

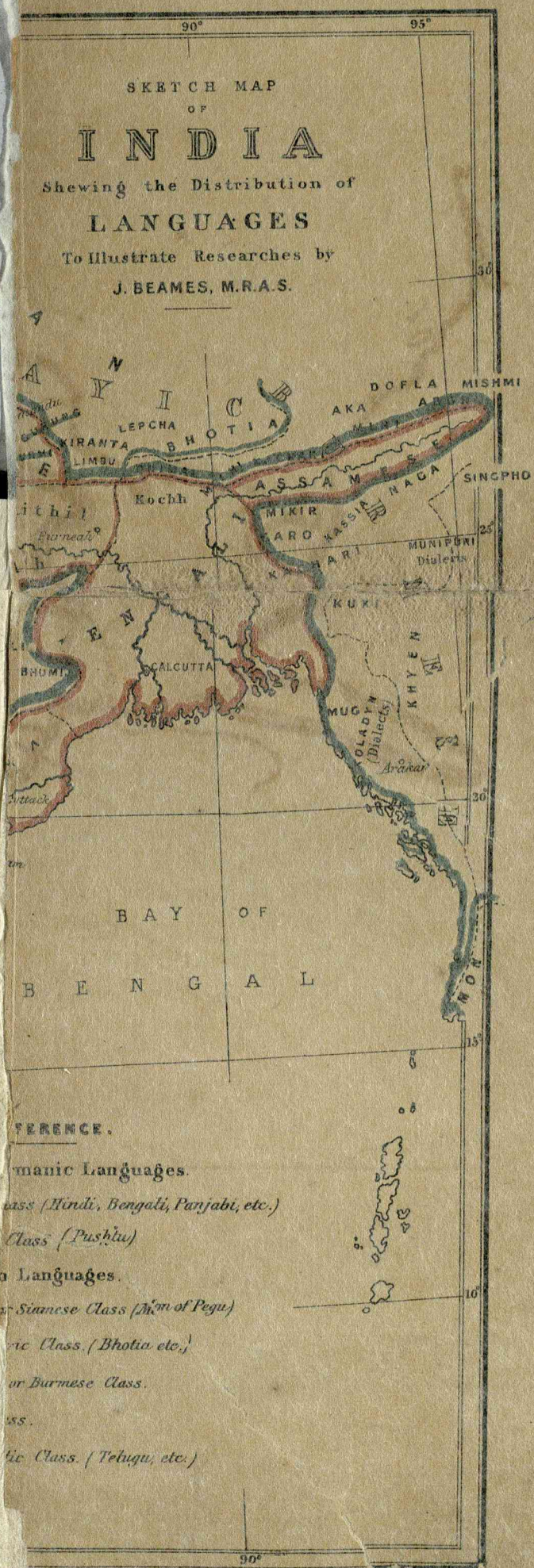


CSL

SKETCH MAP OF INDIA

Shewing the Distribution of LANGUAGES

To illustrate Researches by
J. BEAMES, M.R.A.S.



AS-001822

CSL



FOR CONSULTATION ONLY

OUTLINES

4

OF

INDIAN PHILOLOGY,

WITH

A MAP SHEWING THE DISTRIBUTION OF INDIAN
LANGUAGES.

18

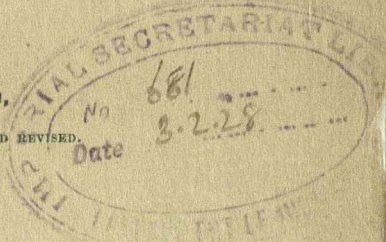
BY

JOHN BEAMES, M.R.A.S.,

BENGAL CIVIL SERVICE, MEMBER OF THE ASIATIC SOCIETY OF BENGAL, THE
PHILOLOGICAL SOCIETY OF LONDON, AND THE SOCIÉTÉ
ASIATIQUE DE PARIS.

SECOND EDITION,

CONSIDERABLY ENLARGED AND REVISED.



LONDON:

TRÜBNER AND CO., 60, PATERNOSTER ROW.

1868.

(All rights reserved.)



CSL

Calcey of Fat William

491.4

Bea M080 ~~B360~~



4914

Bea
M68.0

10792

1613

11/8/13.



5CSL

PREFACE.

THE first edition of this little work, though meeting with a favourable reception, has been said by competent critics to be too meagre. I have, therefore, entirely re-written many parts, enlarged others, and added much new matter. I trust that this second edition will be found in its amended form more generally useful. - I have carefully avoided inserting matter which would be inconsistent with its character as a purely elementary work. In a few cases I have added references to the authorities from whom I derive my statements, but as the work is professedly a compilation, it would be out of place to quote all my authorities. I have also had the map re-engraved, and have written many words in their appropriate Indian type.

CALCUTTA,

February 24th, 1868.



CSL

PREFACE TO THE FIRST EDITION.

THE following pages are a compilation from the best and most accessible books on the science of language, supplemented by facts derived from personal observation. They do not pretend to be anything more than an outline for the use of those who, having no knowledge of Linguistic Science, wish to record and preserve dialects of obscure and uncivilized tribes with whom they may come into contact, or any of the countless local peculiarities of the leading Indian languages which may be spoken in their neighbourhood. Unscientific persons can often render great service to Science by simply recording faithfully and intelligently facts which come under their notice.

It is hoped that this little sketch may be of use to such persons, and its very incompleteness and omissions may lead to a desire to become better acquainted with the useful and fascinating study of the science of human speech.

CHUMPARUN,

*August 25th, 1866.*¹

¹ The first edition was not published till August, 1867, owing to the distance of my station from the nearest printer.



CONTENTS.

	PAGE
CHAPTER I.—Classification of Languages	1
CHAPTER II.—Distribution of Indian Languages	9
CHAPTER III.—Stages of Development of Languages	22
CHAPTER IV.—Family Characteristics of Languages	36
CHAPTER V.—Dialects	52
CHAPTER VI.—Hints on observing and recording a new Dialect	65
APPENDIX A.—List of Numerals in nearly all the Indian Languages	76
APPENDIX B.—Personal Pronouns in the leading Turanian Dialects	90
APPENDIX C.—Abstract of official information concerning the Languages of Assam	91

Map, at end of the book.



NOTE.—In these pages the spelling of Indian words follows, as a rule, the system of Professor Wilson ; but I have departed freely and frequently from that system in the case of names of places, which it would be mere pedantry to write according to a philosophical system, now that the old spelling has become so well known and is so universally used.¹

N.B.—In this second edition I have, in deference to an opinion expressed in the *Athenæum* (No. 2097, Jan. 4th, 1868, p. 19), altered the spelling of the word Marathi, which I formerly wrote Mahrathi. Though the former is more usual, yet the latter is found in many works, and may be defended on the ground of the original name being Mâhârashtri.

¹ I have adhered to the common-sense principle laid down in the following plain, though uncouthly worded quotation :—“ In writing Marâth names of places in the Roman character, names of rare occurrence among Europeans are, of course, spelled as spelled in the original—by Roman symbols equivalent to the Marathi symbols ; but names of common occurrence and familiar appearance in Roman apparel are, equally of course, left intact ; for Bangâla, Kalikatta, Shrirâmpûr, Chandranagar in place of Bengal, Calcutta, Serampore, and Chandernagore, would be not the Roman guise, but a Roman disguise, not an attire, but a metamorphosis.” (Preface to Molesworth’s Marathi Dict. p. xv.)



CSL

ERRATA.

- Page 7, note, *for Bhai read Bghai.*
" " " *for Piro read Pwo.*
" " line 19, *for Tudu read Toda.*
" " line 20, *for Budugur read Badaga.*
" 8, line 23, *for Tuluvu read Tuluva.*
" 18, last line, *Tuluvu read Tuluva.*
" 32, last line, *insert antre after tout.*
" 33, line 3, *for as read sa.*
" 45, note, *for Sindi read Sindhi.*
" 61, last line but one, *for section read selection.*
" 73, line 10, *for tribes read tribe.*
" 77, Marathi line 6, *for षष् shash, read सहा sahá.*
" 95, line 6, *from "The proportionate," etc., down to "districts,"*
page 96, line 17, *is a quotation, and should be marked by inverted*
commas.

ERRATA IN MAP.

- For Tuluvu read Tuluva.*
For Budugur read Badaga.
For Tudu read Toda.



OUTLINES OF INDIAN PHILOLOGY.

CHAPTER I.

CLASSIFICATION OF LANGUAGES.

THE languages of Europe and Asia are divided into three great families:

1. Indo-Germanic, | 2. Semitic.
3. Turanian.

1. The Indo-Germanic family includes the following classes:—

- | | | |
|-------------|--|--------------|
| 1. Indic. | | 5. Teutonic. |
| 2. Iranian. | | 6. Slavonic. |
| 3. Celtic. | | 7. Hellenic. |
| 4. Italic. | | 8. Illyric. |

Of these only the first two classes are found in India.

2. The Semitic family comprises—

1. Arabic. | 2. Hebrew.
3. Aramaic.

Nothing more need be said about this family, as no Semitic languages are spoken in India.¹

¹ The Arabic element, which so largely enters into the spoken dialects of India, is no exception to this rule, because the Arabic words so used are not inflected or conjugated according to strict Semitic canons.



3. The Turanian family is divided into two branches, the Southern and the Northern. It is only with the southern branch that the present work has any concern. It includes the following classes :

- | | | |
|---------------|--|-------------|
| 1. Thaïc. | | 3. Lohitic. |
| 2. Himalayic. | | 4. Kol. |
| 5. Dravidic. | | |

I. Indo-Germanic family.

1. Indic class.

The earliest representative of this class is the language of the Vedas, the most ancient recorded form of Sanskrit. Next in point of time we have the classical Sanskrit, contemporary with the later works of which is the Prakrit, or vulgar dialect probably then spoken by the common people of India, which is nothing more than a corruption of Sanskrit. It has several dialects, named after various provinces, though it is doubtful whether each dialect was really spoken in the province from which it took its name. One of these, the Mágadhi or speech of Magadha, the modern South Behar, was the mother-tongue of the great reformer Buddha, and as such became the sacred language of countries where the Buddhist religion prevailed, especially Ceylon, where it was transported by Vijaya in B.C. 543, and is called Pali. It possesses a considerable literature. Similarly the Sauraseni dialect of Prakrit, spoken in the neighbourhood of Delhi and Agra, has



become the sacred language of the sect of Jains, who are principally Marwáris. The languages of this class at the present day are the following :

- | | |
|-------------|---------------|
| 1. Hindi. | 6. Gujarati. |
| 2. Bengali. | 7. Nepalese. |
| 3. Panjabi, | 8. Uriya. |
| 4. Sindhi. | 9. Assamese. |
| 5. Marathi. | 10. Kashmiri. |

11. Doghra.

The dialects of Hindi are very numerous. The chief are :

- | | |
|--|--|
| (a.) Maithil; spoken in Purneah and Tirhut. | (e.) Brijbhasha; Upper Doab, Agra, and Delhi. |
| (b.) Magadh; in South Behar. | (f.) Kanauji; Lower Doab. |
| (c.) Bhojpuri; in Shahabad, Sarun, Champáran, Goruckpore, Eastern Oudh, and Benares. | (g.) Rajput dialects, very numerous; Rajputana. |
| (d.) Kosali; in Oudh and Rohilkhand. | (h.) Bundelkhand dialect; from the Chambal to the Soane. |

Panjabi has many dialects. In fact, in the Panjab every district has its own dialect, and some districts have more than one.

Sindhi is divided into—

- | | |
|--------------------------------|---|
| (a.) Sirai of Upper Sindh. | (d.) Uch of Multan (transitional to Panjabi). |
| (b.) Vicholi of Central ditto. | (e.) Kachi of Kach (transitional to Gujarati). ¹ |
| (c.) Lari of Lower ditto. | |

¹ Of these the Sirai is considered the purest. Dr. Trumpp in *Zeitschrift d. Morgenl. Gesellsch.* vol. xv. p. 692.



CSL

OUTLINES OF INDIAN PHILOLOGY.

Marathi has four dialects—

- | | | |
|--|--|--|
| (a.) Konkani; spoken in Ratnagiri and along the sea-coast. | | Sawant Wari. This dialect is locally known by the name of Kudali. ¹ |
| (b.) Dakhini. | | (d.) Khaudesi. |
| (c.) Gomantaki; spoken near | | |

Gujarati has three dialects, spoken respectively in

- | | | |
|------------------------|--|-----------------|
| (a.) Surat and Broach. | | (b.) Ahmedabad. |
| | | (c.) Kattiwar. |

Nepalese pure is called Parbattia or Paharia;² slightly different are the dialects of

- | | | |
|--------------|--|---------------|
| (a.) Palpa. | | (c.) Garhwal. |
| (b.) Kamaon. | | (d.) Tharú. |

2. Iranian class.

The parent of the languages of this class is the so-called Zend or early Persian. It is closely akin to Sanskrit. The name Zend, though consecrated by usage, is not really the name of the language, but is derived from the celebrated collection of fragments of old Persian religious works known as the Zend Avesta. This work is written in verses. Each verse contains two parts: the Avesta or text, and the Zend or commentary. The knowledge of the language of the text dying out, the commentary became the better known of the two parts, and eventually gave the name not only to the work itself, but to the language in which

¹ This dialect is gradually dying out. The native Roman Catholic Christians, who are rather numerous round Sawant Wari and Goa, speak a patois which contains many Portuguese words mixed with Marathi.

² Also Khas.



it is written, of which it is the sole specimen. Portions of the work called *gáthás* are supposed to have been written by Zarathustra Spitama or Zoroaster himself.¹ It stands in a position analogous to Sanskrit in the Indic class. Pehlevi, Hazvareh, and the languages intermediate between Zend and modern Persian occupy a similar position to Prakrit and Pali. The modern languages are:

- | | |
|--------------|----------------|
| 1. Persian. | 3. Pushtu. |
| 2. Kurdish. | 4. Ossitinian. |
| 5. Armenian. | |

It is unnecessary to go into details concerning these languages which are beyond the limits of India. Nor is it, for the same reason, necessary to give the languages of the remaining classes of this family. We pass on to the Turanian family.

III. Turanian family, southern branch.

Class 1. Thaïc or Siamese. Contains the following languages:

- | | |
|--|--|
| 1. Siamese or Thaï; spoken in Siam. | 4. Khanti; spoken in Burmah |
| 2. Khó or Kambojan; spoken in Kamboja. | 5. Môn „ Pegu. |
| 3. Laos; spoken in Central Siam. | 6. Shan „ North-East of British Burmah. ² |
| | 7. Palaong „ Northern Burmah. |

¹ This is as accurate a statement as I can gather from the various authorities. I refer the reader for fuller particulars to Haug's *Essays on the Sacred Language of the Parsees* (Bombay, 1862).

