i>Sanskrit-Wörterbuch der buddhistischen Texte aus den Turfan-Funden</i> (1973–)
Überbl	O. von Hinüber, <i>Das ältere Mittelindisch im Überblick</i> (1986)
WZKSO	Wiener Zeitschrift für die Kunde Südasiens

Abbreviations of the titles of Pāli texts are those adopted by the CPD.
Editions are those of the PTS, unless otherwise stated.

IV

Buddhism and Regional Dialects

In the third lecture I dealt with the way in which the early Buddhist texts were transmitted by an oral tradition, and I mentioned that the *bhāṇaka* tradition could have come into effect within a very short time of the Buddha's death, because this did not seem to leave time for the homogenisation of language, which would seem to have been necessary if the Buddha preached, as is generally accepted, in a number of dialects. In this lecture I want to talk about the philological information which we have about the form of the language which was used in the very early stages of Buddhism, and the way in which we can interpret philological material to give us some idea of the way in which Buddhism began to spread from the boundaries of its origin.

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 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 has 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 Norman is still pushing that stupid idea of his”, does not enter my head.

This reference to the Buddha refusing to allow his sermons to be translated is based upon a passage in the Vinaya,¹ but to translate the passage in that way is to ignore the very real difficulties which are presented by two words in the original Pāli, *chandaso* and *nirutti*. The first has been translated “into Vedic”, or “into Sanskrit”, or “metrically”, or “as desired”. I cannot believe that “into Vedic” is a possible translation, but if “into Sanskrit” is a possibility, then this means, of course, that the Buddha’s words were not already in that language. In support of this translation, it has been pointed out that the bhikkhus who made the suggestion were brahmins by birth, who might be thought to favour Sanskrit in preference to the local vernaculars.

On the other hand, we might have thought that such learned people would have the sense to realise that the Buddha preached in Prakrit because that was what the local people spoke and understood, and it would be an act of folly to turn his sermons into something which would be unintelligible to them. There is the additional point that, if *chandaso* does mean “into Sanskrit” this would appear to have been unknown to later translators who *did* turn the *Buddhavacana* into that language. Buddhaghosa seems to have understood the bhikkhus’ suggestions as meaning: “Let us translate, as we translate the Veda into Sanskrit”, presumably referring to a situation where brahmins would be able to explain difficult words in the Vedas, when they recited them, in the same way as we know that Sanskrit commentaries were written upon the Vedas. This, then, might perhaps mean that the ex-brahmins were asking permission to gloss, i.e. explain, the Buddha’s words as they recited them.

If, however, we accept either of the translations “metrically”, or “as desired” for *chandaso*, then this tells us nothing about the language which the Buddha used. In view of this uncertainty, we should perhaps not put too much weight upon this problematic passage, and we should instead try to work the matter out for ourselves in another way.

Is there any evidence that there were, in fact, dialects in use at the time of the Buddha? We have Buddhist texts in both Sanskrit and dialects of MIA. How do we know that the Buddha did not preach in Sanskrit, and the translations from Sanskrit into MIA were not made later?

We must assume that the language(s) of early Buddhism reflected the linguistic background of the times. The beginnings of Prakritic tendencies in Indo-Aryan can be placed very early. In his examination of the history of the Proto-Indoaryans, Burrow pointed out² that in the traces of their language found

¹ *na bhikkhave buddhavacanaṃ chandaso āropetabbam*, Vin II 139, 13–14.

² Burrow, 1973, 123–40.

in the Hittite archives, the word for “seven” is *šatta* (cf. Skt *sapta*), and he referred to this “as having evolved beyond the Proto-Indoaryan stage”.³ We find in the R̥gveda a number of characteristics which we would regard as Prakritic, i.e. they do not present the form which our knowledge of Indo-European philology tells us we should expect in early Sanskrit. Many lists of such forms have been published.⁴ They include the semi-vowel *r* appearing as *u* in *muhur* and in *pitus*; *r* disappearing and making a following dental *-t*-retroflex in *vikāṭa*, beside the expected form *vikṛṭa*. We can see that in some places the metre demands the insertion of a *svrabhakti* vowel, e.g. *Ind^{ra}*. Even though we must make it clear that we are talking about the recension of the R̥gveda which we possess now, which is perhaps no earlier than the seventh century B.C.E.,⁵ nevertheless we can see that such features were already in evidence well before the time of the Buddha. If there were Prakritic features in the literary language of the R̥gveda, the assumption is that the spoken languages of the ordinary people had many more of these features.

Our view about the prevalence of MIA features at this time is reinforced by the fact that about 400 B.C.E., i.e. about the time of the Buddha’s death, the Sanskrit grammarian Pāṇini produced his grammar, the prime aim of which was, we may assume, to establish the form of Sanskrit inflexibly, so that MIA could not affect it—although, as will be clear from what I have just said, to some extent he was closing the stable door a little late. Patañjali’s *Mahābhāṣya*, dating from the second century B.C.E., actually quotes some of the forms which must be avoided, and they include forms which are well attested in MIA.⁶

We need then have no doubt that the MIA dialects had been formed, and were in use during the time when the Buddha was teaching, but we have little knowledge about the dialect geography of the fifth century B.C.E. If we ask: “What was the language of the Buddha?”, then we have to admit that we do not know for certain what language or languages the Buddha spoke.

