

The origin of Pāli and its position among the Indo-European languages.

K. R. Norman

Pāli is a dialect of Middle Indo-Aryan, i.e. one of the dialects which lie chronologically between Old Indo-Aryan (=Sanskrit) and New Indo-Aryan (=the modern languages of North India and Sinhalese). The Indo-Aryan languages of India belong to the Indo-European family of languages, and the name Indo-Aryan stands for "the Indo-European languages of India".

Where and how the Indo-European languages came into being can only be a matter for surmise. The earliest period about which we can come to any conclusions, with any degree of probability, concerns the language family at a time immediately prior to the historical period. It is likely that c. 5000 B.C. a group of people speaking closely allied dialects of Indo-European were living somewhere in the area of southern Russia or northern Anatolia. Since the group of Indo-European languages must have come into existence many thousands of years before that, and since we can surmise that from the very beginnings of that group the various dialects had begun to diverge and develop into different languages, there must at one time been other, less closely related, Indo-European languages. Of these we know nothing.

For reasons not known to us, those people speaking Indo-European languages began at that time to separate and move off in different

directions. It is possible that this movement coincided in time with a change in life-style, whereby they had begun to settle down and practise agriculture. If this was so, then the need for new land as the population began to grow would have provided the cause for their movement. It is also possible that the move was provoked by climatic changes, which led to a shortage of wild animals if the peoples were still hunters, or to a failure of crops if they had already become farmers.

As they moved their language continued to develop and, in the course of time, the dialects diverged so much from those of their old neighbours that they gained the status of separate languages, and eventually turned into the various branches of the Indo-European family known to us: Latin, Greek, Slavonic, Germanic, Celtic, Armenian, Albanian, Iranian and Indo-Aryan, to mention some of those still extant, and Tocharian, Thracian, Phrygian and Hittite, to mention some which have died out.⁽¹⁾

A group who were the ancestors of the Iranians and the Indo-Aryans and who, for that reason, are referred to as the Indo-Iranians, had reached an area to the north of present-day Iran not later than 2000 B.C., moving down from Central Asia. They called themselves *Ārya*, and for this reason they are sometimes known as Aryans. The name *Īrān* is a form of the name *Ārya*, and has been applied to the country where one group of the Indo-Iranians settled. The Indo-Iranians split into two.⁽²⁾ One, the Iranians, remained in the region of the River Oxus. The others, the Indo-Aryans, began to move. Some of them moved down into Iran and then to the West, and we find traces of their language in the documents of the Mitanni kingdom in North Mesopotamia c. 1500-1300 B.C.

A treaty concluded between the Mitanni and the Hittites c. 1350 B.C. invokes by name certain divinities including Indara, Mitrašil, Našatianna and Uruvanaššil, who are clearly identifiable as Indra, Mitra, Nāsatya and Varuṇa.⁽³⁾ A treatise on the training of horses discovered in the archives of the capital of the Hittites includes such technical terms as *aika-vartanna*, which is easily identifiable as being connected with Sanskrit *eka-vartana* (“one turn”).

Others moved eastwards and began to enter India. It is probable that the Iranians began to move south c. 1400 B.C.,⁽⁴⁾ and overcame any Indo-Aryans who were still living in the South of Iran, thus cutting off the two branches of the Indo-Aryans from each other. The Western branch were presumably absorbed into the Mitanni and other Anatolian peoples, and eventually ceased to exist as a separate race. We hear no more of them.

The Eastern branch probably remained for a while on the boundary of Afghanistan and India, and moved into India in two or more waves,⁽⁵⁾ possibly spread over several generations. When they moved into India, they met with the indigenous inhabitants of North India, and in particular the inhabitants of the Indus Valley. From the number of Dravidian loanwords in the Indo-Aryan languages,⁽⁶⁾ dating from the very earliest Indo-Aryan texts, it is clear that they met Dravidians at an early stage of their move into India. Whether these Dravidians were the inhabitants of the great cities of the Indus Valley,⁽⁷⁾ or whether there were people speaking a language or languages in that area, which had completely died out by historic times, is a question which has produced a great deal of discussion in recent years.

Since we can assume that the dialects of the Indo-Aryans were

continuing to develop all the time, it is very likely that each wave of Indo-Aryans who moved into India spoke a very slightly different dialect. It is also likely that, if we are talking about a time span of several generations,⁽⁸⁾ their cultural and religious backgrounds would have also been slightly different. It seems very possible that there would have been conflict of some sort between the invaders and the indigenous inhabitants, but it is also likely that after that initial conflict the two races would have started to become assimilated, and there would have been an inter-action between the language of the inhabitants and the Indo-Aryan language(s).

It is also probable that when the later waves of Indo-Aryans moved into India they would have been confronted by these groups of mixed language and culture users, who represented a mixture of the preceding invaders and the indigenous populace, and there would have been further conflict. The result of this is that besides the dialects which we might expect the various groups of Indo-Aryans to speak, there were also the variations which arose from the fact that they had different amounts of Dravidian substrate mixed in with them. Signs of that Dravidian substrate are already evident in the oldest Indo-Aryan material we possess, i.e. the Ṛg-veda, where we find such words as *kūṭa* "hammer," *daṇḍa* "stick", and *phala* "fruit", for which a Dravidian origin is plausibly claimed.⁽⁹⁾

It seems very likely that the various groups of invading Indo-Aryans had their own literature, in their own distinctive dialects. We find some trace of this in historic times—the Ṛg-veda is in a slightly different dialect from the other Vedas and from the other

forms of Sanskrit which developed into Classical Sanskrit, e.g. there is an alternation between *-r-* and *-l-* in certain words, and *-l-* and *-lh-* occur in place of *-ḍ-* and *-ḍh-*. There is a single example of the change of final *-as* to *-e*, not *-o*, which is perhaps a dialect form in the Ṛg-veda, borrowed from some other source.