² The southern Shans are said to be dependant on the Siamese; the rest are to a certain extent free.



OUTLINES OF INDIAN PHILOLOGY.

Nearly all these languages lie beyond the limits of British settlements or British influence.

Class 2. Himalayic. (Sub-Himalayan of Prof. Max Müller.)¹

- | | |
|---|----------------------------|
| 1. Bhotia or Bhotanta, | 13. Sunwar, western Nepal. |
| 2. Lepcha (Sikkim). | 14. Sarpa, ditto. |
| 3. Limbû } basins of the Sun | 15. Kanáwari or Milchán. |
| 4. Kiranti } Kosi & Arun rivers, | 16. Tibarskad } beyond the |
| Eastern Nepal. | 17. Hundesi } snowy range |
| 5. Murmi, eastern Nepal, | north of Kamaon. The |
| higher ranges. | country of the Hundesis |
| 6. Gurung, ¹ ditto. | is called Narikhursam. |
| 7. Newar, central Nepal. | 18. Darahi or Dorhi |
| 8. Magar, ² lower ranges, cen- | 19. Denwar |
| tral Nepal. | 20. Pahri |
| 9. Bramhu, lower ranges, cen- | 21. Kaswár |
| tral Nepal. | 22. Pakhya |
| 10. Chepang } Oudh Terai. | 23. Thaksya |
| 11. Vayu(Hayu) } (The Hayus | |
| 12. Kusunda } are also found | |
| } in eastern | |
| } Nepal). | |

} Central Nepal.

The above languages form the Himalayan proper or sub-Himalayan class. The trans-Himalayan or Tibetan

¹ I do not know why Prof. Müller calls this class sub-Himalayan, seeing that its home is in the very core and centre of the Himalayas, while the Tibetan, which he calls Himalayan (without the qualifying prefix), is in reality spoken chiefly beyond and outside of the Himalayan chain.

² Dr. Campbell places these two among Hindi dialects. His reason, however, is not a good one. It is that these tribes are Brahmanical in religion. Their language, however, of which I have had personal experience, is undoubtedly of Tibetan origin, though adulterated with words from the Parbattia and Hindi.



proper, it is beyond our province to notice. At the same time, it must be stated that all these languages are dialects of Tibetan, or, at least closely allied to it; and Bhotia is identical with Tibetan.

Lohitic or Burmese class; contains—

- | | |
|--|--|
| 1. Burmese. | 14. Singpho; southern frontier of Assam. |
| 2. Dhimal—Nepal and Bhotan Terai. | 15. Kuki; North of Chittagong; Tipperah, etc. |
| 3. Mechi, ditto. | 16. Mag; Arakan. |
| 4. Borro; Cachar (same as 2). | 17. Khumia „ |
| 5. Gáro; Gáro Hills. | 18. Mru „ |
| 6. Aka, northern frontier of Assam. | 19. Sak „ |
| 7. Abor „ | 20. Tunghlu „ |
| 8. Mishmi „ | 21. Rukheng „ |
| 9. Miri „ | 22. Koladyn river dialects (said to be very numerous). |
| 10. Dofla „ | 23. Kakhyen. |
| 11. Kassia (Cossya), southern frontier of Assam. | 24. Manipuri dialects. |
| 12. Mikir „ | 25. Koreng dialects. |
| 13. Naga { | 26. Karen dialects. ¹ |
| | Rengmá „ |
| | Angámi „ |
| Lotah „ | |

“The Caucasus itself, with all its accumulation of mutually unintelligible forms of speech, within a com-

¹ The Sgäu is the principal dialect of the Karen. The accent is on the last syllable in the word Karén. The various Karen dialects are—

- | | |
|------------|----------------|
| Sgäu. | Taru. |
| Bhai. | Mopgha. |
| Red Karen. | Kay or Gaykho. |
| Piro. | Toungthu. |

(From an article by Dr. Mason in Journ. As. Soc. Bengal.)



paratively small area, is less remarkable for the density of its languages than the parts now under notice. Whether we look to the Gáro, Kassia, and Mikir areas themselves, or to the parts which immediately underlie them, *viz.*, Cachar, Sylhet, Tipperah, and Chittagong ; whether we look to the Naga districts of Assam, and the parts which lie due south of them, or the valley of the Upper Irawaddy and its feeders,—we find an accumulation of actual languages, or possible dialects, such as we rarely find in the old world elsewhere.”—Latham, “Elements of Comparative Philology,” p. 36.

4. The Kól class contains—

- | | |
|---------------------------|-------------------------------|
| 1. Sonthâl. | 5. Kolehan or Hô. |
| 2. Kól of Chyebassa. | 6. Khond of Sambhalpoor, etc. |
| 3. Bhumij of Purulia. | 7. Gond. |
| 4. Mandali, Chota Nagpur. | 8. Uraon of Sirgulah. |
| | 9. Rajmahali. |

5. The Drâvidian class comprises—

- | | |
|-------------------|---------------------------------------|
| 1. Telugu. | 7. Tudu, Nilgiri Hills. |
| 2. Tamil. | 8. Budugur „ |
| 3. Karnátaka. | 9. Irular „ |
| 4. Malayálam. | 10. Kohatar „ |
| 5. Tuluva. | 11. Brahui, Biluchistan. |
| 6. Kodugu, Coorg. | 12. Singhalese ¹ , Ceylon. |

¹ The large proportion of words of Sanskrit origin in the classical Singhalese, or Elu, has led Max Müller to class it with Aryan dialects. I believe it owes its Aryan words to Pali influence. “There are three elements in Singhalese, one in connection with the Sanskrit, a second with the Pali, a third with local elements.”—D’Alwis, in *Journal of Ceylon Br. R. A. S.*, 1865-66, p. 143.



CSL

CHAPTER II.

ON THE DISTRIBUTION OF INDIAN LANGUAGES.

IN the annexed map the languages of the Indo-Germanic family are shown in red, those of the Turanian family in blue, and the area of each language is approximately marked. In the Himalayas the two families, as far as we have data for them, are so intermixed, that it is difficult, if not impossible, to lay down definite boundaries. Especially is this the case in the large kingdom of Nepal, which is still a *terra incognita* in many respects.

It will be seen that the tract coloured red divides the blue into two parts, between which it comes like a wedge from west to east, breaking through at Rajmahal, and spreading out eastwards again into a wide area, till it meets the blue line beyond Chittagong.

Historically, there is little doubt that at an early period the whole of India, in common with all eastern and southern Asia, was held by races speaking languages of the Turanian family. The Aryan race, speaking a language of the Indo-Germanic family, entered India from the north-west, and gradually worked its way down the valley of the Ganges, driving the Turanians



into the then almost impenetrable forests and hills of the south. The tide of expulsion was chiefly southwards. Unconquered Turanian races already held the mountain fastnesses of the Himalayas and the deadly forests of the Terai; whereas the great "Dakshinaranya," or southern forest, was, as far as we know, uninhabited; and consequently the migration of the flying Turanians was less restricted in that direction. It is necessary, however, to notice in passing that there was some little displacement of Turanians northwards to the hills, to account for the fact that some of the traditions of Northern Behar mention tribes (such as the Kichak or Kiratas, for instance) inhabiting the plains in former times who are now only found in Nepal. Still the mass of ancient legendary poems describes the expelled nations as flying southwards; and the modern Turanian races in the Himalayas either came originally from "across the snows," *i.e.*, from Tibet, or round the end of the Himalayas by the valley of the Brahmaputra; chiefly the former.

In spite of expulsions and oppressions, however, there is reason to believe that a considerable number of Turanians remained still in the Valley of the Ganges; and it is to the obstinacy with which they retained certain characteristics of their original speech, that we must ascribe some of the peculiarities of modern north-Indian languages.¹

¹ I allude here chiefly to the Hindi post-positions, especially "Ko," and



At a very much later period the Aryan colonies penetrated the hills of Nepal and the western and central Himalayas, but did not entirely drive out the Turanian populations. Hence the perplexing mixture of dialects which we find in the Himalayas.

In the south, again, the Turanians continue to hold the low ranges of the Rajmahal and Kymore hills, and the wild country which stretches south-east to Orissa and south-west to the Nerbudda. Those Turanian tribes who penetrated into the extreme south were afterwards civilized by Brahmins from Aryan India, but those who lived in the hill ranges remained in their original savage state. This explains why in the Dravidian class of languages we find so many words of Sanskrit origin; and why the Tamil, Telugu, and Canarese peoples are in possession of a civilization so far superior to that of the Kols and Gonds.

To the east, the hill ranges which divide Assam from Sylhet, and the Tipperah and Chittagong ranges, mark the utmost limits of living Aryan extension. I say living, because in the fossilized form of Pali, an Aryan language was carried by Buddhism into the Siamese Peninsula, Java,¹ and the Asiatic Archipelago.

to the method of forming the plural by addition of "*log*"—" *sab*," etc. I do not consider the cerebral or lingual letters as of Turanian origin. They arise chiefly from a suppressed *r*. The initial *ज्ञ* *jh* also, which is by some ascribed to non-Aryan influences, is generally traceable to the Sanskrit *घ* *dhj*, or *ध* *dhm*.

¹ In the Javanese of the present day there are many pure Sanskrit words; and the old poetical dialect or "Kavi" speech is almost wholly Aryan.



The Mahomedan invasions of India did not alter the areas occupied by the two great families above-mentioned. The invaders were a very mixed multitude, consisting of Arabs, Persians, Afghans, Turks of the Chagatai, Uzbek, and other tribes, and Mongolians generally. The only results, as far as language is concerned, of their arrival in India, were the creation of the Urdu or Hindustani, and the introduction of a mass of Arabic words which have established themselves in almost every Indian language, though their influence is more perceptible in those of the Indo-Germanic family than in those of the Turanian.

A brief review of the languages at present spoken in India, describing their geographical limits, will illustrate this part of the subject more fully.

Beginning at the north-west angle of India, we find in the Peshawur and Hazara valleys, and in the district east of the Indus, called Chach Hazara, the PUSHTU language spoken with various local modifications, by a population of about 900,000.

In the mountain ranges between the Panjab and the Valley of Kashmir, the DOGHRA dialect or dialects are spoken, and in the Valley itself KASHMIRI. The population of the Doghra districts is apparently (for the census returns are not very trustworthy) about 400,000: that of Kashmir, in round numbers, three millions.

From the Indus on the west to the Sutlej on the east, and from the mountains to the neighbourhood of



Multan, we find PANJABI. This language is scarcely spoken alike in any two towns. The purest Panjabi is spoken between the Rivers Ravi and Beás, and generally the further south you go the wilder and more remote from the Hindi standard, and nearer to the Sindhi, becomes the speech. Panjabi is really nothing more than a dialect of Hindi, and is probably descended from the Sauraseni-Prákrit, but by virtue of having a different alphabet, it has come to be considered a separate language. It is spoken by about sixteen millions of people.

South and east of the Sutlej, Panjabi fades away imperceptibly into Hindi. The exact boundary cannot be fixed. On the banks of the Sutlej you are among Panjabis; travelling eastwards to the banks of the Jumna you find yourself among Hindustanis.¹

The HINDI covers a greater area than any other Indian dialect. The western boundary may be placed about Sirhind ($76^{\circ} 30'$ long., $30^{\circ} 45'$ lat.), and goes side by side with Panjabi south-westwards through the deserts of Patiala and Bhawalpur, till it meets Sindhi near Jysulmere. It then turns westwards through Udaypur, where it is conterminous with Gujarati and Marathi. The Hindi of these parts is much mixed with Sindhi and Gujarati. About Indore the three

¹ The natives generally fix the boundary of their language and country at the town of Sirhind, which they say is so called from being the sar-i-hind سرہند "head or beginning of India." Sirhind is now a half-ruined cluster of huts of no importance whatever.



CSL

languages meet. From this point the Vindhya and Sâtpura ranges bound it to the south as far as the Sone, which it follows northwards to Sirguja, thence skirting the Sonthal and Rajmahal hills to the Ganges, which it crosses at or near 87° 45' long., and goes in a line due north to the hills. These boundaries are of course approximate only. Except where natural barriers, as mountains or seas, occur, we nowhere find one language leaving off and another beginning at a given spot. Thus the Hindi of Purneah, as you go eastwards, gets more and more tinged with Bengali, till at last you reach a point where there is no Hindi at all traceable in it; but it is impossible to point out the exact spot where you cease to hear Hindi or begin to hear Bengali. Similarly, in Central India, it would be difficult to fix upon a point where the language ceased to be Hindi and began to be Marathi, or Gond, or Kole.¹ The Hindi speaking population is stated in the latest official returns to amount to 60,763,779.²

¹ It would be useful if persons living on frontiers between two languages would collect and publish facts tending to fix the exact limits of each. The few books at present accessible are in many instances, I am convinced, wrong on this point.

² The official census is as follows:—

Behar	{ Patna Division	8,282,559	
	{ Bhagulpore Division ...	3,254,538	
North-Western Provinces.....		30,007,871	
Oudh.....		8,000,000	conjectural.
Delhi and Hissar Divisions		3,218,811	
Central India		3,000,000	conjectural.
Rajputana		5,000,000	conjectural.

60,763,779

The language is spoken over an area of 248,000 square miles. The area



This does not include the large numbers of Mahomedans in other parts of India who speak Hindustani or Urdu. The Hindi-Hindustani is by far the most widely spread and commonly understood of all Indian languages, and is spoken as a *lingua franca* by people whose mother-tongue it is not, all over India.

BENGALI begins where Hindi leaves off, somewhere between Purneah and Dinajpore, and is bounded on the north by the Himalayan range as far as Assam, where it blends into Assamese. Thence turning south it is bounded by the hills which lie east of the Brahmaputra, till it fades away about Chittagong. In this last place the dialect is a curious mixture of bad Hindi and corrupt Bengali. The western boundary of Bengali is formed by the Rajmahal hills, and runs southwards through Bancoorah and Midnapore to the Subanrikha, which it follows to the sea. Population from official returns 20,583,635.¹

ASSAMESE, which is closely akin to Bengali, is spoken along the valley of the Brahmaputra from Gwalpara to Sadiya. There are no data for the population of Assam in any of the records of the Government of India.

URIYA extends along the seacoast from the Subanrikha to near Ganjam; landwards its boundary is un-

of France is 202,125 square miles and that of Austria 255,000 square miles, with populations respectively of 35 millions and 38 millions.

¹ This includes, however, the population of Calcutta, where a large proportion of the people also use Urdu, as well as English, Arabic, Armenian, Chinese, and all the languages under the sun.