Any movement, whether religious or social, which opposed the power and status of the brahmanical caste might be expected to make use of a non-brahmanical language. Not only would this be an anti-brahmanical gesture, but it would also aid the non-Sanskrit-speaking element of the population, i.e. those who spoke a vernacular language. It therefore seems very likely that the Buddha’s sermons were preached in a non-Sanskritic language, i.e. a Prakrit, and

³ Burrow, 1973, 125.

⁴ See Elizarenkova, 1989, 1–17; and Pinault, 1989, 35–96.

⁵ For other evidence see Bloomfield & Edgerton, 1932, who give evidence for voicing and unvoicing (§§ 44–79), aspiration and de-aspiration (§§ 80–124), *kṣ/ṭs/ps* > *cch* (§§ 183–86), *y* > *j*, and *j* > *y* (§§ 192–93), the development of *-ṛ-* > *-a/-i/-u-* (§§ 631–44).

⁶ e.g. *āṇapayati*, *vaṭṭati*, *vaḍḍhati* (*Mahābhāṣya* [ed. F. Kielhorn], I, 259).

from the fact that he moved about preaching in various places we can assume that he preached in a number of dialects, varying his language to suit his audience. Much of his teaching life was spent in Magadha (although none of the great events of his life occurred in that area). We can therefore assume that on some occasions, at least, he used the Magadhan dialect of the time, which we may call Old Māgadhī.

This belief might appear to be supported by the fact that the Pāli commentarial tradition tells us that the Buddha's language *was* Māgadhī. Unfortunately the commentators use this name about the form of the canon which they had before them, i.e. the language which we call Pāli, and as we can be quite certain that Pāli is not Māgadhī, we have to examine carefully just what they meant when they said this. It is probable that they believed that the Buddha spoke the words as they were in the canon, i.e. in Pāli, and as they knew he lived and uttered the words in Magadha, they believed that Pāli was Māgadhī.

We do not know precisely the form which Māgadhī had at the time of the Buddha. We can deduce to some extent what form that dialect had at the time of Aśoka, and we know what the grammarians writing some centuries later said were the characteristics of the dialect. Extrapolating from this, we can gain what we may regard as a fairly accurate idea of the main features of that dialect at the time when the Buddha was teaching. Its main characteristics would have been: *l* for *r*, palatal *ś* for all sibilants, *-e* as the nominative singular ending.

If this is correct, then, as I have said, the language of the Theravādin canon which we have is certainly not Māgadhī. That would indicate that the language has been translated on at least one occasion, presumably to meet the needs of the situation. As Buddhism moved from the land of its origin into areas where different dialects or languages were spoken, and as those dialects or languages developed and changed over the course of time, it would seem to be inevitable that some sort of translation process was needed if the Buddha's teaching was not to become unintelligible to those to whom it was preached. Are there any traces of this? Yes, there are. If we analyse the language of the Theravādin canon, i.e. the language which we call Pāli, we can see that for the most part it has features which we would class as western, using this in a linguistic rather than a geographical sense, since we find eastern forms in the version of the Aśokan inscriptions in the West at Sopārā. Nevertheless, we can detect anomalous forms in it, i.e. forms which do not seem to follow the western patterns of phonology and morphology of the language. Some of these have features which are more appropriate to Sanskrit, e.g. consonant groups containing *-r-*, and the absolutes in *-tvā*. Disregarding these for a moment, we can see that there still remain a number of other anomalous forms. There are forms with eastern *kkh* where we should expect western *cch*, e.g. *bhikkhu* and

bhikkhunī; forms with *l* where we should expect *r*, e.g. verbs with the prefix *pali-* instead of *pari-*; forms with *-e* where we should expect *-o*, e.g. *bhikkhave* instead of *bhikkhavo*, and also some nominative singular endings in *-e*, instead of *-o*; examples of the voicing of consonants, e.g. *yādeti*; forms with intervocalic *y* instead of *k* or *t*; forms with *j* where we should expect *y*, e.g. *jantāghara*; dental *n* instead of retroflex *ṇ*, e.g. in *nibbāna*; *v* where we should expect *y*, e.g. *āvuso* from the noun *āyu(s)*; and a small group of words where a consonant group including a nasal has developed in an unexpected way, e.g. *nt* > *nd*, *nd* > *nn* and *mb* > *mm*.⁷

It is generally agreed that these are dialect forms, from one or more dialects, and they may be regarded as being remnants of dialects through which the *Buddhavacana* was transmitted before it was translated into the language which we find in the Theravādin canon, i.e. Pāli.

In what I say in this lecture I am, to some extent, a prisoner of my own nomenclature. I shall be using “Pāli” as the name of a language, and it has become conventional to do so. But Pāli is an abbreviation of *pāli-bhāsā*, which means the language of the *pāli*, i.e. the texts, or the canon. It follows, then, that every feature of the language of those texts, however strange, and however inconsistent with the rest of the language of the texts, is Pāli. We can define it, if we wish, as a western dialect with some eastern features, although this in itself is a matter for debate, as we shall see, but it would, technically, be incorrect to talk about these possible eastern features as anomalous forms. It is even more incorrect to call them non-Pāli forms. Nevertheless, I must call them something in the course of my discussion about what to call them, so I shall in general call them “anomalous” or sometimes “eastern”, and when I want to stress that the Pāli being used is free from such eastern forms, I shall call it western. This may seem confusing, but I hope that all will become clear as I go along.