There are also some signs of what we might call “Prakritisms”, i.e. the first indications of the divergences from the Old Indo-Aryan dialects which were to develop into Middle Indo-Aryan. So we find that the expected genitive singular ending **-tṛs* of *-tar* stems is replaced by *-tus*, as the vocalic *-r-* sound is replaced by *-u-*, e.g. *pitus* from *pitār-*, instead of **-pitṛs*. There are also examples of the typical Middle Indo-Aryan phenomenon of the retroflexion of a dental *-t-* by a vocalic *-r-*, which then disappears, e.g. *vikāṭa* “horrible”, besides *vikṛta* “changed, mutilated”. It is interesting to note that in the traces of Proto-Indoaryan found in the Hittite archives the word for “seven” is *šatta*⁽¹⁰⁾, showing that the Middle Indo-Aryan feature of assimilation of *-pt->-tt-* had already begun to operate in that language also. This suggests that the development of such features had already begun before the Indo-Aryans entered India.

We know from the evidence of Iranian and from the differences between the Ṛg-veda and later Sanskrit that there must have been dialects of Old Indo-Aryan which turned all *-r-* and *-l-* sounds into *-l-* and others which turned them all into *-r-*, and still others which mingled the two sounds in different proportions. It is very likely that there was one dialect which turned all *-r-* and *-l-* sounds into *-l-*, and had the nom. sg. of *-a* stems in *-e*, and also turned all three sibilants into *ś*. The speakers of this dialect moved to

the east of India, and there is evidence that by historical times the dialect was being used in the region of Magadha.

As the Indo-European speakers moved towards the East of India they came into contact with tribes speaking languages which belonged to another linguistic family, the Munda group. They also borrowed words from this source,⁽¹¹⁾ although not on the scale on which borrowings from Dravidian had taken place. They also borrowed widely, especially in the field of agriculture, from an unknown source.⁽¹²⁾ Some of these borrowings are also found in the Dravidian languages, from which we can deduce that both the Indo-European and the Dravidian peoples were at one time in contact with speakers of a language from yet another linguistic group. If the inhabitants of the cities of the Indus Valley civilisation were not Dravidians, then it is possible that they were this people speaking an unknown language. Clearly they were still in existence in India at some time after the arrival of the Indo-European speakers in the North-West of India, having already been in contact with the Dravidians before the Indo-Europeans arrived.

It seems likely that by about 500 B.C., if we date the beginnings of Buddhism and Jainism then (or about a century later if we follow the later dates⁽¹³⁾ for the beginnings of those religions), the vernacular dialects which, following the terminology of the Indian grammarians, we call Prakrits,⁽¹⁴⁾ were appreciably different from the Sanskrit of the brahmanical priestly class. Not only were there morphological and phonological differences of the sort which I have described, but there were also differences of vocabulary, and variations in the way in which words were formed.

The differences of vocabulary were not simply due to the adop-

tion of words for animals or plants which were unlikely to be a part of the brahmans' language, but also included words which can be shown to belong to Indo-European, e.g. the particle *cia* "indeed" (cf. Latin *quidem*)⁽¹⁵⁾. These serve as additional evidence for the existence of dialects of Old Indo-Aryan besides the language of the Ṛg-veda.⁽¹⁶⁾ The different types of formation are very often based upon a different grade of vowel gradation, e.g. *tuvaṭa* "quick" < **tvṛta*, a past participle showing the weak grade *tvṛ-* of the root *tvar-* "to hurry", and *turita*, showing the weak grade *tur-* of the same root, whereas Sanskrit has only the form *tvarita*,⁽¹⁷⁾ which shows the *guṇa* grade of the root. Since vowel gradation is an ancient feature of Indo-European, we can be sure that such alternative forms are old. It is probable that the adverbs *idha* "here" (which shows an older form of the suffix *-dha* than Sanskrit *iha*), and *sabbadhi* "everywhere" (which shows a suffix akin to the Greek suffix *-θη*) also go back to dialects of Old Indo-Aryan.⁽¹⁸⁾

The brahmanical religion was based upon a very strict adherence to the form of ritual and to the sacred texts upon which the ritual was based. The changes which were taking place in the dialect of the brahmanical caste, as the Indo-Aryan dialects continued to develop, were beginning to cause difficulties, for they led to a situation where there was a danger that rituals were not always being carried out correctly. It was essential that every word used in the recitals which took place at the sacrifices was pronounced in its correct form. Failure to do this might mean that the sacrifice was vitiated, and the goal sought by the sacrificer(s) would not be attained. The brahmans went to great lengths

to ensure that the pronunciation should be correct, by developing special methods of reciting.⁽¹⁹⁾

