India as a Lingustic Area
Author(s): M. B. Emeneau
Source: *Language*, Vol. 32, No. 1, (Jan. - Mar., 1956), pp. 3-16
Published by: Linguistic Society of America
Stable URL: http://www.jstor.org/stable/410649
Accessed: 10/05/2008 10:03

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=lsa.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We enable the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

http://www.jstor.org

INDIA AS A LINGUISTIC AREA

M. B. EMENEAU

University of California, Berkeley

The American anthropologists who have been linguistic scholars as well—I would mention Boas, Sapir, and, last but not least, Alfred L. Kroeber, whom we delight to honor in this issue of LANGUAGE—have been catholic in their approaches to linguistics. Descriptive linguistics on this continent owes a tremendous amount to these men. But none of them has ignored historical problems, and in their various ways and even with radically different points of view on subjects which were open to dispute, they have contributed much to both detailed genetic problems, especially of the North American continent, and to the discussion of certain general questions. One of the latter is a problem that arises again and again in any region of the world where the linguistic picture is complicated, and it is particularly fitting, as will appear, if I attempt to add something to it as an offering to Kroeber.

The problem is that of diffusion of linguistic traits across genetic boundaries. Boas raised the problem (whether it was original with him, does not matter—for I do not intend to be bibliographically complete) in several places, including especially the introduction to *Handbook of American Indian languages* 1.47–53 (BAE, Bull. 40, Part 1; Washington, 1911); *American anthropologist* 22.367–76 (1920); and *Lg.* 5.1–7 (1929). One of his conclusions was that 'in a considerable number of native languages of the North Pacific Coast [of North America] we find, notwithstanding fundamental differences in structure and vocabulary, similarities in particular grammatical features distributed in such a way that neighboring languages show striking similarities. ... It seems ... almost impossible to explain this phenomenon without assuming the diffusion of grammatical processes over contiguous areas' (*Lg.* 5.6). In the preceding exposition he had taken it to be a matter of general agreement (as it patently is) that words may be borrowed, and probably also that phonetic traits may be borrowed. He was particularly concerned to demonstrate that morphology also may diffuse, and he brought forward a rather considerable number of instances that seem convincing. This for us is the important point of his treatment. We may find it more difficult to accept the phraseology of his general theory ('hybridization of languages'), as being a little too simple, too unsubtle, and as ignoring (as we know he did) genetic relationships that were not clear at first inspection; but at the moment we are not concerned with this.

Sapir treated the problem especially in Chapter 9 of his book *Language* 205–20 (1921). He ranged much more widely for his examples than Boas had done, drawing in material from his immensely broad linguistic experience. He, like Boas, accepted the borrowing of words as commonplace, though he pointed out and discussed the varying tolerances of languages for such borrowings. He accepted also the borrowing of phonetic traits; his discussion introduced the bilingual individuals who are the social carriers of change. More important as pre-

figuring what his attitude would be when he took up morphology, was his insistence that phonetic borrowings are not random but are regulated (as by a sort of 'governor') by 'the phonetic drift' of the language (214): 'so long as its main phonetic concern is the preservation of its sound patterning, not of its sounds as such, there is really no reason why a language may not unconsciously assimilate foreign sounds that have succeeded in worming their way into its gamut of individual variations, provided always that these new variations (or reinforced old variations) are in the direction of the native drift.' This is essentially a reluctance to accept such borrowings without qualification; whether the qualification is always demonstrable in Sapir's terms, will depend on the historical data available (and of course on the ingenuity of the scholar). The method is surely that of our present-day structuralists. In the final section of the chapter (215–20) Sapir examined morphological borrowings. His argument ran: English has borrowed a certain number of affixes from French, Latin, Greek (*-ess*, *-ize*, *-able*), but these are merely additions to the old stock of affixes and are hardly different from the borrowings of words. They make no difference 'to the essential build of the language'. To generalize: 'nowhere [i.e. in no language] do we find any but superficial morphological interinfluencings.' He summed up: 'We may infer one of several things from this:—That a really serious morphological influence is not, perhaps, impossible, but that its operation is so slow that it has hardly ever had the chance to incorporate itself in the relatively small portion of linguistic history that lies open to inspection; or that there are certain favorable conditions that make for profound morphological disturbances from without, say a peculiar instability of linguistic type or an unusual degree of cultural contact, conditions that do not happen to be realized in our documentary material; or, finally, that we have not the right to assume that a language may easily exert a remolding morphological influence on another.' In some cases of morphological similarities Sapir pointed out that they are vestiges of genetic relationship, and he was willing (as he showed elsewhere in setting up the superstocks for North America) to use this solution rather freely. Finally, he frankly said of diffusion that we have 'no really convincing examples of profound morphological influence by diffusion', and he characterized language as 'probably the most self-contained, the most massively resistant of all social phenomena'. The well-known conflicting attitudes of Boas and Sapir are clearly at work here—Sapir's 1921 statement is reaction to Boas' skeptical attitude toward genetic relationships; Boas' 1929 article is his rebuttal of Sapir, even though only tacitly so. It is to be emphasized that Sapir makes a distinction between two types of morphological influence, one 'superficial', the other 'profound'. This is, in spite of its specious attractiveness, a highly subjective differentiation and one that it will be very difficult to apply in specific cases; it will undoubtedly give rise to disagreement between scholars who handle the same data, and it may in the long run not be at all a usable distinction, just since it is a value judgment and not quantifiable.