CSL

OUTLINES OF INDIAN PHILOLOGY.

certain, it melts gradually into the Khond and other rude hill dialects and co-exists with them. In Bastár and the neighbourhood, some classes speak Uriya and some Khond. I am informed that Hindustani and Bengali are much used as a means of intercourse between different classes. If this be true, it is only another instance of the strong tendency of Hindustani to supply the place of *lingua franca* in all parts of India. This language is spoken by nearly two millions of people.

The whole of that net-work of low hills, whose northernmost point abuts upon the Ganges at Sikri-galli, while its southern ranges look down upon Nag-pore, the part of India least known to Europeans, is inhabited by the wild Koles, Gonds, and other Turanian tribes. The nine languages of Class 4 of the Turanian family occupy this region. The distribution of them cannot be given with any approach to accuracy. Its more accessible portions are encroached on from the east by Bengali, the west by Hindi, and the south by Telugu. The languages themselves are daily losing ground, and receding before the more civilized forms of speech which hem them in. I need hardly say that no reliable estimate has ever been made of the population of these wild hills, far less any regular census.

The Gond, one of these languages, passes near Nag-pore into MARATHI, which is in point of area and general importance second only to Hindi among Indian lan-



guages. From Nágpur it goes westwards through Sioni and Baitúl to the Sátpurá range, which it follows to the Narmadá (better known as the Nerbudda river). Thence it is bounded to the north by Bhíl tribes speaking a Kol dialect as far as the little river Damán Gangá, which it follows to the sea at the Portuguese settlement of Damán. The western boundary is formed by the Indian Ocean as far as the Portuguese settlement of Goa, where it turns inland through Kolhapur¹ till it touches the Krishna river, which is for some little distance the boundary between it and Canarese, then in a north-easterly direction through Bijápur, Sholápur, and Náldrug to the Godávarí, and so to Mahur, where it is conterminous with Telugu along the Páin Ganga river. Then it turns east and goes as far as Chanda, and on to Wairágar, and so to Nagpur. The boundary line is nearly everywhere irregular and ill-defined. The population is about ten millions.²

GUJARATI is bounded by Marathi on east and south, on the north it merges into Hindi in Marwar and Udaypur. To the west it meets Kachi and Sindhi about 70° 15'. Population about six millions.

KACHI occupies the isolated peninsula of Kach.

¹ In Kolhapur are several classes who use a patois of their own. They are called Ramusis, Mangs, Bagwans, etc. See Report on Kolhapur by Major Graham in Selections from Records of the Bombay Government, No. 8, new series, where several vocabularies are given.

² See Preface to second edition of Molesworth's Marathi Dictionary, p. xxiv.



SINDHI is spoken in the valley of the lower Indus from Multan to the sea ; on the east it merges into the Rajputana dialects of Hindi, and on the west into the Beluch dialects. Sindhi is spoken by somewhat less than two millions. The exact figures are 1,795,594, but these do not include the wild pastoral tribes of the desert.

About Ganjam, on the East Coast, TELUGU is first heard, and it prevails all down the coast to Pulicat, close above Madras. Its northern boundary is conterminous with Uriya, the Kole dialects and Marathi. West of Haidarabad it meets the cognate Canarese, and goes thence to the eastern boundary of Mysore, whence it is conterminous with Tamil as far as Madras.

The TAMIL area is bounded on the north by Telugu, and extends south to Cape Comorin, and along a small part of the western coast to Trivandram. On the west it is bounded by the Ghats and the Nilgiris and the eastern boundary of Mysore, till it meets the Telugu near Kaddapah. It is also spoken in the north of Ceylon.

The MALAYALAM begins about Trivandram, and extends northwards between the Ghats and the sea to Mangalore, where it yields to Tulu and Canarese.

CANARESE is spoken throughout Mysore and Canara, Its northern boundary is conterminous with Marathi in Belgam and Kolhapur.

TULU or Tuluva is spoken in a small area round



Mangalore, along the Malabar Coast; KUDUGU in Kurg.

To return now to the Himalayas, where the various dialects are mixed together in great confusion; on the northern Assam frontier are found, in the following order from east to west, the Aka, Abor, Dofla, Miri, Mishmi; next to these is Bhotia, which carries us as far east as the Tista. Sikkim, or the country between the Tista and the Singhaleela¹ range, contains the Lepcha and Limbû dialects. The Sikkim Terai gives us the Dhimal, Bodo or Mechi, and Koch, which latter also occupy the plains of Koch Bihar, and the northern parts of Rungpur, Dinajpur, and Purneah. The Koch people now speak a bad Bengali. It is exceedingly difficult to find one who speaks Koch.

In Nepal, according to Mr. Hodgson and Dr. Campbell's researches, we find a perfect maze of dialects. Beginning from the Singhaleela range, we find Kiranta, which goes west as far as the Dudkoosi River ($86^{\circ} 44'$, Atlas of India). Sherwill found the Gurungs in the higher parts of Singhaleela, closely connected with whom are the Murmis. Along the lower hills are the Magars, who extend to the west as far as Palpa.

¹ This Singhaleela range is a remarkable spur which runs north and south from the foot of the Kanchanjanga mountain to the plains. It has a general elevation of 12,000 feet, and divides Nepal from Sikkim. It also creates a striking difference between the physical characteristics of the two countries; by checking the moist winds from the Bay of Bengal it renders Nepal open and dry and grassy, while Sikkim is overgrown with dense forests, and always wet and foggy.



Somewhere about here we should apparently place the Brahmú, Chepang, Háyú or Vayú, and Kusunda. The Chepangs and Kusundas are wild forest tribes, very shy and inaccessible. Mr. Hodgson was able to meet on one occasion only, during a hunting excursion, some Chepangs and secured a few dozen words of their language. He is the only European who has ever done so. Kusundas, even he could not get at.¹ In Central Nepal are the Newar, Pahri, and Bhramo (a dialect of Magar), also the Daráhi or Dorhi, Denwár, and Pakhya. The Tháru live in the Terai, between Champáran and the Khatmandu Valley, as far west as the River Gandak. These last four are classed among Indo-Germanic languages. The rest are Turanian, with more or less infusion of Hindi. The Parbattia or Pahária, a dialect of Hindi, is spoken all over Nepal, and is the court language. In Chapter I. it is called Nepalese, and by the people themselves the Khás. West of this again comes the Palpa, then the Thaksya, Sunwar, and Sarpa, the dialects of Kamaon and Garhwal, which carry us on to the Milchan of Kunawur, the Hundesi, and Tibarskad north of it. West of this come the Doghra dialects of the Panjab hills.

On the Southern Assam frontier we have the numerous Naga and Singpho dialects, the Mikir, the languages of the Khassia and Jaintia hill-men, the Boro in Cachar,

¹ See Selections from records of Government of Bengal, No. xxvii.



and the Garo in the hills of that name. The Kukis occupy parts of Tipperah and Chittagong, and the Mugs, Arracan and Chittagong. The wild tribes of the interior are underderstood to speak an infinite variety of dialects, but the published statements are not clear enough to admit of the geographical position of each tribe being distinctly fixed.

Such is a brief sketch of the general distribution of Indian dialects; it will at once be seen how much can be contributed to the sum of knowledge on this subject by persons residing on the boundary lines of languages.

NOTE.—Latham ("Comparative Philology") gives the Angami as a distinct language, but in a communication lately received from Assam, I find them classed as a tribe of Nagas. On many points connected with these wild frontier tribes we have yet much to learn.

Similarly, the arrangement of the frontier tribes of Assam is chiefly taken from Latham, but I find it differs from that given in the map of the north-eastern frontier of Bengal, published at the Surveyor General's Office, Calcutta. In that map the arrangement is as follows:—On the northern frontier of Assam, going from east to west, the tribes are arranged thus—Mishmi, Abor, Miri, Dofa, Aka. On the southern frontier thus, from east to west—Khamti, Singpho, Naga, Cachari, Jaintiah, Garo. I am unable to say which is correct, as my proposal to visit these tribes has been negatived by the Bengal Government; and the Surveyor General's map is professedly based on travellers' reports, which are not always to be relied on.

See Appendix C. for information supplied to me by the officers in charge of the various districts in Assam.



CSL

CHAPTER III.

STAGES OF DEVELOPMENT OF LANGUAGES.

EVERY language is divisible into two parts ; words and inflections. The *word* is the naked abstract statement or sound picture of a thing, a state, or an action ; as man, be, do. The *inflection* is that sound or syllable, or combination of sounds, which is used to modify the abstract word with respect to time, place, or relation.

The amount of cohesion between the word and its inflection varies in different languages from mere juxtaposition to complete amalgamation : that is to say, in some languages the inflections are merely written side by side with the word, while in others the inflection is so mixed up with the word as to be entirely one with it and quite inseparable from it.

The degree and nature of this cohesion form the most striking and simple means of dividing languages into classes, and defining the characteristics of each class.

In the first and earliest state, as exhibited in Chinese and other monosyllabic languages, the inflection is itself a word capable of being used independently, and not incorporated with the root in any way. Modifi-



cations of the primary idea are affected by prefixing or affixing words to the word which is to be modified, and these prefixes or affixes obtain, when so employed, a meaning different from that which they have when used alone, though no alteration takes place in their form.

The following examples will explain the above remarks.

In Chinese, "from" is expressed by prefixing the word *t'sung*, which, when used alone, is a verb, and means to *follow*; and affixing the word *lat*, which means to *come*; thus, *t'sung Peking lat*, means "from Peking." Here we see the modification of the abstract idea of Peking effected by two distinct particles, which are at the same time independent words, and are not incorporated or altered in form.

Yung yi-pà-tai; "by means of a sword."

Yung is a verb meaning "to use," *yi* means "one," and *pà-tai* "sword;" the whole sentence therefore means "use one sword."

The Chinese verb has no moods or tenses. The differences of time are indicated by particles, which are also, when used separately, verbs. Thus, *tseù*—walk.

tseù-liaù [=walk + finish]=walked.

i-kí-tseù [=already + finish + walk]=has walked.

yaú-tseù [=wish + walk]=will walk.

This is the earliest and most primitive state of speech. The next step is that the affirmatives or inflections lost

their meaning as separate words, and only continued to exist as inflections. Examples of this stage occur in Chinese also; thus the syllable *ti* affixed to a word indicates the genitive, and is equivalent to the English "of," but has no separate existence.¹

The most familiar example of a language in this secondary state is the Turkish. In this language the inflection is incorporated with, or rather agglutinated to, the stem-word; but so that the stem and its several inflections can be readily dissected, and the whole word reduced at a glance to its component parts.

Thus the syllable (now meaningless, whatever it may have been anciently) *in* added to a noun means "of;" *eh* or *ah* means "to," *dan* means "from." We have from *ev* a house—

evin—of a house.

eveh—to a house.

evdan—from a house.

If now to the stem-word *ev* we add *ler*, the sign of plurality, we get *evler*—houses, and this is declinable exactly as *ev*; thus:—

evlerin—of houses.

evlereh—to houses.

evlerden—from houses.

The syllable *im* means "my," and is added both to

¹ For these illustrations I am indebted to Summer's Chinese Grammar.



nouns and verbs.¹ Thus from *ev* we get *evim*, "my house." This again may be declined by adding the above particles; as—

اووشگ *evimin*—of my house.

اوومه *evimeh*—to my house.

اوودن *evimdan*—from my house.

A step further gives us *evlerim*, my houses; and this also is declinable, as—

اولروشگ *evlerimin*—of my houses.

اولرومه *evlerimeh*—to my houses.

اولرودن *evlerimdan*—from my houses.

In analyzing this last word, the peculiar genius of the Turkish language comes out clearly. We see each syllable in itself unalterable and indeclinable, but a collocation of these immutable syllables supplies all the necessary modifications of sense. Thus *ev*—house, *ler*—number, *im*—my, *dan*—from; the whole word is therefore house + number + my + from = my houses.

The same principle runs through the verb. Thus *olmak*, to be, makes in the present tense *oloram*—I am; which is *ol* = be and *or* = affix of the present tense and *am* (for *im*) = my; literally, "being my"—or "my being." In the plural this is clearer still; thus, "we"

¹ It would take us too far in a purely elementary sketch like this, to give the reason why *im* is attached both to nouns and verbs. An admirably lucid discussion on the subject will be found in Garnett's "Philological Essays," p. 289 *et seq.*

is *biz*, “ye” *siz*, “they” *anlar*; and we get by combining these with the participle *olor*, “being”—

اولورز *oloriz*—we are.

اولورسئز *olorsınız*—ye are.

اولورلر *olorlar*—they are.

Where the affixed pronouns, though under slightly altered forms, are clearly distinguishable from the verbal root, these terminations are perfectly regular throughout the verb. The results of this simple collocation are seen in the endless forms of the verb. No less than thirty-six distinct conjugations are possible in Turkish; and as each conjugation can have twenty-seven separate participles, each of which can be inflected with all three persons in singular and in plural, the total number of participles alone of a single Turkish verb may amount to five thousand nine hundred and thirty-two! Thus, from *almak*, to take, we have a participle *alah*, taking; this in the ablative is *almakda*, whilst taking; with the pronominal inflection it becomes *almagınızda*, on your taking.

The reflexive form of this verb might be *alınmek*, to take oneself, and negatively *alınmemek*, not to take oneself. The causal impossible form would be *alınderehmemek*, “not to be able to cause oneself to take,” and inflected into the ablative participle given above, we might have a form *alınderehmemeginizde*, which would mean “when you could not cause yourselves to



take!" The analysis of which would be, *al* = take, *in* = self, *der* = cause, *ek* = able, *me* = not, *meg* = when (participial form), *iniz* = you [*in* = thou, *iz* = sign of plurality], *da* = from; where each syllable has its own separate and independent meaning, though incapable of being used alone. I would not be understood to say that this word actually occurs, but simply that it *might* easily occur, and that it would be quite in accordance with the genius of the Turkish language.

I have entered into these details in order to give some idea of the peculiar organization of Turkish words—long strings of live syllables all united together by euphonic modifications; and each syllable having its own distinct meaning.

There are languages, however, in the agglutinated stage where each syllable has not a distinct meaning, though the stem-word is still separable from its suffixes.

As an example of this stage we may take the Telugu, one of the chief languages of the small but important Dravidian class. One of the best and most scientific philological works that the English language can boast of has been written about this class of languages. I allude to Dr. Caldwell's Comparative Grammar of the Dravidian languages, from which I condense the following short sketch of the Telugu.¹

¹ I must also acknowledge my obligations to pp. 73-104 of the Linguistic section of the Voyage of the Austrian frigate Novara, by Dr. Müller—a work which would repay translation into English.