The form of the language, as well as the pronunciation, was changing. The subjunctive was going out of use, and prepositions were being attached to verbal roots, instead of being quite free in their position. Various nominal and verbal forms were disappearing, and the pattern of nominal composition developed, with some types of compound going out of favour while others became more popular. While these changes were not numerous in the speech of the dominant brahman community, they were much more common among the less well educated mass of the population. In their speech the subjunctive mood and the dual number disappeared, except for one or two fossilised forms, and all final consonants disappeared, except for the nasal *anusvāra*. The changes in pronunciation were even greater. All conjunct consonant groups were simplified, either by assimilation or by resolution by means of the insertion of a *svarabhakti* vowel. Intervocalic consonants were weakened, either by voicing or by elision. The pattern of development and change seems to have varied from area to area in North India, and it is possible to categorise the changes which distinguish these Middle Indo-Aryan dialects on a geographical basis, with the Western dialects tending to assimilate consonant groups and develop the group *-kṣ-* to *-cch-*, while Eastern dialects resolved groups and developed *-kṣ-* to *-kḥh-*.⁽²⁰⁾

Faced with these changes, the brahmans set about protecting their language from corrupting influences, and they began to develop systems of grammar designed to define the nature of Sanskrit, and thus exclude non-Sanskritic features. The culmination

of these grammatical productions was the grammar of Pāṇini, who wrote c. 400 B.C. It is likely that such grammars were intended only to describe the form of language which was considered correct, but they were soon accepted as normative, prescribing the form which must be adopted by anyone wishing to produce correct Sanskrit. The result of this was that the form of Classical Sanskrit was frozen at the stage it had reached at the time of Pāṇini, and from then on anyone who wished to write Sanskrit wrote according to the rules which the grammarians had prescribed.

The Buddhist and the Jain texts tell us that Gotama the Buddha and Mahāvira the Jina both preached in Prakrits.⁽²¹⁾ We may assume that when they preached, they used the vernacular dialects of the areas where they were touring. They were, to some extent, attacking the brahmanical religion in their teachings, and the existence of dialects which were quite distinct from the language of the brahmins enabled them to give teaching which differed not only in content, but also in the form of its language. We do not know the precise details of the dialect pattern of the 5th and 4th centuries B.C. in the Magadha region where both the religious leaders lived and preached, but we can identify certain features from the information which we obtain from later times, especially the inscriptions of Aśoka in the middle of the third century B.C., and the writings of the grammarians, etc. This information is not necessarily entirely reliable, since the grammarians were writing at a time some centuries after the various vernaculars had ceased to be current, and they were accordingly basing their observations upon the literary uses of dialects in dra-

mas and other literature.

Even the inscriptional evidence is not entirely satisfactory, since the Aśokan material consists of a number of versions of Aśoka's inscriptions in various parts of India, which we assume were translated into the local dialects to the best of the administrators' ability. In the case of the dialects used in the versions of the Aśokan inscriptions carved in the North-West of India, we can tell from the languages which are spoken in those areas at the present time that the features which we find in the Aśokan inscriptions there, e.g. the restoration of all three sibilants (ś, ṣ, and ś) and the writing of consonant groups containing *-r-*, were genuine features of the area, and we can deduce that the scribe at Shāhbāzgaṛhī produced a fairly accurate version of the local language. Even so, there are a number of Eastern forms, which we call Māgadhisms, included there, which probably represent oversights on the part of the translator, rather than reflecting genuine features of the dialect. The version at nearby Mānsehrā includes even more Māgadhisms, which again may reflect an unsatisfactory translation technique.

The version at Gīrnār in the West shows, for the most part, the expected Western characteristics, e.g. the nominative singular of *-a* stems in *-o*, and the consonant group *-kṣ-* developed to *-cch-*, but there are also a number of idiosyncracies which perhaps represent the scribe's own preferences, rather than the local dialect. There are a number of consonant groups containing *-r-* which we should not have expected in that part of India at that time, and it is possible that the scribe was deliberately including anachronisms or even showing his knowledge of Sanskrit. At nearby

Sopārā the fragments which have been found of two Rock Edicts seem to show that the scribe translated mechanically, turning every *-l-* sound into *-r-*. At Eṛṛaguḍi in the South, the scribe reproduced the Eastern form of his exemplar almost without change, although since the site is in Dravidian-speaking territory the dialect can scarcely represent the language of the indigenous people, although it may well reproduce the language of the administrators, who had probably come from the North of India.

Similarly the later inscriptions at Mathurā in the North of India vary in a way which can scarcely represent the actual dialects spoken in the area, or even the dialects of those who were making the donations which are commemorated in the inscriptions. It seems likely that the scribes were translating, again to the best of their ability, into what was thought to be appropriate at the time. At a time when it was thought preferable to write inscriptions in Sanskrit, the inscribed results probably indicate the competence of the scribe to compose in that language.