Sapir's attitude has come to be widely held in this country. Kroeber in his presidential address to the Linguistic Society in 1940 (*Lg.* 17.287–91 [1941]) calls it a 'usual dictum' that 'words can be borrowed freely between distinct

languages, but grammar with difficulty if at all.' Hoijer (*Lg.* 24.335 [1948]) says: 'Traits of language are not readily borrowed.' A Boasian point of view has been and is, however, apparently more favored in Europe (cf. Sebeok, *Lingua* 2.136 ff. [1949–50], and the bibliography given by him; N. S. Trubetzkoy, *Acta linguistica*, 1.81–9 [1939]; Roman Jakobson, *IJAL* 10.193 [1944]; etc.). It has, moreover, never been totally abandoned here. Leonard Bloomfield in his book *Language* 468–71 (1933) certainly favored it. Kroeber (loc.cit. 290) thought that 'the time has come to reexamine this [i.e. Sapir's] dictum', led to this statement by Ray's hypothesis that the Melanesians have borrowed from Malayo-Polynesian more formal structure than vocabulary. Whatever may be the correct solution for this last specific problem, it would seem that more evidence bearing on the general problem would be welcome. Probably the most useful approach to a solution would be the provision of material of a kind that would allow historical examination with a considerable time-depth. India provides such material. Some of it has been examined from this point of view by earlier scholars, but little of it has penetrated into the realm of general linguistics, since it was published in specifically Indological outlets.[1] Perhaps this reexamination will have a better fate.

The Indian subcontinent[2] is inhabited by a very large population who speak languages belonging to three major families, Indo-Aryan (a subfamily of Indo-European), Dravidian, and Munda. Indo-Aryan speakers in 1931 numbered about 255 million; by 1951 they must have numbered nearly 330 million. Dravidian speakers in 1951 numbered approximately 90 million. Munda speakers must number well over 5 million. This does not take account of all the languages that are included geographically in this area. There are Burushaski in Gilgit, Khasi in the hills of Assam, Nicobarese, Andamanese, and many languages of the Tibeto-Burmese group in the Himalayas and in Assam. Our attention will be focussed primarily on Indo-Aryan, Dravidian, and Munda.

For Indo-Aryan and Dravidian our historical knowledge is considerable. Sanskrit speech, that important member of the Indo-European family, has been in India and recorded voluminously since the second millennium B.C. Middle Indo-Aryan in a number of varieties is well known, and Modern Indo-Aryan, in its dozen or more major languages and innumerable local dialects, is also fairly well known. Of the four literary Dravidian languages, Tamil has voluminous records dating back at least two millennia. These four languages are well known, and the remaining fifteen or sixteen nonliterary Dravidian languages are on record to some extent. The dozen or so Munda languages are on the whole not well

[1] The first phonetic trait that I shall examine below is quoted, e.g. by Bloomfield 469 f., relying on *Linguistic survey of India* (hereafter *LSI*) 4.278 ff. Of the two other suggested influences mentioned by Bloomfield, the first, that the confusion of Indo-European *l* and *r* in Sanskrit is perhaps due to a substratum language which possessed only one of these sounds, has nothing to recommend it, since we know of no language in India with this characteristic. The remaining one will be discussed below; it concerns the use of distinct singular and plural stems to which the same case endings are added.

[2] 'India' and 'Indian' will be used in what follows for the subcontinent, ignoring the political division into the Republic of India and Pakistan, and, when necessary, including Ceylon also.

known—they are not literary languages—but for a few of them there are preliminary accounts that tell us a considerable amount.

The historical relationships between the three families are largely a matter of reconstruction.[3] It is clear from the geographical nature of the boundaries between the three families in Central India that the northern boundary of Dravidian is and has been for a long time retreating south before the expansion of Indo-Aryan, and that the small islands of Dravidian speech north of the main boundary are isolated patches that have not yet become extinct. Similarly with the Munda languages; they are all islands of greater or less extent surrounded by and pressed upon by Dravidian or by Indo-Aryan. This should mean a much greater spread both for Munda and for Dravidian at an earlier period. We know in fact from the study of the non-Indo-European element in the Sanskrit lexicon that at the time of the earliest Sanskrit records, the Ṛgveda, when Sanskrit speakers were localized no further east than the Panjab, there were already a few Dravidian words current in Sanskrit. This involves a localization of Dravidian speech in this area no later than three millennia ago. It also of course means much bilingualism and gradual abandonment of Dravidian speech in favor of Indo-Aryan over a long period and a great area—a process for which we have only the most meager of evidence in detail.[4] Similar relationships must have existed between Indo-Aryan and Munda and between Dravidian and Munda, but it is still almost impossible to be sure of either of these in detail.