The noun is unchangeable in its form. A root or stem-word becomes a noun when nominal suffixes are after added to it, or a verb when verbal signs are written it—but it undergoes in itself generally no change, or at most a slight euphonic one. Thus *gurramu*, “a horse,” becomes *gurramu-yokka*, “of a horse;” and the syllables *nu*, *ku*, or *na-ku*, *cheta*, *lo*, when affixed to the same unaltered word *gurramu*, produce the significations respectively of the accusative, dative, instrumental, and locative cases. By inserting the syllable *la* between *gurramu* and the above suffixes, the idea of plurality is acquired; thus, *gurramu-cheta*, “by a horse,” *gurramu-la-cheta*, “by horses.” Precisely the same process takes place in the pronouns, except that those most hard-worked parts of speech have been somewhat worn away in the course of ages: so that where, as we have in the nominative *nenu* = “I,” we find in the other cases only *nā*, as *nā-ku*, “to me:” *nīvu*, “thou,” but *nī-cheta*, “by thee.”

The verb is merely a stem-word combined with a considerable variety of suffixes, and, as before stated, would become a noun if combined with nominal suffixes.

A verb may be positive or negative, transitive or intransitive, active or passive, or casual.

Thus we have the positive, transitive, and active verb *pampu*, “to send;” this becomes negatively *pampa* (for *pampu-a* by addition of the negative particle *a*), “not to send.” The addition of *inchu* makes it causal, as *pamp-*



inchu, "to cause to send." If we affix the verb *baḍu*, "to suffer," we get the passive verb *pampu-baḍu*, "to be sent" (lit. "to suffer a sending"). In some languages of this class, the passive is formed by adding a particle *un*, meaning "to eat." The construction then resembles that so familiar in Hindustani, *mār khāna*, "to eat a beating," for "to be beaten."

The tenses of the verb are built up in the same way. We have first the stem-word *pampu*, conveying the naked idea "send;" to this is added the syllable *tu*, which gives the idea of present time, and to this again *nānu*, meaning "I." The result is *pamputunnānu*="send-now-I," or "I send;" similarly, *pamputunnāvu*="send-now-thou," or "thou sendest." The same process is followed in the formation of all the other tenses.

This class of languages shows us the affirmative particles beginning to be modified by euphonic considerations, dropping some of their letters in one place changing their vowels in another, and so on; but, though losing their original form, still independent and separable from the stem-word, which itself remains unchanged.

The next stage in language is that in which the words used as inflections have not only lost their original form, but have become so thoroughly incorporated with the stem-word which they serve to modify, as to become one word with it, and to be no longer capable of identification as separate words, except by patient

scrutiny and elaborate analysis. This stage is called inflectional. The languages of the Indo-Germanic family offer abundant examples of this class. In the more ancient languages of this family, the inflection is sometimes so loosely incorporated with the word as to be easily distinguished from it, though even in this case it is rarely that we find the inflection in its pure and primitive form. Thus, in Sanskrit अस्मि *asmi* = I am. This is easily resolvable into अस *as* = be, and मि *mi* = I or me; but although *mi* retains sufficient similarity to the pronoun of the first person to lead us to refer it thither, yet *mi* as a separate word, meaning I, me, or my, has no longer any existence in the language.¹

In Greek *ειμι* *eimi* is *ει* *ei* = be and *μι* *mi* = I, but *ei* as a separate word meaning *be*, or *mi* as a separate word meaning I, are not found; *ei*, again, is a corrupted form of *es*, the Sanskrit *as*.

In Latin *'sum* is *'s* = be and *m* = I; the short vowel *u* being inserted to facilitate pronunciation, and the *'s* is again contracted like the Greek from *es* the equivalent of the Sanskrit *as* "to be."

In Gothic, *im* = is; *i* = be, and *m* = I; *i* contracted from *is* = Sanskrit *as*.

Now in none of these languages does the affix retain its original form, while in most of them the stem-word also has been changed in some way.

¹ And, in fact, the discovery that the inflection *mi* is related to the pronoun of the first person is of recent date.



In the forms of the second person singular, Sanskrit असि *asi* [for *assi*] is अस *as* and सि *si*, but *si* for *thou* nowhere occurs as an independent word. In Greek, *ei* is for *εσι* *esi* [*essi*] or *es*=*be* and *si*=*thou*, neither of which exist alone.

Again, in the following forms we see the change that has taken place very clearly.

Sanskrit भरति *bharati* (*bhar*=bear, *a* euphonic, *ti*=he).

Greek φερεi . . . *pherei* (*pher*=bear, *e* euphonic, *i* [for *ti*] he).

Latin *fert* (*fer*=bear, *t*=he).

Gothic . . . *bairith* (*bair*=bear, *i* euphonic, *th*=he).

English . . . *beareth*—still further modified into *bears*, where a lengthened analysis is requisite before we can see the connection between the final *s* and the pronoun *he* (Sanskrit *sa*).

It is unnecessary to give a specimen of a language in this stage, as we need not go further than the common Latin or Greek grammars for an example.

There is yet one more stage in languages; the last which any language has yet reached. It is that in which the inflection has become so abraded and destroyed that not a trace of it remains, and with the trace of the inflection the modification of sense which it effected is in danger of being lost also. Additional words have to be called into use to retain the distinctions of sense, and languages in this stage resemble



at first sight those in the first or syntactical stage, in so far as they again express modifications of sense by particles having an existence independent of the stem-word. Thus in French the terminations *o* and *at* in the Latin *porto* and *portat* being lost, there was nothing to distinguish *porte* = *porto* from *porte* = *portat*; consequently it became necessary to prefix the pronouns *je* and *il*. But here we have a repetition, because if *porte* = *porto*, and *porto* means "I carry," then *je porte* means "I, I carry;" so also *il porte* means "he, he carries." In English something of the same sort occurs, though not to such an extent as in French. *Goes* = *goeth*, and that again = *go* + *he*, therefore *he goes* contains the pronoun twice over.

A great change in the nature of languages accompanies the transition from the inflectional stage to the analytical. When the relation which exists between words is expressed by changes in the form of the words themselves, it is immaterial what place in a sentence the words take; but when you have no terminations at all, it is only from the order of the words that the meaning can be discovered. In the words of a recent French writer—"On sait que la grande différence qui distingue le latin de notre langue présente est que le français exprime le rapport des mots par leur ordre, tandis que le latin l'exprime par leur forme. L'idée ne change point en latin si au lieu de dire : *canis occidit lupum*, on dit, *lupum occidit canis*; elle devient tout en



français si l'on dit *le loup tua le chien* ou *le chien tua le loup*; en un mot le français reconnaît le sens du mot à sa place; le latin, à sa désinence."¹

This rejection of terminations is, I think, in some respects, a great loss to a language. It is chiefly owing to this, that Englishmen as a rule write such bad grammar. The sense depending entirely on the order of the words, a long or involved sentence in English requires so much care and thought, that few writers are sufficiently clever to avoid falling into mistakes; whereas the German, who has retained a considerable number of his inflections, can safely indite a sentence half a page in length without fear of confusion.

The practical application of the above remarks is simple. In previous chapters it has been shown that there are in India languages of many classes, and a knowledge of how these languages are classified with regard to their development is essentially necessary to that study of them which it is the aim of the present remarks to encourage; namely, the observation and recording of languages hitherto unstudied and unknown, and in which consequently the observer must rely entirely on his own resources.

We have given above four stages of development in language.

1st. The collocational or syntactical stage, as seen in Chinese.

¹ Brachet, *Grammaire Historique de la langue Française* (Paris, 1867.)



2nd. The agglutinated stage, as seen in Turkish and Telugu.

3rd. The inflectional stage, as in Sanskrit, Greek, and Latin.

4th. The analytical stage, as exemplified in modern French and English.

In the first stage may be placed the bulk of the languages in classes 1, 2, and 3 of the Turanian family; that is to say the Thaïc, Lohitic, and Himalayan classes. It may be safely laid down that the most cultivated of these languages has not got beyond the agglutinated stage, while the wilder, and less cultivated, are as yet in the syntactical stage only. The literary languages of these classes are well known to science, it is only in the ruder ones that fresh discoveries remain to be made; the observer therefore is not likely to err who classes any new dialect of the above languages as syntactical, and in his records treats every syllable as a distinct word with a distinct meaning of its own.

In the second, or agglutinated stage, we have the languages of classes 4 and 5 of the Turanian family, the Kōle and Dravidian. In the Kōle languages agglutination pure is found; in the Dravidian, euphonic tendencies have operated so extensively as to give them in some cases almost an inflectional appearance. The same remark holds good here as in languages of the first three Turanian classes; namely, that while the most cultivated have not advanced beyond, nay have



hardly attained, the inflectional stage, the ruder ones are still in the agglutinated.

To the third, or inflectional stage, belong the languages of the Indo-Germanic family, classes 1 and 2; but with this exception, that some of the more advanced dialects present points of similarity with the last or analytical stage.

Again we find the twice enounced rule hold good, that the more advanced of these languages are almost analytical, while the less advanced are purely inflectional. Thus quasi-Sanskrit case-endings and verbal forms are found in greater frequency in Bengali than in Hindi; and Bengali is therefore less advanced than Hindi; the latter being in many respects analytic, the former almost purely inflectional.



CSL

CHAPTER IV.

ON THE FAMILY CHARACTERISTICS OF LANGUAGES.

THE classification of languages according to their stages of development is not sufficient for purposes of analysis.

It is only one step in the process.

Two languages may be in the same stage of development and yet differ widely ; for instance, Sanskrit and Hebrew are both in the inflectional stage, but no two languages could differ more entirely in structure or characteristics.

It is necessary, therefore, to give a slight sketch of the principal characteristics of each family in each stage of development.

With regard to the Indo-Germanic or Aryan family, in the first place there is this difficulty—that it occupies so extensive a portion of the earth's surface, and the Aryan race has from the earliest times enjoyed so rich and progressive a form of civilization, that it is difficult to give a complete and exhaustive resumé of its distinctive characteristics in the narrow limits of this sketch.

I shall therefore only indicate the more salient and striking points of the subject.



Aryan languages are found in the inflectional and analytical stages, but the transition from the former to the latter stage is not marked by any clear and definite line. In languages, as in all other departments of nature, the various classes, genera, and species merge gently into one another; there are no harsh lines of demarcation, no barriers; nature is not a box divided into compartments, it is a chain of many colored links.

We find languages which, though not yet purely analytical, are still no longer purely inflectional; and consequently it would be easy to divide this stage into two, early and late inflectional.

It will be sufficient, however, in this place, merely to state the fact, as the distinctions between the two, if worked out fully, would lead us into details embarrassing to one making for the first time an acquaintance with the subject.

The leading structural characteristics of the inflectional Aryan languages are as follows:—

Nouns are distinct from verbs. The same word cannot be at the same time a noun and a verb; it is only in the syntactical and agglutinated stages that this can take place.

The noun possesses three numbers—singular, dual, and plural; and numerous cases, each distinguished by a peculiar and inseparable termination.

The number of these cases in the singular and plural varies from five to nine in different languages. In the



dual no language has preserved all its cases; the Sanskrit and Zend have only three forms for eight cases, the Greek two, the Latin has entirely lost its dual.

The case terminations cannot be dispensed with, or elided, or separated from their roots. They are not constant, but vary according to the nature of the noun to which they are appended.

Thus there arise numerous declensions.

In all declensions, however, the fundamental rules are the same; and the differences arise merely from the operations of the laws of euphony which regulate the method in which the case-endings are added to nominal roots terminated respectively by vowels and consonants.

These laws are still clearly traceable in some languages, as Sanskrit and Zend, less clearly in Latin, Greek, and Armenian; but even in those languages where most effacement has taken place, it is still possible by a rigid analysis to find traces of these operations.

The verb displays great wealth of terminations. It has, like the noun, three numbers,—singular, dual, and plural; though as in the noun, so also in the verb, the dual has perished from some languages of this family.

Each tense possesses three separate forms for three separate persons in each number, but, as a rule, has not separate forms for the genders. Thus “I speak,” “thou speakest,” “he speaketh,” are expressed by three different forms:—but “thou (man) speakest” is not different in form from “thou (woman) speakest;”



this is mentioned because in the Semitic family there are separate forms for genders.

The personal terminations of the verb are abraded pronouns, or rather pronominal types.

There are separate sets of forms for each time to which the action of the verb refers, also for the conditions under which the action is performed. Each of the sets of forms (or moods and tenses, as they are called), possesses its full set of personal forms.¹ The verbal root first undergoes certain changes to qualify it to express the necessary modifications of sense, and to this modified form the abraded pronoun, which performs the functions of a personal termination, is added. The laws of euphony are, of course, in full operation in the junction of the verb with the termination.

Past time is usually denoted by augment and reduplication; augment being the prefix of a short *a* or its equivalent, reduplication the prefix of the first consonant and vowel of the root, sometimes with a phonetic variation in the vowel, and sometimes in the consonant.² Some languages of this class have lost both the augment and the reduplication. The pronouns, though they differ very much in details, yet retain the same characteristics. These are—

m for the 1st person.

t for the 2nd.

s for the 3rd.

¹ The imperative in some languages is an exception to this rule, as it has lost its first person singular.

² Sansk. *gam* to go, *jagama*; *ji* to conquer, *jigáya*.



Another striking characteristic of this family is its power of expressing complicated ideas or strings of ideas by compounds. Several words are joined together, and the case and tense-endings are added to the last word only, the first member of the compound being either a preposition or a noun, or even a verb. This power is not possessed by other families.¹

The above remarks will be found to apply, strictly speaking, only to the ancient Aryan languages, as none of the modern tongues are altogether inflectional; nearly all of them are, at least, in some respects analytical. The same process which, in the European members of this family, has changed the Latin into Italian, French, and Spanish, has, in the Indian members, operated to transform the Sanskrit into Hindi, Bengali, and Marathi.

There are, however, several points of difference in the position of the two classes.

1. In the European group we have so much historical evidence, and so many documents written in successive ages to adduce as examples, that we can trace the whole process of growth and conversion from the pure classical Latin to pure modern Italian. In the Indian group we cannot do this. We have Sanskrit at one end of the line and the modern languages at the

¹ Of course great differences occur in the different members of this class, but the above remarks will serve briefly to indicate the leading characteristics of the inflectional Aryan languages; for details the reader is referred to Bopp's Comparative Grammar.



other, but the intervening space is enveloped in darkness which has not as yet been dispelled. Consequently, we are often in doubt as to the origin of any particular form or inflection, and are at present very much in the dark in many respects.

2. There are in the Romance languages many words which are not traceable to a Latin source, being importations either from the Teutonic branch of the family or from some other language. Similarly, there are evidently many words in the modern Indian tongues, not of Sanskrit origin, but whence these words come we cannot decide. The most probable hypothesis—it is no more than a hypothesis at present—is that they are relics of an earlier speech, used by those races who were driven out by the Aryan invaders.