We know from the Jain canonical texts that the Jina toured and preached in Magadha, and it seems likely that some, at least, of his teaching was in a dialect which corresponded exactly to the Māgadhi dialect as described by the later grammarians, i.e. it had nom. sg. in *-e*, all *-r-* and *-l-* sounds appeared as *-l-*, and all sibilants as *-s-*. The inscriptional material in this dialect is very restricted in quantity, but some does exist. There is no Jain scripture exactly in this dialect, but there are traces of it in later Jain texts, and we must assume that this “old” or “genuine” Ardha-Māgadhi was originally more wide-spread. The older part of the Jain canon, as we have it now, is in a dialect which

the Jain commentators call Ardha-Māgadhi. There are different explanations given for this name, some commentators stating that it was so called because it was the language of “half of Magadha”, but it seems more likely that the name was given because it had only some (“half”) of the features of the true Māgadhi.⁽³¹⁾ So it has nom. sg. in *-e*; most, but not all, *-l-* sounds become *-r-*; all sibilants become *-s-*.

Since the Buddha lived and preached in an area which largely overlapped that of the Jina, it seems quite likely that the pattern of language of the Buddha’s preaching followed similar lines, and we can assume that in the course of his preaching tours, he varied his language to suit his audience, and used the dialect appropriate to the region in which he found himself.⁽³²⁾ At that early date, the difference in the dialects was probably not great, and the differences would be of three kinds: (1) the phonetic or phonological variety, i.e. whether the dialect had *-r-* or *-ś-*; or whether consonant groups were retained, or assimilated, or resolved; (2) the morphological kind, with nom. sg. in *-e* or *-o*, and loc. pl. in *-ehi* or *-esu*; (3) vocabulary, with different words being used for identical objects, e.g. pots and bowls having different names in different places.⁽³³⁾

It can be assumed, therefore, that from the beginning the teachings of the Buddha and the Jina were not restricted to a single dialect or language. It is probable that their hearers would repeat the sermons which they had heard in their own dialects, and it is also likely that as the Buddha’s followers went on preaching tours they too would recite the sermons in the dialects of their audiences. As time went by there would also be a need to change the language of the sermons, as the language in which they had

been recited began to become archaic.

As Buddhism became more established in North India, it began to centre around various *vihāras*, and it is possible that each *vihāra* began to build up a collection of sermons, which were probably still in different dialects, as they had been remembered and handed down from teacher to pupil. It is likely that a certain amount of homogeneity of language was imposed upon such teachings, at least to the extent of making them easily intelligible to all the inmates of a particular *vihāra*. The same sort of homogeneity was probably imposed upon these collections by the group recitations (*saṅgītis*) which the Buddhist tradition tells us took place from time to time, the first immediately after the death of the Buddha.⁽³⁴⁾

As Buddhism began to split up into various sections as a result of doctrinal and other differences, the separate *vihāras* became the seat of different sects, and as missionary expeditions set out, so the dialect which formed the basis of the teachings of the *vihāra* which had sent out the missionaries became the dialect or language of the off-shoot which was founded in this way. The homogeneity or consistency of the language of the "scriptures" of each sect or *vihāra* depended very much upon the ability of the monks to translate (or "transform", since the difference between each dialect, at least in the early period of Buddhism, was probably not great enough to merit the use of the word "translate"). It is likely that at first each monk made the necessary changes as he moved from region to region, but it is probable that after a while the language of the scriptures was rendered more standardised.

The tendency towards standardisation probably began very early in the history of Buddhism, because we read that after the First Council the various parts of the Buddhist canon were handed over to different theras and their pupils for safe keeping and transmission. This was the beginning of the system of *bhāṇakas* (“reciters”)⁽³⁵⁾, and although it is questionable whether such a system could have started as early as tradition says, it must have started fairly soon in the first century after the Buddha’s death, while members of the audiences who had heard the Buddha and his chief disciples could still remember what they had heard. Such acts of memory were not necessarily restricted to *bhikkhus*. The Buddha himself mentions the possibility of *bhikkhus* having to go and listen to a layman reciting a *sutta* to ensure that it would not be lost at his death.⁽³⁶⁾

The system of *bhāṇakas* meant that not only was the language of each text standardised to some extent by the recitation process, but also the contents. Recitation must have led to the introduction of stock phrases and lists, as the original inconsistency which must have existed in the earliest *suttas*, as they were first remembered and recited, was edited out and a standard form and order of epithets, etc., was introduced into the texts. Although all the transmission of the Buddhist texts was done orally in the early period, the use of writing for administrative and literary purposes must have been increasing, and there is evidence for its growing use for religious purposes.⁽³⁷⁾ The fact that a fully fledged writing system was available for Aśoka to use indicates that writing must have existed in India for some considerable period before Aśoka, and it is probable that his decision to have his edicts

publicly inscribed served to encourage its wider use.

At some time after the introduction of Theravādin Buddhism into Ceylon, war and famine and the destruction of *vihāras* led to a breakdown in the *bhāṇaka* system, and to a situation where some texts were known to a very few *bhikkhus*. Buddhaghosa records the fact that there came a time when only one *bhikkhu*⁽³⁸⁾ knew the Niddesa, and from fear of its disappearing completely the *thera* Mahārakkhita was persuaded to learn it from this one *bhikkhu*, and other *theras* learnt it from Mahārakkhita. This made the *theras* in Ceylon realise the fact that the whole canon could disappear if the oral tradition died out. The result of this was that after a *saṅgīti* held in the Āloka-vihāra during the reign of Vaṭṭagāmiṇī in the first century B.C., the entire canon which had been collected together by that time was set down in writing.