The presence of these anomalous forms in the Theravādin canon was, of course, noticed a long time ago, and their value as indicators of an earlier form of the *Buddhavacana* has been much studied and debated. We have to ask ourselves what their significance is. Why they were retained from an earlier version, and what conclusions can we draw from their retention? Some scholars have regarded them as evidence of an Urkanon, (“a primitive canon”),⁸ while others have seen them as pré-canonique (“pre-canonical”).⁹ The first of these views pre-supposes that there was a canon in existence at some pre-Pāli time, and while this is possible and indeed quite likely—although this depends to some extent upon the definition which is given to the word canon (something to

⁷ See Norman, 1989C, 369–92 (= CP IV, 46–71).

⁸ e.g. Lüders, 1954.

⁹ e.g. Lévi, 1912, 495–514.

which I shall return in the eighth lecture)—these forms cannot be taken as proving this. All they prove is that some texts, i.e. the ones in which we find the anomalous forms, existed at an earlier date in a dialect or dialects other than Pāli. Even this is overstating the case, because it is quite clear that the use of some of these forms was extended by analogy. I shall return to this in a moment. At this point I will only say that there is probably no one reason for their retention.

If these anomalous forms are remnants of dialects through which the *Buddhavaṇana* was transmitted, then our task is to identify the dialects and the areas in which they were used, in the belief that this will give information about the regions of North India through which Buddhism spread. The problem is to see whether our anomalous forms exhibit any of the characteristic features of any of the dialects about which we have knowledge, in the hope that this will tell us about the regions through which the texts, or parts of them, were transmitted. To guide us in our search we have two main aids: the inscriptions found in India, from the time of Aśoka onwards, and the statements of the grammarians. Both sources are to some extent unreliable, and in any case refer to a time later than the Buddha. I mentioned in the second lecture the belief of some that the Buddha's death is to be dated c. 400 B.C.E. This means that the Aśokan inscriptions are about 150 years after that date, and so we have to allow that much time for linguistic change, and we have to accept that the Aśokan inscriptions cannot be entirely satisfactory as a guide to earlier dialects and dialect forms. Furthermore, although we can surmise that Aśoka intended his inscriptions to be carved in the dialects appropriate to each site, there is clear evidence that this was not always done.

The grammarians were writing many centuries later, from the fifth century C.E. onwards, when the dialects had, in any case, been standardised by being used for literary purposes, especially in the dramas, although a literary dialect or language very often represents a fossilised version of a language used as a vernacular long before, and so may retain very old features. As a third aid, there is also the information which can be gained from Jain texts, which underwent the same type of language changes as the Buddhist texts, i.e. different dialects were used as Jainism spread. As in the Buddhist texts, there is the same difficulty of identifying the languages and the areas where they were used. Nevertheless, we can, to some extent, use Jain texts to help with the identification of the dialects of Buddhist texts, and vice versa.

We know nothing about the translation techniques which were employed by those taking the *Buddhavaṇana* to an area where a different dialect was used, and who were therefore faced with the task of translating into a new dialect. If the donor dialect (i.e. the dialect from which the translation was being made) and the receiving dialect (i.e. the dialect into which the translation was being made)

had differences of phonology and morphology, then much of the translation process could have been done almost mechanically. For example, if the translation was being made from a dialect which did not voice intervocalic consonants into one which did voice them, then the translator simply had to voice all unvoiced intervocalic consonants. If the translation was being made from a dialect where the nominative singular of short *a*-stem nouns was in *-e* into a dialect where the nominative singular was in *-o*, then the translator had simply to change all nominative *-e* endings into *-o*.

There would, however, have been certain items which remained unchanged: 1) words which had a specific sanctity attached to them because they were regarded as technical or semi-technical terms, and were therefore, despite the air of strangeness which they must have presented, too important to change; 2) words which had no equivalent in the receiving dialect, and which therefore had to be retained; 3) words retained by an oversight.

The accuracy of such a translation process depended on the knowledge and ability of the translator. Clearly, the translator had to know something about the characteristics of the donor and receiving dialects. If, by any chance, the donor dialect was one which voiced intervocalic consonants, and the receiving dialect did not voice them, then a mechanical translation technique was likely to produce errors, because it would lead to a situation where consonants which should be voiced in the receiving dialect might become unvoiced. It was all very well changing all past participles in *-ida* in the donor dialect into *-ita* in the receiving dialect, but what about a form like *uppāda*? Since there is a root *pad-* as well as a root *pat-*, should this be translated into *uppāta*, or left as *uppāda*? Only the sense of the passage could determine this, and if the passage was ambiguous, or for some other reason it was difficult to understand the meaning of the passage, then a translator was left to guess, and his guesses might not always be correct. If the receiving dialect had the nominative singular of short *a*-stem nouns in *-o*, then to turn all *-e* endings into *-o* may well have meant that forms which should have had the *-e* ending in that dialect, e.g. locative forms, were also changed, incorrectly, into *-o*.