3. The inflections of the Romance languages accurately reflect, especially in the verb, the features of the Latin; so much so as to be unintelligible without it. Thus the Italian tense

ebbi	<i>I had, etc.</i>	avemmo	<i>we had, etc.</i>
avesti		aveste	
ebbe		ebbero	

seems at first sight to contain words from two entirely different verbs; but when we place side by side with it the

Latin tense	habui	habuimus
	habuisti	habuistis
	habuit	habuerunt

we see that the Italian tense is only a corruption of the



Latin, and that the difference in the form of the first syllable in the Italian arises from euphonic considerations. In the second persons singular and plural, the Latin terminations *isti* and *istis* are heavy, *i.e.* they contain a double consonant which causes the stress or accent to be laid on those syllables, consequently the simple or light form *av* is used. In the other tenses the terminations of the Latin are light, and the accent falls on the initial syllable, which therefore requires to be strengthened, and the form *ebb* is used.¹ It is hardly, if at all, possible to find a parallel to this in the Indian languages. Their verbs do not as a rule at all reflect the Sanskrit verb; indeed, I know nowhere two things more dissimilar than the complicated but euphonic Sanskrit verbal inflections, and the few simple tenses of the Hindi verb. It is clear that when the purity of the Sanskrit type was lost, degeneration went on far more rapidly in the east than in the west.

And yet as regards stages of development, we can by no means place the European languages on one side and the Indian on the other. The next few pages will shew that both in Europe and India some modern languages

¹ It must be remembered that for a long time,—as late as the 12th century, in fact,—Latin remained the language of literature, and Italian was considered as a mere jargon unfit to be written. When this prejudice was overcome by Dante and others writing in the “lingua volgare,” the words were, in many cases, written down as they were pronounced, without any regard to their derivation. Much of the peculiarity of Italian spelling is due to this fact.



have preserved an inflectional form, while their immediate neighbours are analytical. In either case, however, a large proportion of the forms of the older languages has ceased to exist in its fulness, and their place has been supplied either by pre- or post-positions or by combinations of words, technically called "auxiliaries."

In the noun the dual has disappeared entirely, and the number of cases in the singular and plural has diminished so much by abrasion or elision of case-endings, that modifications of the sense have to be rendered clear by prepositions, some of which are used with one form, some with another; or, technically speaking, some govern one case, some another.

In some languages, as in English, French, and Hindústání, case-endings have entirely disappeared; in others, as German, Bengali, and modern Greek, they still exist, though in a smaller number than formerly.

Similarly the verb has lost the greater part of its long array of moods and tenses, and has to express its modifications by the additions of the verbs *to be*, *to have*, etc.; the personal distinctions of each tense have been so much dropped, that it is generally necessary to prefix a pronoun to indicate the person.

The second person retains its individuality more distinctly than the other persons.

The amount of abandonment of ancient forms varies very much; some languages are analytical as regards their nouns, but inflectional as regards their verbs; as



the modern Persian, where the noun is declined by pre- and post-fixes, while the verb still retains its forms in a fair state of preservation; *e.g.*—

روم ravam, <i>I go.</i>	رويم ravem
روي ravi etc.	رويد raved
رود ravad	روند ravand

and the Armenian singular of the Indic. pres.

սիրեմ sirēm.

սիրես sirēs.

սիրէ sirē.

which will bear comparison on the score of integrity of pronominal affixes with भवामि bhavāmi, भवसि bhavasi, भवति bhavati, भवावस bhavāmas, भवथ bhavatha, भवन्ति bhavanti, or with amo, amas, amat, amamus, amatis amant, or with τυπτω, τυπτετε, τυπτετε, τυπτομεν, τυπτετε, τυπτοσσι.

Modern high German again possesses a large amount of nominal forms, both for substantives and adjectives, while its verbal system proceeds almost entirely by auxiliaries.

English has lost all its case forms in nouns (unless we take the 's of the genitive in, *e.g.*, *man's hand* to be a relic of a case-ending); it has also lost all tenses but two, *e.g.*, *love*, and *loved*: it has lost the plural personal endings, and is fast losing those of the singular,—*thou lovest, he loveth*, have given way to *you love, he loves*.



Hindústání has lost the Sanskrit nominal inflections, and declines through affixes; the separate form for the plural is rapidly going out of use,—*báten*, *baton* as the plural of *bát*, are restricted to certain parts of the country.¹ The verbal forms still remains, but in an extremely mutilated and scarcely recognizable dress; so much so, that it is a question whether the forms of the Hindi verb can be derived from those of the Sanskrit. I am of opinion that they can, but the question is still *sub judice*. The aorist, future and imperative, are the only true *tenses*, the presents and preterites are, strictly speaking, participial forms, thus the aorist is—

मैं मारूँ main mārún	I beat
तू मारै tú máre	thou beatest
वुह मारै wuh máre	he beateth.

The terminations here, *ún*, *e*, *e* are derived either from the Sanskrit terminations of the present indicative, *ami*, *asi*, *ati*, or from those of the potential *eyam*, *es*, *et*.²

¹ The plural in Hindústání is now most usually indicated by the affix लोग *log* = people, or सब *sab* = all, or some such word. This, however, is only the case when it is necessary to bring out the idea of plurality with great distinctness. In ordinary cases there is no difference made between the singular and plural—thus we say, ایک بیل *ek bail*, ‘one bullock,’ and بیس بیل *bis bail*, ‘twenty bullocks.’

² More probably the indicative for (1) the *ún* of the first person could not come from *eyam*; and (2) in old Hindi the forms of the 2nd and 3rd persons end in *ai*, which is probably a contraction of *asi* and *ati*, by leaving out the unsupported consonant between two short vowels, according to a well-known Prakrit rule. (3) In the Sindi the form of the 3rd person plural



The future differs from the aorist only by the added syllable गा *gá* (fem. गी *gi*; pl. m. गे *ge* f. गी). This is probably derived from the Sanskrit root या to go (not गम as has been suggested), on the same principle as the English phrase, "I am going to beat," or French, "je vais frapper."

The present definite is really a participle, मार्ता *mártá*, "beating," and मैं मार्ता = "I (am) beating." So also with the preterite मारा = "beaten," or "having beaten."

In Sindhi, Marathi, Panjabi, Gujarati and Bengali, the verb is formed in a manner closely akin and similar to Hindi, only in some cases with a fuller, in others with a scantier development.

It will be necessary to sketch briefly the structure of the Semitic family here, because the contrast will bring out in a clearer light the peculiar characteristics of the Aryan family.

The first and most striking point is that in a Semitic tongue all words are derived clearly, and by an easily traceable process, from one root consisting of three letters. This trilateral root is the all-pervading and sustaining element of Semitic speech.

Whether consisting of three *syllables*, as in Arabic and Ethiopic, two as in Hebrew, or one as in Syriac,

is *márinī*, which could in no case come from the optative of Sanskrit, which makes its 3rd pers. pl. in *eyus*. The sense of "I may beat," etc., given to this tense in most Hindústáni grammars, is entirely wrong. No native ever uses it in this sense. It is really the indefinite present indicative, or, as I have called it in the text, aorist.



the root is invariably composed of three *letters*; these three letters in their simplest form represent the third person singular of the past tense; קָטַל *Katal*, Heb., قَتَلَ *Katala*, Arab., قَتَلَ *Ktal*, Syr., "he killed," RaKaBa "he rode," KaTaBa "he wrote": from this trilateral root, by lengthening or shortening, inserting or eliding, the vowels which accompany the three immoveable consonants, and by prefixing or affixing syllables, are formed all the possible modifications of the primary idea.

Under no circumstances do the three consonants of any given root change their relative positions; vowels and consonants may be inserted, prefixed, affixed, elided, lengthened, shortened, but through all these changes the three radical letters preserve their relative positions unmoved.

Three letters only, representing the vowels *a*, *i*, and *u*, or their corresponding consonantal sounds, are subject to changes of a phonetic nature; and in verbal roots which contain any of these letters it cannot be said that the three radicals remain unchanged.¹

The noun has only faint indications of case, and that only of three cases; there are no case-endings; simply the final vowel of the noun alters, being *u* for the

¹ The three letters ا, و, and ي are considered as consonants, and from their liability to modification are called in Arabic حُرُوفٌ عِلَّاتٌ *huruf-illat* 'weak letters.'



nominate, *i* for the dative, and *a* for the accusative. Hebrew and Syriac have lost these terminations entirely, and modern Arabic only occasionally employs them. Case modifications are generally indicated by prefixes. There are three numbers, singular, dual, plural, and two genders.

The verb has only two tenses, and much confusion exists in the employment of them with regard to time. The preterite in Hebrew is used with a prefix for the future, and the future with the same prefix becomes a preterite, while both are used at times for a present.¹

There is no form for present time, nor for possible, conditional, optative, or contingent modes of action.

Each tense has separate forms for gender. "He spoke" differs from "she spoke," and so throughout the tense, except in the first person, where the presence of the speaker renders the distinction unnecessary.

The personal terminations of the verb are pronouns, somewhat abraded, but not so much as in Aryan languages. There is this peculiarity, however, that whereas in the preterite these person-endings are affixed to the root, in the future they are prefixed. The reason of this appears to be, that in the tense which we generally call the preterite, but which might more properly be called the definite, the attention is more particularly called to the idea of the verb, which therefore comes first; while the personal-endings, being less

¹ I allude, of course, to the *vav conversive*.



important, are subjoined : whereas in the future or indefinite tense the idea of the verb is less strikingly impressed on the mind, and the personal particles become more prominent, and are therefore prefixed. In support of this view it may be pointed out that in Syriac, the most mutilated of the tongues of this family, the personal-endings are far more abraded in the preterite than in the future.

Pronouns have two forms, separate and affixed, the affixed forms are attached both to nouns and verbs, and are contractions or alterations of the separate forms. Where the object of an action is a pronoun, it is treated in Aryan languages as a noun, and placed in its proper case and place in the sentence. In a Semitic language, on the contrary, the pronoun is affixed to the verb of which it is the object; thus, in Hebrew, קטלתני *ketal-tani*—"thou hast killed me," is = "קטלת *ketalta*—thou hast killed," and "ני *ni* (shortened from *ani*—I)—me."

Similarly in nouns, where a possessive pronoun would be used in an Aryan language, an affixed pronoun is used in Semitic; thus, "my book" (*liber meus*) is in Arabic کتابي *kitabí*; where *kitab* = book; *í* = my (literally, book-my). All the languages of this class are in the inflectional stage; modern Arabic alone is beginning to show a slight tendency to become analytic.

Languages of the Turanian family have been exemplified fully in the preceding chapter, and, besides, from their vast range, and the little that is, as a rule,



known of them, it is almost, if not quite, impossible to give, in a few words, a clear idea of their peculiar type.

The following, however, may be noted:—

1. A system of accent or tone of a very subtle and delicate nature, varying in amount and degree from the elaborate tones in Chinese and Burmese to the simple laws of vowel collocation in Magyar.

2. The formation of cases of nouns and tenses of verbs by adding to an unchangeable monosyllabic root certain modifying syllables, whether distinct words capable of being themselves used as roots, or words which have lost the power of separate existence. The former belong to the syntactical; the latter to the agglutinated stage.

3. The absence of any modification of the root of a verb to express persons, or of the root of a noun to express cases.

4. As a natural consequence of the above, a tendency to a monosyllabic style of structure, which further results in an absence of compound consonants. As a rule, each consonant requires its vowel; such a word, for instance, as the Sanskrit *smṛiti*, or the English *strength*, would be almost beyond the power of Turanian organs to pronounce. This remark is more true of the southern than of the northern branches of this family.

Some Turanian languages—as, for instance, Karén—have a system of affixed pronouns, similar in some respects to that of the Semitic family.



These are more fully developed, and appear in greater number in the northern branch of the family.¹ In the Dravidian class, as well as in the Himalayan, they are almost, if not entirely, absent.

¹ Especially in the Altaic class; there is a valuable work by Castrén on this subject. "De affixis personalibus linguarum Altaicarum."



CSL

CHAPTER V.

ON DIALECTS.

MANY as are the languages spoken in India, yet an enumeration of them by no means exhausts the list of varieties of human speech in this country. No language is spoken in precisely the same way over the whole of its area. In fact, it may be said that no language in the world, however small the extent of country in which it is spoken, is free from dialectic variations. Much discussion has taken place on the question of what constitutes a language, and what a dialect. What amount of deviation from the classical or central standard of a language is compatible with merely dialectic variation, and at what point is the boundary passed, and a new language constituted? It appears probable that no determination will ever be arrived at on this subject, because it is one on which it is impossible to lay down a general rule. Geographical situation, political and physical accidents, education, habits, religion, all have their bearings on language.

As an illustration of the first we may take England. Sea-girt on every side, English glides imperceptibly into no other language. The Celtic of Scotland, Ireland, and Wales is marked by definite limits. No one ever



hesitated whether to call the Lowland Scotch an English or a Celtic dialect. Even towards Holland and Flanders, peopled by races closely allied to our own, the line is very clearly marked; though Flemish is extremely like the dialect of Yorkshire, yet it would never occur to any one to call it a dialect of English, or deny its claim to be considered a language.

Again, the political accident of Runjeet Singh's succeeding in establishing for a few years an independent monarchy in the Panjab, has led to the speech of that country being considered as a language, though it has intrinsically no more claim to the title than Bhojpúri or Brij-bhashá. In the case of Panjabi, the influence of religion also comes in. The Sikh religion gave a sacred character to the Gurmukhi letters, in which the Holy "Granth"¹ was written, and thus partly also a distinct existence to the language in which it was composed.

The physical accident of the singular conformation of the Himalayas has changed a number of Tibetan dialects into so many distinct languages; that is to say, the total absence of valleys with easy passes over the intervening mountain ridges, and the perpetual recurrence of long spurs rising steeply up from impenetrable forest depths to a height of seven, eight, and twelve thousand feet, make the Himalayas the most difficult mountains in the world to travel in. Locked up in its

¹ The Scriptures of the Sikh religion.



gloomy wooded glens, cut off from all communication with its fellow tribes, each little section of the Southern Tibetans has worked out for itself such variations on the now scarcely traceable mother tongue, that to apply the name of dialect to any one of the hill *bhāshās* would be to lose the very meaning of the word language.