The Theravādin canon, when written down in this way, was called *pāli*, in distinction to the commentaries, which were called *aṭṭhakathās*. An examination of the language of the canon, which the commentators referred to as *pāli-bhāsā* “the language of the canon”, shows that, although for the most part the language is homogeneous, with a preponderance of forms which we would describe as “Western”, e.g. the nominative singular of short *-a* stems in *-o*, it is not a single consistent language, but shows a number of dialect features which we can identify as belonging to various dialects which were spoken at various times in North India. It is clear, therefore, that the language of the canon, which we call Pāli, through a misunderstanding of the term *pāli-bhāsā*,⁽³⁹⁾ is a mixture of dialects.

It is possible that this mixture reflects the variety of dialects

which were used by the Buddha and his followers as they moved around North India, using the various local dialects as seemed appropriate in their preaching tours. So the relatively small number of Māgadhī features which we can identify would be the remnants of the sermons which they preached in that area. In fact however, the Māgadhī features consist mainly of individual words which show *-l-* instead of *-r*, e.g. the prefix *pali-* instead of *pari-*, or a number of nominative singular forms in *-e* instead of *-o*. It appears that a number of the phrases where these forms are found are quotations of the views of other teachers. It is probable, therefore, that these are genuine remnants of other sects' texts,⁽⁴⁰⁾ but many of the features seem rather to be the result of an inadequate translation technique, whereby those who were responsible for adapting the collection of material into one standard dialect omitted by oversight, or because the features were ambiguous, to make a correct "translation". The Eastern nominative singular and the Western locative of short *-a* stems were both in *-e*, and in a context where either case would make sense,⁽⁴¹⁾ a translator might be forgiven for failing to remember that, in the version from which he was making his translation, the *-e* form was a nominative, as a result of which he mistook it for a locative and left it untranslated.

We can also see that some of the dialect anomalies which occur happen as a result of hyper-forms, whereby editors or recensionists misunderstood forms which they found in the versions from which they were making their translations, and made a translation change which was in fact unnecessary. Such a thing could, for example, occur if they were making a translation from a dialect in which

the voicing of intervocalic vowels occurred, i.e. *-k-* and *-t-* became *-g-* and *-d-* respectively. If the exemplar contained the word *up-pāda*, then the translator did not know, in an ambiguous context, whether this was a formation from the root *pat-* or the root *pad-*. If he worked on a mechanical principle, then he would assume that, since the dialect changed *-t-* into *-d-*, *-d-* in the exemplar should be changed to *-t-* in his own dialect. He therefore wrote *uppāta* in his translation, instead of the form *uppāda* which the Sanskrit form of the word confirms is the correct form.

Similar translation problems could arise when translating from a dialect where intervocalic consonants were elided, i.e. *-k-*, *-g-*, *-t-* and *-d-* became *-y-*, and an original *-y-* remained unchanged. Once again, in an ambiguous context, a translator would not know how he should interpret a word containing *-y-*. If, for example, his exemplar contained the word *anindiya*, he had the choice of regarding this as a future passive participle <Sanskrit *anindya*, or a past participle <Sanskrit *anindita*, and so he had to decide whether to write *anindiya* or *anindita* in his translation.⁽⁴²⁾

The preference for the use of the various Prakrits, which had been stimulated by the use of those dialects by the founders of Buddhism and Jainism and their followers, and which had reached its culmination in their use for administrative purposes by Aśoka, began to wane after Aśoka's death, and the resurgence of Brahmanism led to a re-assertion of Sanskrit as the language of literature, administration and religion. The Prakrits continued to be used for several centuries, but their importance gradually diminished, as they were replaced for cultural purposes by Sanskrit, and as a medium of current usage by the vernacular languages

of the common people. The Prakrits remained in use only as the languages of the early texts of the non-brahmanical religions, and even there they were subjected to a great deal of influence from Sanskrit.

The early works of Buddhism had been written in various Prakrits, but the growing prestige of Sanskrit led to an attempt to re-write these texts in Sanskrit. The success of such a translation process varied from sect to sect, and we find a variety of Sanskritised Prakrits which are generally referred to as Buddhist Hybrid Sanskrit, although it must be made clear that there is no one form of that language. The writing down of the Pāli canon in the first century B.C. probably put an end to the wholesale restoration of Sanskrit forms into the language of the Theravādin canon, but it is likely that the Sanskritic features which had been introduced by that time included: the restoration of *-r-* into certain consonant groups, especially in the locative pronominal ending *-tra* and the group *br-*, probably starting with the word *brāhmaṇa*; and the restoration of *-tv-* in the absolutive ending ⁽⁴³⁾ *-tvā*.

It is probable that these Sanskritisms were introduced into the canonical language in a haphazard way at first, in accordance with the ability of individual *bhikkhus*, who made the appropriate changes as they were preaching, or as they were reciting or copying texts. It is probably, however, that in the same way in which the form and content of the canon began to be standardised, so too the extent of Sanskritisation also began to be fixed.