Let us look at some of these anomalous forms and see what we can deduce from them.

It has been pointed out that the Buddha’s mother *Māyā* was probably not called “Delusion”, but “Mother” (*Mātā*).¹⁰ This development of intervocalic *-t- > -y-* is a characteristic of Māhārāṣṭrī, but it is scarcely conceivable that this detail of the Buddha’s life story was transmitted from Magadha to Mahārāṣṭra (where our evidence is in any case from a later time—probably after the writing down of the Pāli canon) before it entered the Theravādin tradition.

The change of intervocalic consonants to *-y-* is not a feature of the eastern dialect of the Aśokan inscriptions—there are only one or two examples, which can probably be explained otherwise¹¹—but the *-iḱya* forms found at Kālsī, where the other versions have *-ika*, may reflect some such change. We do not know precisely what the writing of the *ky* ligature means, but it is possible that the scribe was writing it because he received *k* in his exemplar, but wanted to show that his own pronunciation or the pronunciation in his area was nearer *y*. Lüders has given some examples of *-ika/-iya* alternations,¹² and the fact that the change of *-k- > -y-* occurs after *-i-* seems to support the view that *iḱya* at Kālsī does indicate that the change *> iya* had taken place or was beginning to take place.¹³ It is, therefore, possible that the sound change found in Pāli *Māyā* reflects the fact that in some part of Magadha, perhaps in the West, towards the Kālsī area, this change was operative.

Moreover, it has been shown that some of the etymologies in the Sabhiya-sutta of the Pāli Sutta-nipāta depend upon their being first pronounced, i.e. composed in, not just transmitted through, a dialect where some, at least, of the intervocalic consonants had developed into *y*.¹⁴ The age of the sutta is shown by the fact that a Buddhist Hybrid Sanskrit version of it exists in the Mahāvastu, where the bhikkhu is called Sabhika—the *-iya/-ika* variation of the name in the two versions being in itself an example of the very change I am talking about. If the age of the text supports the view that the version underlying the Pāli and Buddhist Hybrid Sanskrit versions was composed in Magadha, then this is additional evidence for the existence of this sound change in Magadha at an early date.

Similar considerations would apply to the examples of voicing found in Pāli. It seems clear that they are evidence for transmission through a dialect where voicing was usual, and such hyper-forms as *uppāta* < Sanskrit *utpāda* show that the translator had knowledge of a dialect where voicing occurred. Voicing is a typical feature of Śaurasenī, but there would be considerable problems if we had

¹⁰ von Hinüber, 1991A, 183–93 (187).

¹¹ See Norman, 1970, 132–43 (136–37) (= CP I, 93–107 [98–100]).

¹² See Lüders, 1954, §§ 89–90.

¹³ See Lüders, 1954, §§ 88.

¹⁴ See Norman, 1980B, 173–84 (177) (= CP II, 148–61 [154]).

to assume transmission through that dialect. Although it is not a consistent characteristic of any of the Aśokan dialects, nevertheless there is some evidence for voicing in the Aśokan inscriptions, e.g. *libi* for *lipi*, *thuba* for *thūpa*, and the hyper-forms which occur, e.g. the root *pat-* written for *pad-*, show that the scribes had knowledge of a dialect where voicing occurred.

We find a very small number of forms which show an anomalous development of a consonant group containing a nasal: *hanta* > *handa*; the Buddha's charioteer name *Chanda* > *Channa*; *ālabana* > *ārammaṇa*. These changes are typical of the Gāndhārī dialect as seen in the Gāndhārī Dharmapada, but although this text is said to show the characteristics of the Gāndhārī dialect several centuries earlier than the manuscript itself (which dates perhaps from the third century C.E.), these features are not found in the version of the Gāndhārī dialect which we find in the Aśokan inscriptions in the North-west. They therefore seem to be a later development in that dialect, and it is most unlikely that these forms could be borrowings into Pāli from the Gāndhārī dialect. I should rather favour the view that since our knowledge of early dialect geography in India is so unreliable, there may well have been unattested dialects which left some mark upon early Buddhist texts.

The ending *-e* in place of *-o* is perhaps the most widely attested of the anomalous forms. It is a standard feature of the Māgadhī dialect and the eastern versions of the Aśokan inscriptions, and it also occurs in the Gāndhārī dialect and the Sinhalese Prakrit. We find in the Dīgha-nikāya descriptions of the teachings of the six teachers who were contemporary with the Buddha, and some of these descriptions include nominative singular forms in *-e*,¹⁵ as well as a number of other anomalous forms.