Education retards the formation of dialects as much as isolated position assists it. An unwritten language possesses, as a natural consequence, no standard of purity by which all other forms may be measured. From the earliest times we find that nations who were possessed of a written character have kept their language comparatively united. The Sanskrit written at Ujjayin did not, as far as we know, differ in any respect from that written at Ayudhya. The Hebrew Scriptures give us but few indications of dialectic varieties in Palestine. And although, in the first instance, it may be objected that the numerous forms of Prakrit¹ were contemporaneous with Sanskrit, yet it must be observed that Prakrit was spoken by the illiterate masses; Sanskrit, if at all in its most elaborate form, by the educated only. If the plays are to be taken as any guide to the actual practice of those times, the number of persons

¹ The number of dialects of Prakrit is variously given by different writers. Vararuchi, the oldest and leading author on the subject, knows only of five; called respectively Māhārāshtri, Sauraseni, Paisāchi, Māgadhi, Apabhransa. The tendency to minute subdivision so characteristic of the Indian mind has led later grammarians to multiply dialects ad infinitum. So that Rāmataraka Vagisha, whom Lassen happily calls "bonus nugator," exhibits a list of no less than twenty-two kinds of Prakrit.



sufficiently educated to speak Sanskrit must have been very small. Only kings and Brahmins are made to talk in it. With the decline of Sanskrit literature, and during the darkness of mediæval times, when learning, in India as in Europe, was scarcely kept alive, the processes of corruption were commenced and developed which have resulted in the numerous dialects which we now meet in India.

The test which has been proposed, and, in fact, insisted on with much certainty of manner by some, is that of mutual intelligibility. If two persons using different forms of speech can understand each other, then those forms are to be called dialects of a common language; if not, they are separate languages. The weak point of this rule, and that which renders it entirely untrustworthy, is that the human intellect is a very varying organ. A rustic addressed by a citizen will not understand a word that is said to him, while an educated person with a little reflection will generally make out the peasant's meaning after a while.

Nor is this all; cases occur in which the words, or a large proportion of them, are common to two languages, and yet the grammatical forms are entirely different. This is the case in Hindi and Bengali, both of which possess an extensive stock of words of Sanskrit origin, and quite identical in all respects, yet the grammatical system of the one is as different as possible from that of the other, and the result is that they are not mutually intelligible.



A second case is where the grammatical forms are almost the same, while the words are very different. This is the case with Hindi and Panjabi; though the latter has many words in common with Hindi, yet it retains so many Sanskrit words which have dropped out of Hindi, and *vice versâ*, and has so many purely local terms of its own, that the vocabulary is very different from that of Hindi. And, again, mutual intelligibility does not exist.

There are besides instances where two forms of speech are identical, or nearly so, both as to words and inflections, and yet the pronunciation of the one differs so much from that of the other, as to produce on the ear the effect of a different language. Syriac and Chaldee, Lowland Scotch and English, are examples of this sort.

Now, in all three of the above cases we have mutually unintelligible forms of speech, and yet Bengali is undoubtedly a separate language from Hindi, Panjabi is getting to be accounted once more a dialect, while Syriac is always held to be a dialect of Chaldee, and Lowland Scotch of English.

Where so much obscurity and uncertainty, both in theory and practice, exists, it will save trouble to ignore the question altogether for the present as a general question of philology, and in each language to note the varieties of form, whether in vocables or grammar, as fully as possible. Subsequently, when proceeding to



OUTLINES OF INDIAN PHILOLOGY.

CSL

consider the languages of India at one view, each one of the subordinate forms may be affiliated to that one among the classical and cultivated languages to which it most approximates. When once every form of speech in India has been investigated, it will be easy to settle whether any given form is an independent language or only a variation of a language; but, as said before, no rule can be laid down in the present state of knowledge.

In the first chapter of this work, I have classed the languages and dialects according to the usually received arrangement, but I have thought it necessary to add the foregoing remarks, and those which follow, to prevent misconception on the subject.

In the languages of the Aryan family, the existence of a well-known common origin in the Sanskrit renders it unnecessary to enquire closely into the line which separates languages from dialects. It is convenient to range the Bhojpuri as a dialect of Hindi, and the Bengali as a distinct language, because, although the former differs in grammatical forms as widely from Hindi as the latter, yet, in the country where it is spoken, it is confined entirely to the peasantry, and every one who possesses a little education drops it for Urdu. Indeed, in Behar, and the valley of the Ganges generally, a man's social status and respectability may be pretty accurately fixed by the amount of the peasant dialect which he uses. This line of argument will not, however, hold good in other cases. The Magars, for in-



stance, who live in the Nepalese hills, are getting ashamed of their little language;¹ they are a warlike race of mountain shepherds, and as such have been largely drafted into Jang Bahadur's army. There they have learnt to speak the Khas or Court Nepalese, an Aryan dialect, and only use their own Tibetan among themselves. We have, therefore an example perfectly analogous to the Bhojpuri. The educated and higher classes of Magars are gradually dropping their native speech for Nepalese, and yet it would be far from correct to call Magar a dialect of Nepalese.

The object of the above remarks is this. In attacking the countless forms of speech which prevail on our Indian border, it is important to have some definite principle to go on in classing them. Are, for instance, the Aka, Dofla, Abor, Miri, and Mishmi all separate languages; or can they, by means of any of the foregoing considerations, be brought under one or more general heads? If they can, the benefit, both scientifically and practically, will be great; scientifically, because the existence of one language broken into dialects gives us only one definite fact to deal with instead of many, and because if these forms of speech can be proved to stand to each other in the relation of dialects, and not merely cognate languages, their peculiarities will throw immense and direct light upon each other, and our eventual grasp of the whole will be firmer and

¹ The Nepalese call it a **लेहरा बोली** "*lehra boli*" or "dirty language."



clearer, and the solution of the question of their position in the general family easier ; practically, because it is less difficult to learn one language than twenty.

At the same time, the consciousness that to prove these forms of speech to be dialects rather than languages would be an advantage, ought not to lead any one to enter on the study of them with even the wish to obtain this result. Nothing is more fatal to the proper investigation of a scientific question than approaching it with a pre-conceived determination to see facts in a certain light only. I do not go so far as to deprecate the deliberate twisting of facts to suit a theory, because this is little short of dishonesty, and it is hardly necessary to point out that it ought to be avoided. But without going to this extreme, many persons do not sufficiently guard against the temptation to grasp too eagerly at facts which seem to make in favour of a pre-conceived idea, and to slur over those opposed to it. The mind of one who enters on the study of a new language or a new science should be absolutely free from bias, or if this is impossible, there should be an ever-present stern watchfulness over oneself. Facts should be ascertained beyond doubt, calmly weighed and recorded, and not until the whole investigation is complete, or at least till it has been carried as far as the student's circumstances permit, should enquiry as to the bearing of the facts, and the inferences to be drawn from them, be entered on.

In conclusion, I will here sum up, at the risk of



partial repetition, the principal points to be borne in mind with regard to dialects.

1. The test of mutual intelligibility is a very unsafe one, as it depends on the intelligence of individuals. The savage and the peasant will exaggerate it; the man of education will make too light of it.

2. By taking into consideration certain influences which have operated on the people, the mutual intelligibility test may, however, be brought to bear to this extent, that it may be fairly said of two forms of speech that if they are not mutually intelligible, they ought to be, and in fact they may often be so much alike, that the student, who is master of one, would almost, if not altogether, understand the other, though two natives could not.

3. These influences are, geographical position, civilization, political and physical accidents, religion, difference of pronunciation, education.

4. Mispronunciation of words by the uneducated, where the educated pronounce them rightly, is a peculiarity which should not be mistaken for a dialectic one, as is too often done.

5. The fact that a form of speech is used only by a small number of people is no argument against its being really an independent language. It may be that the tribe which speaks it was once larger, and has shrunk to its present small dimensions from war or other causes; it may also be that, like the Hebrews, a special religion



has marked and set apart the tribe, and prevented its spread ; or, in the third place, it may be that, like the Basque, all its congeners have been swept away, and their places supplied by tribes from other families of the human race.

6. It is a mistake to suppose that rustic dialects are degenerate or debased forms of a language. In those languages with which we are most familiar, it has generally been found that many different dialects have existed side by side from the earliest times. One of these by some accident has been taken up and cultivated, has produced a literature, and been enriched with additions from other sources, while the others have remained in their original obscurity. Far, however, from being debased, they often retain early and pure forms of words which have dropped out of the cultivated dialect.

Thus, in England, the accident, so to speak, of London being selected as the capital, has led to such prominence being given to the Southern English as to make it the foundation of the classical speech ; while the Northern English, at one time its rival, not only in speech, but in literature, has sunk almost into complete obscurity.

The publication of Luther's translation of the Bible into the Saxon dialect of High-German, in like manner, rendered that dialect the classical and literary type of the language. So also the selection of Delhi as their capital by the Moghul invaders, is the reason why



modern Hindústání takes the form in which we now know it. It is the Hindi dialect of Delhi, and the parts adjacent, polished and mellowed, and supplemented by a large stock of Arabic and Persian words. Had the Mussulmans fixed their head-quarters at Patna, for instance, Hindústání would probably have had the Bhojpuri dialect as its basis.

From the above remarks it will be seen that there is much to be said and studied under the head of dialects. With special reference to the object of these outlines, it may be pointed out that the difficult problem of tracing the origin of many words and grammatical forms in modern Hindústání, may receive great aid from a study of the various Hindi dialects, which preserve ancient and transitional forms with great fidelity. Thus the process by which the Sanskrit भू *bhú*, to be (भवामि *bhavámi*, भवसि *bhavasi*, भवति *bhavati*, etc.), became होना *honá* is more readily understood, when we discover that in Bhojpuri भा *bhá* and भया *bhyá* are still used.

Those, therefore, who do not live in places where hitherto unexplored languages are spoken, may yet do good service to the cause of Indian philology by noting and investigating the local dialects of their district. These dialects are fast disappearing, and in a few years perhaps they will be extinct; it is important, therefore, to record them ere they pass away, and so secure all the aid that may be derived from them while they are yet in existence.



Under the head of dialects, I may introduce some remarks on a state of things observable in many parts of India; namely, the existence of local peculiarities of speech which do not amount to dialects, and yet cannot fairly be overlooked. In some districts in the Hindi area we find all the ordinary implements of agriculture and household furniture known by names peculiar to the place, while the rest of the local speech is quite in accordance with the ordinary type of the language. I will give a few examples of this.

The educated natives of Delhi and Agra, although speaking an extremely good Hindústání as a rule, always say *vishá*, *visko* for *uská*, *usko*; and the lower class in that neighbourhood use *jáke* for *jiske*, etc.

In Purneah, the small bullock-cart used by the better class of merchants and others, and elsewhere universally called रथ *rath*, is known as रेहू *renrhú*.

The shed in the wilderness built by cowherds to rest in at night is called on one side of the Gandak बथान *bathán*, on the other अरार *arár*,¹ while in the neighbouring district the words घारी *ghári* and सार्घर *sárgar* are used for the same thing. The word घारी again is in some parts of the country used to signify a cowshed attached to a house; for which also गोशाल *goshál* is used in Tirhút, and गोहाल *gohál* in Purneah.

¹ Or अड़ाड़ *arâr*. The spelling of words in the local dialects is very uncertain; the latter form is probably the correct one, as I connect the word with the Hindi आड़ *âr*, "a shelter."



CSL

OUTLINES OF INDIAN PHILOLOGY.

Hindi is peculiarly rich in these provincial words, so are Panjabi and Bengali, and I doubt not Marathi and Gujarati also ; a series of glossaries of the words peculiar to various rural districts is much wanted, and the published dictionaries might be very largely enriched from local sources, an immense accession to our knowledge of the Indian languages might be derived from such works. It is in the power of nearly every Englishman in India to make notes of such words as they come to his notice from time to time. The peculiar value of such words consists in their generally archaic form. Many, nay, it may be safely said, *all* the words found in Sanskrit authors are still in existence in India, though often in very mutilated shapes; and if the great task of tracing the origin of the modern languages is ever to be satisfactorily accomplished, if the veil of darkness which hides the intermediate space between Sanskrit and the speech of to-day is ever to be removed, it must be done, I am convinced, through the medium of the rural dialects; and that, not only by investigating those dialects which present great peculiarities, but those also which are very little removed from the accepted standard of the language. There is no part of India where new words and senses may not be picked up, and it will often be found that quite unexpectedly problems of the most difficult nature will receive a solution from the neglected and despised expressions current in the mouths of the lower classes.



CHAPTER VI.

HINTS ON OBSERVING AND RECORDING A NEW LANGUAGE.

FROM the remarks made in the last chapter but one, it results that in acquiring a new language two things have to be considered: first, the words or vocabulary; secondly, the inflections or grammar.

I place the vocabulary before the grammar for this reason, that in modern languages inflections do not at first sight present so ready a means of comparison as words. In the Indo-Germanic family the inflections, though, in the main, originally identical in all languages, have by the lapse of time become so abraded, softened, hardened, dropped, inverted, and otherwise changed, that the recognition of them is a task which often baffles the keenest scholar.

In the Turanian family, on the other hand, inflections were, as we have seen, originally independent words added to the stem-words, and their selection was led up to by different lines of thought in different languages. Thus a plural might be formed by adding words signifying "crowd," "flock," "swarm," "number." Now if one Turanian language chose to express "men" by "man + swarm," while another, expressed it by "man +



CSL

OUTLINES OF INDIAN PHILOLOGY.

crowd," and a third by "man+number," it is obvious that there could be in the inflections no identity of sound to strike the observer, or to help him in classifying the three languages.

It is therefore in the vocabulary that the data for comparative philology must, in the first instance, be looked for. If two languages can be shewn to possess a certain number of roots in common, one step is gained towards proving their connection, and an analysis of their inflections will then follow.

Again, if a certain proportion of the roots of a hitherto unexplored language can be shewn to be connected with those of some known language, while a certain proportion presents resemblances to another language, it is from an investigation of the inflections alone that we can hope, in the absence of historical data to determine its exact place in the general family. A good example of a language of this sort is the Magar language, which is Turanian at base, with a semi-Aryan vocabulary.

But it will be asked, how are we to know what words are roots and what inflections? To answer this point it will be necessary here to give some practical hints as to how to proceed in noting down a new language from the mouth of a native, and these remarks, I hope, will be found more useful than a long dissertation on the scientific dissection of languages; because, after all, if the traveller can only record with sufficient accuracy and copiousness what he hears, he will render more



service to science by handing over his note-book to the untravelled philologist to dissect, than by making it the subject of his own crude, and often erroneous, speculations.

The first thing in listening to a man who speaks a tongue unknown to you is, to be sure you catch accurately the sounds he utters. This reads like a truism, but it is in reality no easy matter to catch the blurred, semi-articulate utterances of a savage, or the elaborate half-musical cadences of the Turanian tongues. Much time and valuable learning has been wasted by philologists¹ in reasoning on the connection of languages from the basis of notes furnished by travellers who have not recorded correctly the words they heard, because their ear wanted training.