It is clear from the way in which the word *brāhmaṇa* always appears in that form, instead of the form *bamhaṇa* or *bambhaṇa*

which, as we know from inscriptions and other sources, would have been more appropriate to a dialect of Middle Indo-Aryan in the first century B.C., that the Pāli canon has been subjected to a great deal of editorial work, whereby the language has been made consistent. Nevertheless, there are indications that those responsible for such editorial work were not infallible. There is a small number of examples where the context does not make it clear that the ending *-ttā*, which the editors must have found in the texts they were “translating”, was an absolute. In one or two cases the ending had become *-tā*, for metrical reasons.⁽⁴⁴⁾ In these circumstances the editors either left *-tā* or *-ttā* unchanged, or translated it as a past tense form in *-ittha* or *-ittho* rather than as an absolute in *-tvā*.⁽⁴⁵⁾

In addition to this, we sometimes find that the editors back-formed into Sanskrit incorrectly, probably because they misunderstood the context. We find, for example, that they produced a word *atraja* meaning “son”. Such a word is not known from Sanskrit, but it appears very likely that this is a wrong Sanskritisation of a Middle Indo-Aryan word *attaja*, which is really derived from *ātmaja*. Similarly we find a word *vyap̄patha*, meaning “speech”.⁽⁴⁶⁾ This would appear to be derived from Sanskrit *vāk-patha*, but was not recognised as such by the recensionist. He accordingly took the form *vap̄-patha* which he had received, and “restored” *vy-* at the beginning of the word, because he knew that many words which begin with *v-* are in fact derived from Sanskrit words in *vy-*.

It seems that both the language and the contents of the Pāli texts were altered very little after the time when they were writ-

ten down. If changes had been introduced, we should have expected a growth in the number of words and phrases which showed the characteristics of either the Sinhalese Prakrit or the Dravidian language, but there are very few traces of either of these in the canon. Some of the examples which have been claimed as showing the influence of the Sinhalese Prakrit can probably⁽⁴⁷⁾ be explained in other ways.⁽⁴⁸⁾ Some of the nominative singular forms in *-e*, which are usually regarded as “Māgadhisms”, have been explained as traces of the Sinhalese Prakrit, which also had a nominative singular in *-e*. Some of the anomalous forms where *-d(h)-* appears in place of an expected *-j(h)-*, e.g. *addhābhavati*, have also been claimed as being influenced by the Sinhalese Prakrit, where the change of *-j->-d-* occurs. The same claim has been made for the Pāli form of king Prasenajit’s name, i.e. *Pasenadi*, and the word *udu-mana*, where *udu-* has been thought to stand for *uju-*.⁽⁴⁹⁾

The evidence seems to show that the change of *-j->-d-* in the Sinhalese Prakrit did not take place until the fourth century A.D.,⁽⁵⁰⁾ and we should therefore have to assume that these Sinhalese forms were introduced into the canon after that date during the course of the transmission by scribes who inadvertently introduced features of their own vernacular language in place of the canonical Pāli forms which they were copying out.

On the other hand, the evidence for the existence of Sinhalese and Dravidian forms in works which were composed in later times in Ceylon or South India is stronger, and the probability of this happening is much greater. Besides the Dravidian words which Pāli inherited from Sanskrit or some other dialect of Old Indo-

(51) Aryan, there are several words which appear for the first time in the post-canonical Pāli texts, e.g. *aṭṭa* “lawsuit”⁽⁵²⁾. There are also certain grammatical and syntactical features which seem not to have been inherited from Old Indo-Aryan, and a Dravidian origin has been postulated for these with a fair degree of probability,⁽⁵³⁾ e.g. the use of participial constructions in substitution for relative clauses which would normally be introduced by a relative pronoun or pronominal adverb, and also the employment of a new type of compound verb based on the use of particular action auxiliaries, e.g. the use of the verb *deti* with the absolutive of another verb.

In medieval times, there was considerable influence exerted upon Pāli by both Sinhalese and Sanskrit. A number of Pāli works dating from that period were translations from Sinhalese, and included borrowings from the Sinhalese language.⁽⁵⁴⁾ The borrowings from Sanskrit were not restricted to vocabulary, although such borrowings were numerous, since the lexicographical work *Abhidhānappadīpika* was based largely upon the Sanskrit *Amarakośa*, but there was also much imitation of various genres of literature. The works of the Sanskrit grammarians had a particularly strong influence upon writers of Pāli. Following the example of the Sanskrit grammarians, the Pāli grammarians began to classify and categorise the features of Pāli, and it seems clear that some of the Pāli commentators also had a knowledge of Sanskrit, which might well have had an affect upon the phonology and morphology of the texts they transmitted.

It was doubtless their knowledge of *sandhi* in Sanskrit which led to their writing a long *-ā-* before a doubled consonant as

the result of the crasis of two vowels in a compound, and also in such words as *yvāssa* (=yo+assa) and *tyāssa* (=te+assa) as products of a non-historical *sandhi* process, as well as *gavāssāca* (=gavā assā ca). These examples seem to be based upon a knowledge of Sanskrit, since they go against the general rules of Middle Indo-Aryan. A knowledge of Sanskrit seems also to underlie the writing of *agy-antarāya* (=agg' -antarāya <aggi+antarāya), on the basis of Sanskrit *agny-antarāya*, and *Khatyā* (=Khattā <Sanskrit *kṣatra*), possibly in the belief that it was a contraction of *khattiyā*. We might postulate that it was a knowledge of Sanskrit, and a predilection for Sanskritic formations, which led to the complete disappearance from the canon of the word *bārasa* "twelve". It must have been in the canon at one time, since it is quoted by the grammarians, but it has now been replaced everywhere by the Sanskritic form *dvādasā*.⁽⁵⁵⁾

We find, in a strange repetition of what had happened in the case of Sanskrit over a millennium and a half before, that the Pāli grammarians, in describing the language of the Pāli canon, had in fact written a normative study of the language, which served to guide scribes in succeeding centuries in the way in which the language should be written. There was consequently a tendency for the scribes to "correct" what they were copying, in the light of the grammarians' statements.