Another place where such *-e* forms occur is in the framework of the Kathāvatthu, a work which is acknowledged by the Theravādin tradition to be one of the last additions to the canon, since it is said to have been recited at the time of the third *saṅgīti* which was held at Pāṭaliputra during the reign of Aśoka. We know, then, both where and when it was composed, and it is therefore not surprising that we find in it a number of features of both phonology and morphology which coincide with the eastern versions of the Aśokan inscriptions, e.g. the way in which *evaṃ* occurs with an emphatic *h-*, i.e. *hevaṃ*.¹⁶

One of the interesting features of this text is the way in which the translators seem to have understood very well the difference between the predominantly western Pāli forms and the eastern forms which are found on almost every one of its 600+ pages. The text consists of a discussion of certain statements, of which 500 were orthodox and 500 unorthodox according to the commentary, although

¹⁵ See Norman, 1976C, 119–21 (= CP I, 240–42).

¹⁶ See Norman, 1979A, 279–87 (282) (= CP II, 59–70 [64]).

these numbers must have been used in a typical “round number” way, since, in its present form the text, in fact, contains less than 300 unorthodox statements. They are set in a framework of question and answer. All the statements, heretical or otherwise, are free from anomalous eastern forms in *-e*, except for one stock phrase which is repeated a number of times. The framework, however, shows two speakers using *-e* and *-o* dialects which were probably completely differentiated at one time, but which, by the time the text had become fixed in its present form, had become mixed in a consistent way. We cannot say whether from the time of the text’s composition one of these dialects used exclusively *-o* forms, as it does in the present version. If it did, then this would be evidence for the existence of a dialect with *-o* forms in Magadha c. 250 B.C.E. The original differentiation could, however, have been between two sub-dialects of Māgadhī, which later translators thought were too similar to be easily distinguishable by non-Māgadhī speakers.

Among the anomalous forms in Pāli, attention has been drawn recently¹⁷ to the word *je*, which is a particle used when addressing women of a lower class. It is explained as being a shortened form of the word *ajje*, which would be an eastern form of *ayya* < Sanskrit *ārya* “noble”. I have great doubts about this etymology. It may be possible to find evidence for a word which originally meant “noble” being used in a pejorative sense, but I would find it difficult to accept this derivation when the word *ajje* from which it is said to have been derived was still in full use as an honorific form of address in the eastern dialect in which *je* must have developed its pejorative sense. I personally believe that *je* is the emphatic particle, which appears as *ye* in Pāli and the Aśokan inscriptions,¹⁸ and I believe that it is an example of the very rare change of initial and intervocalic *y > j* in Pāli, which is seen also in the word *jantāghara*. My doubt about the derivation < *ajje* is supported by the occurrence in Jain Prakrit of *hamje*, which is used in a very similar sense to *je*.¹⁹

There are also hyperforms based upon the development of *y > j*. John Brough²⁰ remarked: “They were aware of the Prakritic tendency to voice intervocalic stops of the literary language, and in attempting to combat this tendency, they occasionally overreached, and produced monstrosities such as *Yamataggi* for *Jamadagni*”. From the existence of the name *Yamataggi* in the Theravādin canon we can deduce that the person responsible for the production of this form was aware of the fact that the dialect from which he was translating sometimes showed initial *j*- where his own dialect showed *y*-, and intervocalic

¹⁷ By von Hinüber, 1993, 101–13.

¹⁸ See Norman, 1967A, 160–70 (= CP I,47–58 [50–51]).

¹⁹ See Schwarzschild, 1961, 211–17 (= *Collected Articles*, 104–16).

²⁰ Brough, 1962, 249.

-d- where his own dialect showed *-t-*. When he came across the form *Jamadagni* (or more likely *Jamadaggi*), he did not know its correct form in his own dialect, doubtless because it was a name new to him, whose meaning and etymology were unknown. In the absence of any knowledge of the correct form of the name in his own dialect, he was obliged to back-form by rule, which led to the hyper-form. This shows that there was a pre-Pāli dialect where the changes *y- > j-*²¹ and *-t- > -d-* occurred. This conclusion supports the possibility that the words showing the change of initial *y- > j-* which I have already mentioned are dialect words in the Theravādin canon.

There is another word in Pāli which seems to be an example of this anomalous sound change, i.e. *niya* < Sanskrit *nija* “own, belonging to oneself”. This may simply be an example of an intervocalic consonant being elided and replaced by a glide *-y-*, i.e. it may be a genuine Māhārāṣṭrī-type form < *nija*, but it is also possible to explain it as a hyper-form, i.e. a translator who knew that intervocalic *-y-* became *-j-* in the donor dialect, wrongly back-formed the *-j-* which he found in his exemplar into *-y-*, presumably not recognising the fact that in this particular case *nija* was correct in the receiving dialect also.