In India especially this caution is necessary, because the dialects of the Indo-Germanic family spoken here are, on the whole, so well known and so constantly used in the great political and commercial centres of the country, that it is not necessary for chance travellers to record them; manuscript and printed books, too, exist in sufficient numbers to render it unadvisable for scholars to rely upon travellers' note-books; and the English residents of a district whose help would be most valuable are not so likely to write a word erroneously as a traveller, because however bad their ear may be, they

¹ Dr. Latham's otherwise valuable book, "Elements of Comparative Philology," is cram-full of errors arising from this cause.



hear the words so often that they are not likely to misspell them.

It is in the Himalayas, on the Assam, Cachar, and Chittagong frontiers, in Chota Nagpore and Gondwana, that so much remains to be cleared up; and the languages of these places are, with few exceptions, Turanian and unwritten, and consequently full of pitfalls for the unwary.

In such a position much benefit will be derived from inducing the native to enter into a long conversation with some one who speaks his language, trying at the same time to make him talk as slowly as possible. By listening carefully to this conversation, the ear gradually gets familiarized with the tones and accents of the dialect, and even catches and retains a few constantly-recurring words. Note these words as the conversation goes on, writing them by sound. The French-vowel system will be found singularly copious and well-adapted to express Turanian vowels. The sounds of *eu* in *bleu*, *œi* in *œil*, *eui* in *feuille*, *œu* in *cœur*, *sœur* are of frequent occurrence. Thus, in Lepcha, the word *bleu*, "a small hill," is pronounced exactly like the French word *bleu*, blue.

Having got your ear into training by listening to the native talking, ask him the meaning of the words you have written down, pronouncing them to him. He will probably correct your pronunciation repeatedly long after your own ear is satisfied. The reason of this is that all Turanian languages have a delicate and subtle



system of euphony, which a foreigner does not easily acquire. A very good description of these euphonic tones will be found in Summers's Chinese Grammar, which is lucidly and simply arranged.

In English, the same word may bear many different tones, according to the sentiment which the speaker wishes to convey. In the monosyllabic languages, on the contrary, the tone is fixed and inherent in the word. Some words, for instance, must always be pronounced with an accent of surprise, as in English, "Ha!" "Halloa!" A second class has a plain even tone, such as we generally use in English in ordinary emotionless talk. The third tone is a rising tone, such as in English would imply a question, as "Who are *you*?"—"Well, what *then*?"

The fourth is a descending tone, such as we should use to express dismissal, as in the Hindi "chale jáo;" or despair, as "No! all is *lost*!"

The fifth is an abrupt stop, as in "No! certainly not!"

It is difficult for us, accustomed to regard tone and emphasis as marks of emotion, which may be applied to any word at pleasure, to understand a system in which each word carries its own tone, whatever may be the sentiment expressed, but that such is really the case may be readily seen from five minutes' talk with a Lepcha or Bootea. The peculiar sing-song accent of the Burmese, too, is due to the presence of these tones. It is necessary to bear this in mind, because the native



with whom you are conversing will persist in correcting your pronunciation till you catch the right tone, and a knowledge of what he is driving at may save the enquirer from being hopelessly bewildered, and giving up the task in despair.

When the preliminary difficulties are overcome, secure first the numerals from one to ten; next, if you find it possible, which it is not in all languages, those from eleven to a hundred; also the way in which the composite numerals twenty-one, twenty-two, etc., are formed; noting any irregularities in formation, such, for instance, as occur in the Hindi numerals, बावन *bāwan*, तिरपन *tirpan*, etc. The ways of forming numbers are very various. Some languages have only distinct words for the numbers from one to five, and form those from six to ten by saying "five + one," "five + two," and so on.

Thus in the Kambojan language we have—

one, moe	six, pram-moe <i>i.e.</i> (5+1)
two, pir	seven, pram-pil (5+2)
three, bai	eight, pram-bai (5+3)
four, buan	nine, pram-buan (5+4)
five, pram	ten day

Some languages again express twenty by "four fives" or 4×5 , instead of, as with us, "two tens" or 2×10 . In the composition of the higher numbers many irregularities exist. This will be readily understood when it is remembered that so close at home as France the old simple words "septante," "octante," "nonante," have been discarded for the cumbersome paraphrases "soixante-



dix" or $60+10$; "quatre-vingts" or 4×20 ; "quatre-vingt-dix" or $4 \times 20+10$; so that as a climax we get for ninety-nine, instead of, as in English, $9 \times 10 + 9$, "quatre-vingt-dix-neuf" or $4 \times 20 + 10 + 9$.

Next secure the personal pronouns I, thou, he, etc., also their cases (if they have any) of me, to me, etc.

The reason why these should be looked after so early is because, if you should be prevented from recording more than this, you will at any rate have secured those portions of the language which are, *a priori*, most likely to afford resemblances and connections with other languages.

Having got the numerals and pronouns, inquire the names of as large a number of common objects as your time allows; also names of relationships, parts of the body, and so collect a vocabulary of nouns. This will not be difficult. In the case of verbs, however, there is this difficulty, that in most Indo-Germanic languages the verb is quoted under its infinitive form in our dictionaries. Thus, in looking out a word we should look for *marná* not *martá*, *giriftan*, not *gíram*, *aimer*, not *j'aime*, and so on. Now to the savage, whose language is unwritten and uncultivated, grammatical distinctions are unknown, and it would be useless to ask him any questions about them. If you use the most familiar tense in Hindústání, the imperative, and ask, say, a Kole the equivalent of "bolo," he will tell you "abbenkakaji" or "kaji'ben," from which it would be difficult to extract



the fact that the word "kajitea" means to speak, and that all the forms of the verb are derived from the root "kaji." Any attempt to make a Kole understand what you want to know in this respect, would only result in intense bewilderment. The best way is to set to work with a lot of sentences in which the verb occurs in a variety of forms; a subsequent analysis of these sentences will give the result sought for. In fact, this is almost the only course that it is possible to pursue with savages. The system on which sentences are formed will also be learnt in this way with great ease. And from this will be derived a facility of extracting the root from any number of forms, while the sentences themselves will afford most, if not all, of the forms of the other parts of speech.

The changes which occur in the Turanian languages from euphonic causes are very numerous and perplexing; the chief rule is that the vowels in a word follow the first vowel, thus in Turkish the termination "unji" becomes "inji" when added to a word containing the vowel *i*, as "birinji," first; and in Hungarian vowels are divided into two classes, and vowels of two different classes may not occur in the same word.

Languages of the Turanian family are particularly rich in participles, and express by this means many sentences which in other languages would have to be expressed by relative and other pronouns.

With regard to languages of the Aryan family, all



that is necessary is to find out to which of the six or seven principal languages of the Indic class they are to be affiliated. With a grammar of this principal language in one's hand, it will be easy to note the divergences from the classical standard of the language, which is about all that remains to be investigated.¹

To complete the record of a language it is necessary to note the exact locality in which it is spoken, and the area over which it extends; also anything that can be gathered as to where the tribes originally came from, and whether it is in any way connected with neighbouring tribes.

It is also important to observe the habitual style of pronunciation and accent; whether short, sharp, and chippy like Hindústání, or soft, mumbling, and drawling like Panjabi: or broad and requiring much muscular action of cheeks and jaws like Bengali. The habit of pronunciation peculiar to a people often accounts for dialectic variations, such as that which has led the Panjabi to say *kahdá*, *kahndá*, and *kahn'á*, where the Hindústání says *kahtá* and the Bengali *kahitechhe*.

In India it is generally found necessary to make use of an interpreter in collecting a vocabulary. And this interpreter is usually a man who speaks the language in question and Hindústání or Bengali. In proposing sentences therefore to a native through the interpreter you run the risk of having the arrangement of the

¹ But see the chapter on Dialects, *passim*.



sentence affected, or even the meaning altered by passing through the medium of Hindústání. Especially is this the case with regard to the order of words in a sentence; the Hindústání order being generally the reverse of the English and other Indo-Germanic tongues, and a very artificial order too; there is generally good reason to suppose, when you find that a sentence of the language you are studying agrees in the order of its words with the Hindústání, that this is not the genuine order of the words as they would naturally be arranged by the savage; but that it has been altered by its transmission through Hindústání. The only way to detect this is to have the sentence or similar sentences repeated over and over again, when you will probably get at the right order,—that is to say, the way in which the savage would naturally arrange his words, especially in Turanian dialects.

In conclusion it seems advisable to say a few words on the vexed question of transliteration,—that is to say, on the way in which native words should be spelt in English characters. The system invented by Sir W. Jones and improved by the illustrious scholar, Professor H. H. Wilson, is that which is now generally used, and I have no hesitation in recommending it.¹ It will, how-

¹ The paper on this subject, published by the Asiatic Society of Bengal, has been carelessly drawn up, and contains several errors; among others the glaring one of placing the Persian خ as in use in Urdu on the same line as ख as though they were equivalent, while ک the real equivalent of ख is put on a line by itself; also representing थ as the same sound as the English th in 'think.'



ever, be of no use in transcribing languages of the Turanian family, and I know of no good system for such languages. The missionary alphabet of Lepsius is intricate, and inadequate to express vowel sounds. The best way to act is to give at the commencement of your notes a statement of the method you adopt, referring to known sounds either in English, French, or some other familiarly-known language, and, above all things, having once fixed upon your system and given the clue to it, be careful to adhere to it yourself throughout your notes, otherwise endless confusion and mistakes will arise.¹

¹ In the supplement to Transactions of the Philological Society for 1867, Part I., will be found an essay by Mr. Ellis, on what he calls "Palæotype." He professes to have invented a system of transliterating foreign languages into English, which is universally applicable. His system is one of formidable complexity, but well repays attentive study, being evidently the result of much thought, aided by a delicate ear, and extensive knowledge of languages.



CSL

APPENDIX A.

I give here the first ten numerals of nearly all the Indian languages for purposes of comparison, and to show the connection which exists between the various languages, families, and dialects.

The numerals are selected as being those parts of speech which retain their forms with the greatest tenacity, and offer the most obvious similarities.



I. INDO-GERMANIC FAMILY.—I. INDIC CLASS.

CSL

English.	Hindi.	Bengali.	Panjabi.	Sindhi.	Marathi.	Gujarati.
one	एक ek	अक ek	हिक hik	हिकु hiku	एक eka	एक ek
two	दो do	दुइ dui	दो do	ब bba	दोन don	बे be
three	तीन tīn	तिन tin	तिन tin	टे te	तीन tīn	तण taṇ
four	चार char	चारि chāri	चार chār	चारि chāri	चार chār	चार chār
five	पांच pānch	पाँच pānch	पंज panj	पंज panja	पांच pānch	पांच pānch
six	छः chhah	छয় choy	छे chhe	छ chha	षष् ṣhaṣh	छ chha
seven	सात sāt	সাত sāt	सत् sat	सत sata	सात sāt	सात sāt
eight	आठ āṭh	আঠে āṭh	अठ aṭh	अठ aṭha	आठ āṭh	आठ āṭh
nine	नौ nau	নয় noy	नाँ nauṇ	नँवँ nauṇvaṇ	नौ nau	नव nav
ten	दस das	দশ daś	दस् das	डह dāha	दस das	दश daś

I have put these lists into Devanāgarī type to facilitate comparison, though several of the languages use a special character of their own.



INDIC CLASS—continued.

CSL

78

OUTLINES OF INDIAN PHILOLOGY.

English.	Uriya.	Singhalese.	Kashmiri.	Sanskrit.	Zend.
one	eko	eka	akh	एक eka	aêva
two	dui	deka	zuh	द्वि dwi	dva
three	tini	tuna	trae	त्रि tri	thri
four	chári	hatara	tsór	चतुर् chatur	chathwâre
five	páncho	paha	pant	पञ्चन् panchan	pancha
six	chho	haya	sheh	षष् shash	khshwas
seven	sháto	hata	sat	सप्तन् saptan	hapta
eight	átho	afa	aát	अष्टौ ashtau	asta
nine	no-o	namaya	naon	नवन् navan	nava
ten	dosho	dahaya	dah	दशन् daśan	daça



INDO-GERMANIC FAMILY.—2. IRANIC CLASS.¹

CSL

OUTLINES OF INDIAN PHILOLOGY.

English.	Persian.	Kurdish.		Pushtu.	Armenian.	Ossetian.
		Kurmanji.	Zazo.			
one	یک yak	yek	yau	یو yo	mek մեկ	iw
two	دو du	du	du	دوه dwah	erknq երկուք	dua
three	سه sih	sisé	hiryé	دری dre	eri երի	arta
four	چهار chahár	chár	tehér	چلور tsalor	chorq չորք	tzupar
five	پنج panj	panj	panj	پنجه pindza	hinq հինգ	fonj
six	شش shash	shash	shesh	شپږ shpag	vez վեց	achsaz
seven	هفت haft	haft	haut	اوه awa	evthn եւեթ	awd
eight	هشت hasht	hesht	hesht	اته ata	out ութ	ast
nine	نه nuh	nuh	nau	نه nah	inn հին	farast
ten	ده dah	dah	des	لس las	dasn տասն	das

¹ Although none of these are Indian languages, yet I give them to complete a general view of the Indo-Germanic family in its eastern division.



III. TURANIAN FAMILY.—2. HIMALAYIC CLASS.¹

English.	Tibetan. ²	1. Bhotia.				2. Lepcha.
		Kunawar.	Lhopa.	Takpa.	Sumchu.	
one	<i>gchik</i>	chik	che	the	it	kat
two	<i>gnyis</i>	ni	nye	nai	nish	nyet
three	<i>gsum</i>	sum	sum	sum	hum	sam
four	<i>bji (zhi)</i>	zhi	zhi	pli	pu	phalat
five	<i>lnga</i>	nga	nga	liangi	gna	phanyong
six	<i>drug (thu)</i>	tuk	dhu	kro	tuk	truk
seven	<i>bhun</i>	dun	dun	nis	shunish	kakyok
eight	<i>brgyad (gye)</i>	gye	gye	gyet	ket	kaku
nine	<i>dgu</i>	gu	gu	dugu	gu	kakyot
ten	<i>bchu thamba</i>	chathamba	chatham	paki	sa	kati

¹ The numbers affixed to the names of each language refer to the general classification in Chapter I.

² The letters in italics are not pronounced in the modern language.



HIMALAYIC CLASS—continued.