The effect of the Pāli grammatical works differed somewhat from that of Pāṇini's grammar. His work had halted the development of the Sanskrit language, which remained unchanged, frozen and fossilised, for the next two and a half millennia. The Pāli grammarians did not halt the development of the Pāli language.

That had already been done more than a millennium before, when the canon had been committed to writing. What they did was to define certain forms or spellings as correct, to the exclusion of others. The scribes, therefore, following their prescriptions, replaced some of the forms they found in the manuscripts they were copying by others favoured by the grammarians, even though the latter might be artificial creations, and inferior to those that were discarded.⁽⁵⁶⁾

Helmer Smith was correct when he stated⁽⁵⁷⁾ that the Pāli of the canon as we have it now is a reflection of the Pāli of the twelfth century, when the influence of the Pāli grammarians was at its highest.

Notes.

- (1) For a survey of the Indo-European languages see T. Burrow, *The Sanskrit Language*, London, 1955, pp. 7—9.
- (2) See T. Burrow, "The Proto-Indoaryans", *JRAS*, 1973, pp. 123—40.
- (3) See P. Thieme, "The 'Aryan' gods of the Mitanni treaties", *JAOS*, 80, 1960, pp. 301—17.
- (4) See Burrow, *op. cit.* (in n. 2), p. 140.
- (5) See Burrow, *op. cit.* (in n. 1), p. 31.
- (6) For words in Indo-Aryan languages for which a Dravidian origin seems likely, see T. Burrow and M.B. Emeneau. *A Dravidian Etymological Dictionary*, second edition, Oxford, 1984, pp. 759—68.
- (7) See Y.V. Knorozov, M.F. Albedil and B.Y. Volchok. *Proto-Indica: 1979*, Moscow 1981. and A. Parpola, "Tasks, methods and results in the study of the Indus script", *JRAS*, 1975, pp. 178—209.

- (8) See Burrow, *op. cit.* (in n. 2), p. 126.
- (9) For a discussion of the words in the R̥gveda which are possibly of Dravidian origin, see Franklin C. Southworth, "Lexical evidence for early contacts between Indo-Aryan and Dravidian", in *Aryan and Non-Aryan in India*, ed. Madhav M. Deshpande and Peter Edwin Hook, Ann Arbor, 1978, pp. 191—233 (pp. 194—95).
- (10) See Burrow, *op. cit.* (in n. 2), p. 125.
- (11) See F.B.J. Kuiper, *Proto-Munda words in Sanskrit*, Amsterdam 1948.
- (12) See Colin P. Masica, "Aryan and non-Aryan elements in North Indian agriculture", in Deshpande and Hook, *op. cit.* (in n. 9), pp. 55—151. He states that about one third of such items in Hindi seem not to be of Indo-Aryan, Dravidian, Munda or known foreign origin (p. 137).
- (13) See H. Bechert, "The date of the Buddha reconsidered", *IT*, X. 1982, pp. 29—35 and "A remark on the problem of the date of Mahāvira", *IT*, XI. 1983, pp. 287—90.
- (14) From Sanskrit *prākṛta*, derived from *prakṛti* "origin". i.e. "connected with, derived from, an origin", viz. Sanskrit.
- (15) See J. Bloch, "Prākṛit cia, latin quidem", *Language*, 29, 1953, pp. 229—30.
- (16) See K.R. Norman, "The dialectal variety of Middle Indo-Aryan", in W. Morgenroth (ed.), *Sanskrit and World Culture*, Berlin 1986. pp. 389—96.
- (17) See K.R. Norman, "Middle Indo-Aryan Studies V", in *Journal of the Oriental Institute* (Baroda), XV, pp. 113—17 (p. 114).
- (18) For a discussion of the dialects of Old Indo-Aryan, see M.B. Emeneau, "The dialects of Old Indo-Aryan", in Henrik Birnbaum and Jaan Puhvel (edd.), *Ancient Indo-European dialects*, Berkeley and Los Angeles, 1966, pp. 123—38.
- (19) See Burrow, *op. cit.* (in n. 1), pp. 47—48.
- (20) For a survey of the distinguishing features of the dialects of Middle Indo-Aryan, see J. Bloch, *Indo-Aryan from the Vedas to modern times*, Paris 1965. pp. 15—20.
- (21) See K.R. Norman. "The dialects in which the Buddha preached", in H. Bechert (ed), *The language of the earliest Buddhist tradition*, Göttingen, 1980, pp. 61—77 (pp. 66—67).