Brough was doubtful about the date when this sound change was operative. He said: “The mere existence of the [Pāli] form *Yamataggi* then forces upon us the conclusion that parts at least of the Pāli canon were translated from a MIA dialect in which initial *y-* had already become *j-*. This seems to demand a seriously late date. But I can only pose the question...”²²

Brough’s problem was that he knew of no evidence for the existence of a dialect which turned *y* into *j*, either initially or intervocalically, at the time of the Buddha. We know that some at least of the Jain canon was transmitted in such a dialect, where the stem of the relative pronoun is, for example, *ja*, where Pāli has *ya*, and the optative ending is *-ejja*, where Pāli has *-eyya*, but the Jain tradition tells us that the canon was not written down until some nine centuries after Mahāvīra’s death, in the fifth century C.E., and although parts, at least, of the Jain canon certainly existed before that date, there is no reliable way of dating any of the phonological features of the languages of the Jain canon. The answer to Brough’s question is that we must surmise that there was an earlier dialect where this change took place, and we have, by chance, a minute portion of evidence to support this view. It has been noted that there is one occurrence of

²¹ The antiquity of the change of *y- > j-* is shown not only by the examples given in *Vedic Variants* (see Bloomfield & Edgerton, 1932), but also by the fact that at *Rāmāyaṇa* 7.4.12 the etymology of *yakṣa* implies a form with *j-* (cf. Pkt *jakkha*). See T. Burrow, review of U.P. Shah: *The Vālmiki-Rāmāyaṇa*, Vol. I, Baroda 1972, in *JRAS* 1974, 74.

²² See Brough, 1980, 42.

an intervocalic *y* becoming *j* in the Aśokan inscriptions,²³ and that is in the word for “peacock” in RE I(G). It occurs as *majūla* at Kālsī and Jaugaḍa, with the eastern *l* replacing the *r* of Sanskrit *mayūra*, and as *majura* in Sh and M, with the expected western *r*. The appearance of *j* at four sites (the fifth, Girnār, has *mora*, just like Pāli) suggests very strongly that *j* was taken over without correction by the scribes at all those sites from the exemplars which they received. The dialect at Sh has a tendency to turn *-j-* into *-y-*, and would therefore be unlikely to do the opposite, unless there was a good reason for doing so.

The change of *-y- > -j-* is, however, anomalous in the Aśokan inscriptions. It seems not to be a feature of the dialect which Aśoka’s secretariat at Pāṭaliputra used, and it presumably represents a remnant of Aśoka’s own dialect, which was probably a sub-dialect of Māgadhī, i.e. a dialect spoken in just the area where we would expect the early language of Buddhism to be in use. It seems very probable, then, that these words with *j* in place of *y* came into Pāli from an eastern dialect. This might have been the actual dialect used by the Buddha, but it was at any rate one of the dialects into which his teachings were translated at an early stage in the history of Buddhism.

It is to be noted that this deduction, made on the basis of one single word, implies that Aśoka’s own dialect differed from the eastern dialect attested in his inscriptions, because if we assume that his dialect changed *y > j*, then the nominative singular of his form of the relative pronoun would be *je*, whereas we know that the eastern versions of his inscriptions have a form without initial consonant, i.e. *e*. This implication need present no difficulties, because there is also evidence that Aśoka’s own dialect had palatal *ś* for the sibilant,²⁴ whereas no version of the Aśokan inscriptions, other than those from the North-west, has this sound except by scribal idiosyncrasy.

From this, then, we can deduce that at the time of Aśoka, and earlier in the case of some Pāli texts which we can, with great probability, date to a pre-Aśokan time, almost all the sound changes which we have marked out as anomalous in Pāli can be shown to occur somewhere in the East, although sometimes only as a single example in the Aśokan inscriptions.

We must therefore conclude that the information we have hitherto had about dialects in Aśoka’s time, let alone at an earlier time, is deficient, since it by no means tells us about all the dialects which were in use. The fact that, if we look carefully, we can augment this information by detecting hints of other dialects is, in itself, of great importance, because it means that we can use the same technique of marking out anomalous forms as we have employed in the Pāli canon, to mark out the anomalous forms in the eastern versions of the Aśokan

²³ See Norman, 1980A, 61–77 (74, note 43) (= CP II, 128–47 [144, note 1]).

²⁴ See Norman, 1980A, 61–77 (65) (= CP II, 128–47 [133]).

inscriptions, in an attempt to find the dialect which Aśoka himself used—a dialect which is not represented, in its entirety, in any version now extant.

If our conclusion is correct, then, we can assume that the dialects in use, even in a limited area such as Magadha, showed variations. This should not surprise us, because we should expect neighbouring areas, even neighbouring villages, to have very slightly divergent dialects. We can also assume that traces of this linguistic diversity were retained when the sermons, which had been preached in different areas, were first collected together and their language was homogenised. We can guess—but it is nothing more than a guess—that when the first collection was made and homogenisation began to take place, examples of divergence were even more numerous, since, as sub-dialects of the Magadha area, these variations were probably not sufficiently great to cause difficulties for speakers of other sub-dialects, and there was therefore no need to remove them.

Our question must be: on the assumption that the process of homogenisation was continued, while Buddhism was still, for the most part, confined to the Magadha area, so that fewer and fewer of the anomalous forms remained, why do we find that some of those forms survived the major translation into a western dialect which must have taken place at a later date?

Some of these anomalous forms that I have been discussing are called, by some, “Māgadhisms”, because they conform to the pattern of Māgadhī, as described by the grammarians. Others, however, say that the term Māgadhism is misleading, because it takes for granted that these forms are taken over from Māgadhī.²⁵ It is true that it is probably not entirely accurate to regard all such anomalous forms as Māgadhisms, in the sense of their being words which were originally in a Māgadhī version of the *Buddhavacana*, but were retained deliberately or accidentally at the time when translation took place into a non-Māgadhī dialect.