CSL

English.	3. Limbu.	4. Kiranti. ¹	5. Murmi.	6. Gurung.	7. Newar.	8. Magar.
one	tik	kwong	kik	ku	chi	kat
two	nechi	niksi	gni	ni	ni	nis
three	somchi	sam	som	song	sun	som
four	lishi	lé	pli	pli	pi	buli
five	gnáshi	ngo	gna	gna	gna	banga
six	tukshi	rukba	dhu	tu	khu	(the rest of the numerals are borrowed from Hindi.)
seven	nashi	channi	nis	nis	nhe	
eight	etshi	yá	pre	pre	chya	
nine	phangshi	ghu	kuh	kuh	gunh	
ten	thibong	kwaddyum	kan	chuk	sanho	

¹ Only one of the numerous Kiranti dialects is here given; viz., Báhing.



HIMALAYIC CLASS—continued.

English.	9. Bramhu.	10. Chepang.	11. Vayu.	12 Kusunda.	13. Sunwar.	14. Sarpa.
one	de	ya	kolu	goisang	ka	chik
two	ni	nhi	nayung	ghigna	nishi	nyi
three	swom	sum	chuyung	daha	sang	sum
four	bi	ploi	bining	pinjang	le	zhyi
five	banga	puma	(The rest are wanting.)	(The rest are wanting.)	gno	gná
six	(The rest are wanting.)	kruk			ruk	tuk
seven		cháná			chani	dyun
eight		prap			yok	gye
nine		taku			guh	guh
ten		gyib			sashi	chúh



HIMALAYIC CLASS—continued.

CSL

English.	15. Milchán.	16. Tirbarskad.	17.		18. Dorhi.	19. Denwar.
			Hundesí.	Márchhá. ¹		
one	it	ti	chig	tik	(The numerals are taken entirely from Hindi.) ²	(Hindi as in Dorhi.)
two	nish	nishi	ní	nis		
three	sum	sum	sum	sum		
four	pu	pi	zhi	pi		
five	nga	ngai	ngá	ngai		
six	tuk	tuki	dug	(Rest taken from Hindi.)		
seven	stish	nashi	dun			
eight	rai	gyai	ged			
nine	sgoi	goi	gu			
ten	sai	choi	chyú			

¹ Marchha is a sort of Hundesi gibberish used by the Bhotia traders among themselves.

² Nos. 18, 19, 21, and 22 are spoken in the Terai at the foot of the hills, and in consequence of their position are being gradually encroached on by Hindi.



HIMALAYIC CLASS—continued.

English.	20. Pahri.	21. Yakha. ¹	22. Kulungya. ¹	23. Thaksya. ²
one	chi	ikko	ubám	di
two	ni	kichh	nih	ngi
three	sung	sum	sup	som
four	pi	li	li	bla
five	ngo	ngá	nga	nga
six	khú	tuk	tuk	tu
seven	nhe	nu	nu	nges
eight	chya	phang	re	bhre
nine	gun	yech	bong	ku
ten	gi	ibong	ukbong	chyu

¹ These are Kiranti dialects. They add *chi* to all numerals, as Sum *chi*, Li *chi*, etc.
² Kaswar and Pakhya have adopted Hindu numerals.



III. TURANIAN FAMILY.— 3. LOHTIC CLASS.

CSL

English.	1. Burmese.	2. Dhimal.	3. Mечи.	5. Caro.	6. Kachari.	7. Abor.
one	tit	e	che ¹	sha	chie	ako
two	hnit	nhe	nye	gining	nai	ani
three	thong	sum	tum	gatham	tham	amgon
four	lay	dia	bre	bri	bre	ápi
five	nga	na	bha	bongá	ba	pilángo
six	chouk	tu	dho	brok(d)	ro	akye
seven	ko-hnit	nhi	chini	sini	sni	konange
eight	chit	ye	jokanii	chet	jat	pini
nine	ko	kuhá	(8 and 9 are wanting.)	skhu	chku	kinide
ten	ta-sai	te		skang	ji	iinge

¹ No numerals beyond eight. The syllable *man* is prefixed to all numerals, as *man-che*, *man-nye*, etc.



LOHITIC CLASS—continued.

English:	9. Miri.	10. Dofla.	11. Khasía.	12. Mikir.	13. Naga.	
					Angámi.	Rengmá.
one	ako	aken	uwai	nisi	po	katang
two	aniko	ani	ár	hini	kana	anna
three	aumko	an	lai	kithom	se	asam
four	apiko	apli	saú	phili	da	pazr
five	angoko	ango	san	phanga	pengu	pungu
six	akengko	akple	hinreú	therok	shuru	tank
seven	kinitko	kanag	hinieú	theroski	thena	tanet
eight	piniko	plagnag	prah	nerkep	thata	te
nine	konangko	kayo	khâindai	sérkep	teku	taku
ten	uyinko	rang	shipeú	kep	ker	tar



LOHITIC CLASS—continued.

CSL

English.	14. Singpho.	23. Kakhyen.	25.		
			Sgau Karén.	Pwo Karén.	Bghai Karén.
one	dima	nge	ta	lá	ta
two	nkhong	onkong	kee	nee	ki
three	masum	mesong	theu	theru	theu
four	meli	meli	lwee	lee	lwi
five	manga	menga	yai	yai	yay
six	kru	kan	ku	ku	theu to 34
seven	sinit	senit	nwi	nwi	theu to ta 34 2
eight	macat	matsat	khon	kho	lwi to 41
nine	tseku	tiekho	kwi	kwi	lwi to ta
ten	si	shi	tachi	lachi	tashi



4. KOL CLASS.

English.	1. Sonthal.	5. Ho.	6. Khond.		7. Gond.	
			Yerukali.	Sourada.	Kuar.	Northern.
one	midh	miad	vondu ¹	aboy	mia	andi
two	barria	baria	rendu	bagu	bária	rand
three	appia	appia	mume	yagi	appi	mand
four	ponia	upunia	nalu	vuji	áplean	nálo
five	mone-gotang	moya	anju	mollayi	manke	saiyan
six	turin-gotang	turuia	aru	kudru	terame	sáran
seven	lair-gotang	âya	yegu	gulgi	aiya	ero
eight	iral-gotang	íriia	gethu	tamuji	ilhár	ármár
nine	are-gotang	area	ombadu	tinji	arhe	armáh
ten	gel-gotang	gel	pothu	galliji	gail	path

¹ Note the similarity between this and the Dravidian, especially Telugu, the most northern of that class of languages:



DRavidIAN CLASS.

English.	1. Telugu.	2. Tamil.	3. Karnátaka.	4. Malayalam.	5. Tulu.	6. Kudugu.
one	okati	onđu	ondu	onna	onji	ondu
two	rendu	irandu	erađu	rańta	erad	dandu
three	mđu	mđu	mđu	mđu	muji	mđu
four	nálu	nálu	nálu	nála	nalu	nálu
five	ayidu	aindu	aidu	ancha	ayinu	anji
six	áru	áđu	áru	ára	aji	áru
seven	yđu	éđu	éđu	éla	el	éđu
eight	yenimidi	eđu	eđu	eđu	evame	eđu
nine	tommidi	ońbadu	ombhattu	onpata	ovambu	oyimbadu
ten	padi	pattu	hattu	patta	paltu	pattu

For these lists I am indebted to Latham, Caldwell, Dr. Campbell and others. The Burmese and Sran and Pwo Karen numerals were kindly supplied to me by the Rev. B. C. Thomas, Karen Missionary, to whom I am also indebted for some valuable information on the subject of the Karen and Burmese languages generally. The Karen lists are from Leitch's work, and most of the Himalayan from B. Hodgson. The Indo-Germanic lists are derived from personal knowledge, or from the many well-known works on those subjects. I have taken every possible pains to attest each language and authenticate it; but I must confess that in a few instances only I have had to take my words on trust from Latham, who has borrowed them in his turn from some obscure source. This unfortunately I could not avoid, but such I am happy to say, are few in number.





CSL

APPENDIX B.

I subjoin a list of the personal pronouns in a few of the leading Turanian dialects. The pronouns, like the numerals, possess wonderful tenacity, and offer one of the best means of determining the class to which a language should be referred.

Turkish. ¹	Tibetan. ²	Tamil. ³
I ben	nga	nân
thou sen	khyod	ni
he } ol		{ avan
she } ol	kho	{ aval
it } ol		{ adii
we biz	ngachag	nâm
ye siz	khyodchag	nîr
they anlar	khochag	avar

Gond. ⁴	Brahui. ⁵	Milchan. ⁶	Magar. ⁷
I nunnâ	I	goo	nga-ngoï-ngoichü
thou ima	ni	ki	nang
he } wur			
she } wur	da	no	hose
it } wur			
we amat (mâr)	nan	kishang	kan
ye imat (imâr)	num	kina	nahakei
they wurg	dafk	nogonda	hosko

¹ This is given as a specimen of the Turanian family, Turkic class.

² ⁶ and ⁷ Turanian family, Himalayan class.

³ and ⁵ Turanian family, Dravidian class.

⁴ Kol class.



CSL

APPENDIX C.

ABSTRACT OF INFORMATION ON THE LANGUAGES OF
ASSAM.*I.—District of Kámrúp, capital Gáoháti.*

The Kachári, Mikir, Gáro, and Lallong¹ tribes speak their own languages. The rest of the population Assamese.

Kachári is spoken in the village of Bácsá and Gorkhola, 44 miles north of the capital Gáoháti; also in Dekeli, Jhargánw, Betta, and Chapáshuri, 30 miles from the capital. In Beltola, four miles S. of Gáoháti Mikir, Kassyá and Kachári are spoken. In Lukí, Chhaiganw, Pantán, Baradwar, Boko, Bangánw, and Moirapur, 44 miles S.W. from Gáoháti, the Kassyá and Kachari are used. In Danrúa and Panbária, 20 miles S.E., the Lallong and Mikir are spoken.

II.—District of Darrang, capital Tezpúr.

The Kachári, Miri, Dofla, and Bhotanta are spoken in this district.

Kachári between 26° 15' and 27° N. Lat.; 91° 45' and 92° 45' E. Long. Miri and Dofla from 92° 45' to 93° 45' E. Long., and 26° 45' to 27° 5' N. Lat. Bho-

¹ Probably a tribe of Kacharis.



tanta by only a few stragglers between $26^{\circ} 40'$ and $26^{\circ} 50'$ N. Lat., and $91^{\circ} 45'$ to $92^{\circ} 15'$ E. Long. The country inhabited by these tribes is accessible to a limited extent.

Besides the above, there is a small tribe of Akas and one of Miris. The Miri settlers came here from the Abor hills, north of the Lakhimpúr district. The Doflas extend as far as $94^{\circ} 30'$ E. Long.

III.—Naga Hills, capital Golaghát.

Naga, Kúki, Kachari, and Mikir are the languages of this district.

Three tribes of Nagas are found here—the Rengmá, Angámi, and Lotah.

The Rengmá Nagas inhabit the eastern face of that range of hills, which runs almost parallel to the course of the Dhansiri from the Brahmaputra river to the source of the Jamuná river. Some of their villages are only two days' journey west from Golaghát.

The Angámi Nágas occupy the country between the Dhansiri river (the eastern most branch of the Dihong river) and Manipur. Their nearest village is four days' journey S.E. from Golaghát.

The Lotah Nagas extend under different names from the Dansiri to the Desoie river eastward, having the district of Sibságar on the north and wilder Naga tribes on the south. Their most distant village could be reached from Golaghát in five or six days.



The Kukis live on the upper waters of the most western branch of the Dhansiri, seven or eight days journey from Golaghát.

The Mikirs and Kacharis are on the upper course of the Jamuná and in the hills where that river rises. They are distant five or six days' journey from Golaghát and the same from Naogánw.

IV.—Kassya and Jaintia Hills, capital Shillong.

Kassya and Mikir are the languages of this district. The Jaintias speak a patois of Kassya. Mikir is spoken along the northern frontier of the district.

V.—Naogánw (Nowgong), capital Naogánw.

Mikir, Garo, Kassya, Kúki, Kachari, Miri, Lalang, and Naga (Rengma chiefly) are spoken by the inhabitants of this district. "The Naga range of hills extends from Muhandijúa—a place about 46 miles south-east of Naogánw,—and runs in an easterly direction towards Sibságar and Cachar." The various tribes appear to be much mixed up together throughout the district.

VI. Sibságar, capital Sibpúr.

Only Assamese and Bengali spoken in this district.

No replies were received from the remaining districts, viz., Lakhimpúr, Gwálpára, and Dibrúgarh. This is the more to be regretted, as from the situation of the



first and last of these districts, it is probable that much valuable information might have been supplied by the officers in charge. It is unfortunate that the study of languages is not more popular among our countrymen in India; though it is perhaps scarcely fair to expect much from so hard-worked a class of men as Indian officials, especially in the non-regulation provinces.



CSL

ADDENDUM.

AFTER the foregoing pages had been printed, I discovered that I had accidentally omitted all mention of the population of the Dravidian languages. It being too late to insert it in the text, I give here the following extracts. The proportionate numbers of the several races by whom the languages and dialects mentioned above are spoken appear to be as follows:—

1. Tamil	-	-	-	-	-	10,000,000
2. Telugu	-	-	-	-	-	14,000,000
3. Canarese	-	-	-	-	-	5,000,000
4. Malayalam	-	-	-	-	-	2,500,000
5. Tulu	-	-	-	-	-	150,000
6. Tuda	}	-	-	-	-	500,000
7. Kota						
8. Gond						
Total	-	-	-	-	-	<u>32,150,000</u>

According to this estimate, the Dravidian race numbers upwards of thirty-two millions of souls. There cannot be any doubt of their number amounting to, at least, thirty millions; of whom about twenty millions are British subjects, and the remainder belong to the native states of Hyderabad, Nagpore, Mysore, Travancore, and Cochin; the gross population of which, including all races, is estimated at twenty millions, but is probably much greater. In this enumeration of the



CSL

OUTLINES OF INDIAN PHILOLOGY.

Dravidian languages, I have not included the idioms of the Ramûsis, the Korawars, the Lambâdis, the Vedars, the Male-arasars, and various other wandering, predatory tribes. The Lambâdis, the gipsies of the Peninsula, speak a dialect of the Hindustani; the Ramûsis, and the majority of the Korawars, a patois of the Telugu; the tribes inhabiting the hills and forests, corrupted dialects of the languages of the contiguous plains. None of these dialects are found to differ essentially from the speech of the more cultivated classes residing in the same neighbourhood. The Male-arasars, "hill-kings" (in Malayâlam, *Mala-araans*), the hill-tribe inhabiting the southern ghauts, speak corrupt Malayâlam in the northern part of the range where the Malayâlam is the prevailing language, and corrupt Tamil in the southern, in the vicinity of Tamil-speaking districts. (*Caldwell's Comparative Grammar of the Dravidian or South-Indian family of languages*, p. 9).

This refers to Chapter II. p. 18.