- (22) See E. Hultzsch. *Inscriptions of Asoka*, Oxford 1925. pp lxxxvi-lxxxvii.
- (23) See Hultzsch, *op. cit.* (in n. 22), p. xcvi.
- (24) See K.R. Norman. "Lexical variation in the Aśokan rock edicts". *TPS* 1970, pp. 121—36 (p. 134).
- (25) See L. Alsdorf, "Contributions to the study of Aśoka's inscriptions". *BDCRI*, XX. pp. 249—75 (p. 251).
- (26) See D.C. Sircar, "Erragudi edicts of Asoka", *Ep. Ind.*, XXXII, 1957—58, pp. 1—28 (p. 5).
- (27) See Th. Damsteegt, *Epigraphical Hybrid Sanskrit*, Leiden 1978.
- (28) See K.R. Norman, review of Damsteegt, *op. cit.* (in n. 27), *Lingua*, 48, 1979, pp. 291—94 (p. 293).
- (29) See T. Bloch, "Caves and inscriptions in Rāmgarh Hill: Jogimārā inscription", *Archaeological Survey of India Annual Report*, 1903—4, pp. 128—31.
- (30) See L. Alsdorf. "Ardha-Māgadhī", in Bechert, *op. cit.* (in n. 21), pp. 17—23.
- (31) See Norman, *op. cit.* (in n. 21), pp. 67—68.
- (32) See Norman, *op. cit.* (in n. 21), p. 70.
- (33) In the Majjhima-nikāya (Pali Text Society edition, Vol. III, pp. 234—35) the Buddha lists seven words all meaning "bowl", which he says are *janapadanirutti* "country dialect".
- (34) See K.R. Norman, *Pali Literature*, Wiesbaden, 1983, pp. 7-14.
- (35) For references to *bhāṇakas* in the Pāli commentaries, see E. W. Adikaram, *Early History of Buddhism in Ceylon*, Migoda, Ceylon, 1946, pp. 24—32.
- (36) For details see the Vinaya-piṭaka (Pali Text Society edition), Vol. I, pp. 140—41.
- (37) For evidence that texts were already being written down before the first century B.C., see J. Brough, *The Gāndhārī Dharmapada*, London 1962, p. 218.
- (38) Samantapāsādikā (Pali Text Society edition), pp. 695—96.
- (39) For the way in which the compound *pāli-bhāsā* was understood to mean, not "language of the canon", but "Pāli language", see K.R. Norman, "Pāli literature and language", §12. 1—3, to appear in the proceedings of a symposium on the Buddhist Heritage, held at the School

- of Oriental and African Studies, London, 28—30 November 1985.
- (40) See K.R. Norman, “Pāli and the language of the heretics”. *Acta Orientalia*, 37, 1976, pp. 117—26.
- (41) See the discussion on Theragāthā 1229 (*sacce atthe ca dhamme ca āhu santo patitṭhitā*) in K.R. Norman, *Elders’ Verses I*, London 1969, p. 292.
- (42) See K.R. Norman. “Dialect forms in Pāli”, to appear in the proceedings of a colloquium on dialects and dialect forms in the Indo-Aryan languages, held in Paris, 16—18 September 1986.
- (43) See Norman. *op. cit.* (in n. 39). § 6. 3.
- (44) See K.R. Norman, “Four etymologies from the Sabhiya-sutta”, in Somaratna Balasooriya et al. (edd.), *Buddhist Studies in Honour of Walpola Rahula*, London 1980, pp. 173—84 (p. 183 n. 21) and “Pāli etymological studies III”, *JPTS*, 1985, pp. 23—36 (pp. 32—35).
- (45) See O. von Hinüber, “Pāli as an artificial language”, *IT*, X, 1982, 133—40 (pp. 136—37).
- (46) See K.R. Norman, “Two Pāli etymologies”, *BSOAS*, XLII, 1979, pp. 321—28 (p. 326).
- (47) See H. Bechert, “Über Singhalesisches im Pālikanon”, *WZKS-uO*, I, 1951. pp. 71—75.
- (48) See Norman, *op. cit.* (in n. 40), pp. 119—21.
- (49) See O. von Hinüber, *Das ältere Mittelindisch im Überblick*, Wien 1986, § 248.
- (50) See W. Geiger, *Grammar of the Sinhalese language*, Colombo 1938, § 46.
- (51) See M. D’Onza Chiodo and E. Panattoni, “Dravidian contribution to an interpretation of Pāli *gāmapoddava* and Buddhaghosa’s explanation of the term”, *IT*, XII, 1984, 355—63. If the claim for a Dravidian origin of the word *poddava* is correct, then since it occurs in the Vinaya-piṭaka (Pali Text Society edition, Vol. II, p. 105), but does not occur in Sanskrit, it is likely that it was taken over into Middle Indo-Aryan in North India, before the advent of Buddhism to Ceylon.
- (52) See M. D’Onza Chiodo and E. Panattoni, “Pāli *aṭṭa*: ipotesi di un’ influenza dravidica su una controversa etimologia”, *IT*, V, 69—84.
- (53) See K. de Vreese, “Dravidian idioms in later Pāli”, in *Orientalia Lovaniensia Periodica*, XI. pp. 179—222.

The origin of Pāli and its position among the Indo-European languages. 27

- (54) See A.P. Buddhadatta, *Sahassavatthu Pakarana*, Maradana 1959, p. xxv.
- (55) As O. von Hinüber points out, in *Notes on the Pāli tradition in Burma*, NAWG 1983. 3, Göttingen 1983, pp. 74—75 [12—13].
- (56) See von Hinüber, *op. cit.* (in n. 55). p. 73 [11].
- (57) H. Smith *Saddaniti*, Vol. I, Lund, 1928. p. vi.