Some of these forms are certainly found in other dialects, e.g. nominative singular forms in *-e* occur in the Gāndhārī Prakrit, probably as the remnant of a linguistic area which was split into two by a later tribal movement, after which the two halves migrated to very different areas, one to Gandhāra and the other to Magadha. Nevertheless, to suggest that the North-west was the source of such forms would demand a complete re-appraisal of our ideas about the way in which the Buddha’s message was spread.

It has also been pointed out that other explanations can be given for some of these forms, but some of these alternative explanations seem unnecessarily complicated. To say that *pure* < Sanskrit *purah* was taken over into Pāli as *puro*, but then became *pure* by analogy with *agge*, etc., or by the dissimilation of the

²⁵ See Bechert, 1991A, 3–19

vowels *u* and *o*, leads to a situation where we have to assume that a Māgadhī form was converted into the correct western form, but then underwent a change which converted it back to the same form which it had in Māgadhī, but we are not, nevertheless, to regard it as a Māgadhism.²⁶ There is unlikely to be agreement about such views, since there is no way of proving or disproving either hypothesis. In such circumstances, I should like to propose a theory of “the economy of development”, which means that an explanation of any sound change by a single stage of development is more likely to be correct than an explanation by a multiple development. If we adopt this theory, then to take it as a Māgadhism provides a simpler explanation.

What is, however, clear is that every Māgadhism does not prove that the passage in which the Māgadhī form occurs is old and dates from the Māgadhī period of Buddhism, and certainly those who say²⁷ that we can rely too much on such details are correct. Once a standard procedure had been adopted, e.g. of writing *bhikkhave* in a specific context, and it was applied to newly created texts, the occurrence of such Māgadhisms tells us nothing about the original language of the text in question, i.e. the occurrence of a Māgadhism in a text does not prove that the text was originally in the Māgadhī dialect. There was a great deal of extension of use by analogy. In just the same way, as we shall see in the eighth lecture, phrases in common use in canonical texts do not prove that a text is canonical, because such phrases were used by medieval writers to give their texts a veneer of canonicity.

A recent discussion of some of the features I have just been talking about was entitled “From colloquial to standard language”.²⁸ I find the use of the word “colloquial” slightly strange, because to my ear it has something of a pejorative sense. If the Buddha preached in the dialect of the local people wherever he went, then we would say, I think, that he made use of the local vernaculars. Many of these features also occur, as we have seen, in the language of the eastern versions of the Aśokan inscriptions, which is normally identified as the administrative language of the secretariat, and which can scarcely be described as colloquial. The same discussion describes the difference between the particle *je* and the feminine vocative *ajje* as representing a distribution of colloquial and standard language.²⁹ Again, this usage of “standard” is strange, in the first place because it is not justified, but it is simply taken as fact that there *was* a standard language in Buddhism, and in the second place because the term “standard” is not defined. We are left to deduce that it means a language without

²⁶ See Bechert, 1991A, 13.

²⁷ See Bechert, 1991A, 12.

²⁸ von Hinüber, 1993, 101–13.

²⁹ von Hinüber, 1993, 102.

“colloquialisms”, on which, perhaps, all existing versions of Buddhist Indo-Aryan dialects and languages are based. It is hard to see how Pāli could fit into such a pattern, because, as we have seen, Pāli does include such forms. Perhaps “standard” means a literary type of language, but no evidence is given that such a dialect ever existed. I would doubt that changes were made by translators simply because some words were thought to be “colloquialisms” and therefore “non-literary”. It seems much more likely that eastern forms were unacceptable, for phonological and morphological reasons, to the western Prakrits into which the early Buddhist texts were translated, and they therefore had of necessity to be changed, except when there were reasons for their retention.

It is clear that many of these anomalous forms were retained because they were technical or semi-technical terms, e.g. *bhikkhu*, *bhikkhunī*, *nibbāna*, *bhūnahū*,³⁰ with the last three showing a replacement of *-ṇ-* by *-n-* which in the Aśokan inscriptions is typical of the eastern dialect(s). These words were, in fact, part of the basic vocabulary of early Buddhism. We can surmise that the spelling of *khāṇa* < Sanskrit *kṣāṇa* “opportunity” was retained because of its repeated occurrence in the phrase “do not let the opportunity pass you by”. The expected western form *chaṇa* exists, but in the sense of a particular sort of opportunity, namely “a festival”. The particle *je* was doubtless retained in its eastern form because the genuine western form *ye* was not used in this particular sense in the receiving dialect into which the sermons were translated, although *ye* is used as an emphatic particle after infinitives in Pāli in exactly the same way that *je* is used in Jain texts.³¹ A word like *āvuso* also had a semi-technical sense, in as much as the Buddha had laid down the specific circumstances in which it was to be used.³²

It seems to me that the words attributed to the six teachers probably reflect (despite the view of some to the contrary³³) the actual dialects of their teachings, at least as they were remembered at the time of the composition of the texts. In support of this suggestion is the fact that comparable views ascribed to heretical teachers in Jain texts show close verbal similarities. Other forms were probably kept by mistake, or because the translators did not recognise the verbal root, e.g. *āvudha* in place of *āyudha* “weapon”. In some cases it seems clear that the sense was ambiguous, and the translator could not determine the correct form, e.g.

³⁰ See Saksena, 1936, 713–14.

³¹ von Hinüber is misled by the *Āgamaśabdakośa* (see *Überbl* § 49) into believing that this usage is only attested once. As Schwarzschild (1961) makes clear, it is quite common. For additional examples see Oberlies, 1993, 78, s.v. *je*. For Pāli *ye* after infinitives see Norman, *EV* II 418.

³² D II 154, 9–15.

³³ Bechert, 1991A, 13.

there is a verse in the Theragāthā,³⁴ where we have three forms ending in *-e*, *sacce*, *atthe* and *dhamme*, which can be eastern nominative or western locative case forms, and we cannot be certain how we should translate them. The Pāli commentary takes all three as locatives, but this seems rather forced, and if we look elsewhere for an interpretation we have a choice of translating “In truth the meaning and the doctrine are grounded” or “Truth is grounded in the meaning and the doctrine”.³⁵ The BHS translator took it in the second way, and changed one *-e* form to *-am*, making a nominative *satyam*.

After examining the anomalous forms in Pāli, we can therefore say that we have evidence that the texts of the Theravādin canon was transmitted through a mixture of dialects or sub-dialects, almost all of which can be shown, or can be surmised, to have been employed in the East at the time of Aśoka, and probably earlier.

I said earlier that the great majority of forms in Pāli are what we would call western, using this in a linguistic rather than a geographical sense. This would indicate that the last recension before the writing down of the canon was in an area where a western style Prakrit was in use, at least for literary purposes. This was likely to be in the West of India, but not necessarily so, since, if an eastern dialect could be used in the West, as we have seen at Sopārā, there is no reason why a western dialect should not be used in the East of India. Theoretically, the recension could have been made at the time of writing the canon down in Sri Lanka, but this suggestion causes problems, because the Sinhalese Prakrit which we find in the inscriptions of the first century B.C.E. does not resemble Pāli very closely—it has, for example, eastern nominative forms in *-e*—and if we are looking for an area where a western dialect was used, then this does not seem to be a likely candidate.

The presence of Sanskrit forms in Pāli used to be taken as evidence for the belief that Pāli was one of the oldest of the MIA dialects. This idea was based upon a view that MIA showed a steady progress from Sanskrit to the last form of MIA, i.e. Apabhraṃśa—the very latest stage before the emergence of the New Indo-Aryan languages. A dialect which retained a proportion of Sanskrit forms was therefore thought to be closer to Sanskrit, not only in form but also in time. As more was learned about MIA, however, and as more and more texts were found in Sanskrit or various forms of Sanskrit, owing a smaller or larger debt to underlying MIA dialects, it became clear that this view was wrong. As some of these Sanskrit forms in Pāli were recognised to be incorrect, and not genuine Sanskrit forms at all, e.g. *attaja* is Sanskritised as *atraja*, instead of *ātmaja*, it

³⁴ *sacce atthe ca dhamme ca āhu santo patiṭṭhitā*, Th 1229cd.

³⁵ In EV I 1229, I took the former interpretation; Udāna-v 8.14 takes the latter, writing *satyam*.

became clear that the Sanskrit forms in Pāli were late additions to the texts, i.e. they represented the result of a limited re-introduction of Sanskrit or quasi-Sanskrit forms into the MIA original, not the preservation of old forms.

We have no direct evidence as to where and when the Sanskritisms were introduced into the Theravādin canon. Although a start had probably been made before Theravādin texts were taken to Sri Lanka, nevertheless, we believe that the greater part of the Sanskritisms were introduced in Sri Lanka, if only because we would start to date the start of Sanskritisation rather late, probably not before the second century B.C.E.

Whether other changes, beside Sanskritisation, were being made to the language of the canon at that time we do not know. Since it is clear from the commentators' explanations that updating of the language did occur, e.g. present participles with the old historical nominative singular ending *-am* in the canonical texts are explained by younger forms with the analogous ending *-anto* in the commentaries, we may well be correct in believing that such things could also happen in the canon, although not in verse texts, where the metre would act as a constraint upon any changes which would alter the metrical length of a word.

We have no evidence that Pāli, either with or without Sanskritisms, coincided with any historical language or dialect. It is not clear what we should call such a language. Some³⁶ call it “artificial”, and certainly there are artificial features, such as the incorrect back-formations just mentioned. The English language, however, has similar artificial features, e.g. the *g* in *sovereign*, by analogy with *reign*, the *s* in *island* by analogy with *isle*, or the *p* in *receipt* because it is derived from Latin *receptum*, but no one regards English as an artificial language, reserving the term for such invented languages as Esperanto. Others call the Pāli language “literary”, and certainly it is the language of a literature, although it is not literary in the sense that it represents the refined form of a popular dialect.

However we choose to describe Pāli, the chronicles tell us that in the first century B.C.E. the Pāli canon was written down in Sri Lanka, and it is with the use of writing and the effect which it had upon Buddhist texts and Buddhism itself that the fifth lecture is concerned.

³⁶ See von Hinüber, 1982, 133–40 (140).