The question of vocabulary borrowings between the three families need not be more than mentioned. The Dravidian languages all have many Indo-Aryan items, borrowed at all periods from Sanskrit, Middle Indo-Aryan, and Modern Indo-Aryan. The Munda languages likewise have much Indo-Aryan material, chiefly, so far as we know now, borrowed from Modern Indo-Aryan, though this of course includes items that are Sanskrit in form, since Modern Indo-Aryan borrows from Sanskrit very considerably. That Indo-Aryan has borrowed from Dravidian has also become clear; T. Burrow, *The Sanskrit language* 379–88 (1955), gives a sampling and a statement of the chronology involved.[5] It is noteworthy that this influence was spent by the end of the pre-Christian era, a precious indication for the linguistic history of North India: Dravidian speech must have practically ceased to exist in the Ganges valley by this period. Borrowings from Munda into the other two families must have taken place, but are difficult to identify (Burrow 377–9).

Other features than vocabulary items are of more interest in this connection. It has long been recognized that even our earliest Sanskrit texts show features

[3] For an attempt at reconstruction, see my article Linguistic prehistory of India, *Proc. Amer. Philosophical Society* 98.282–92 (1954).

[4] This is the historical process to be invoked, rather than the too facile and unrealistic one of a general displacement of populations through expansions. Undoubtedly there were expansions involved, in the shape of marauding bands and of missionaries, but neither of these agencies had an interest in getting rid of earlier populations; it was to their advantage, political, economic, religious, to have subjects and proselytes. Absorption, not displacement, is the chief mechanism in radical language changes of the kind we are considering.

[5] See also my article referred to in note 3 for a sketch of the history of this part of Indic scholarship and a bibliography.

that historically are un-Indo-European in their nature, but that resemble features of the Dravidian languages, and that as time went on, more such features appeared in Indo-Aryan. The late Jules Bloch collected and discussed all such points that he knew in the concluding chapter of his book *L'indo-aryen du Veda aux temps modernes* 321–31 (1934).[6] I shall discuss these and can add several more and more detail. Three general tendencies emerge: either specifically an 'Indianization' of Indo-Aryan, or, in a few instances, the appearance of a trait in contiguous languages (but not all the languages) of all three major families, without the possibility of one's being sure where it originated, or a similar situation to this last, but with evidence for the original source.

Most of the languages of India, of no matter which major family, have a set of retroflex, cerebral, or domal consonants in contrast with dentals. The retroflexes include stops and nasal certainly, also in some languages sibilants, lateral, tremulant, and even others. Indo-Aryan, Dravidian, Munda, and even the far northern Burushaski, form a practically solid bloc characterized by this phonological feature; since, however, one of the Munda languages, viz. So·ra·, which there is a good chance is archaic, does not have it, it is at least possible that it is not Proto-Munda (so Burrow 95). Even our earliest Sanskrit records already show phonemes of this class, which are, on the whole, unknown elsewhere in the Indo-European field, and which are certainly not Proto-Indo-European. In Sanskrit many of the occurrences of retroflexes are conditioned; others are explained historically as reflexes of certain Indo-European consonants and consonant clusters. But, in fact, in Dravidian it is a matter of the utmost certainty that retroflexes in contrast with dentals are Proto-Dravidian in origin, not the result of conditioning circumstances. In Southern Dravidian, moreover, several languages have three phonemic series in the front of the mouth—dental, alveolar, retroflex —a possibility hardly envisaged by the framers of the International Phonetic Association's alphabet; the comparative evidence looks to similar distinctions in Proto-Dravidian. This being so for Dravidian, it is beyond doubt that, even where Indo-European material yields Sanskrit retroflexes, pre-Indo-Aryan and pre-Dravidian bilingualism provided the conditions which allowed pre-Indo-Aryan allophones to be redistributed as retroflex phonemes. Certainly as time went on, Middle Indo-Aryan showed more such phonemes than old Indo-Aryan, and in consequence Modern Indo-Aryan does so too. This is a clear instance of Indianization of the Indo-European component in the Indic linguistic scene.

A phonological example of more limited inter-influence—an isogloss, for which the historical solution is not yet at hand, is the following. In Marathi the palatals of Old Indo-Aryan are represented by *tš* and *dž* affricates before front vowels, by *ts* and *dz* affricates before back vowels; there are so many exceptions to this statement of distribution (because of recent borrowings from Sanskrit which always have *tš* and *dž*, and for other reasons) that it is necessary to postulate two sets of phonemes. A similar distribution is found in southern Oriya (Indo-Aryan; *LSI* 5.2.379), in Telugu and northern Kannaḍa (Dravidian; T. N. Sreekantaiya, Affricates in Kannada speech, *Indian linguistics* 1954.83–90), and, according to Bloch (following the *LSI* 4.169, but the distribution is completely

[6] Earlier in *BSL* 25.1–21 (1925) and *BSOS* 5.730–44 (1930).

uncertain), in Kurku (Munda). These languages form a continuous band across central India, and the trait undoubtedly originated in one language and spread to the others from it; which was the originator is, as I said, unknown. One can, however, guess from the distribution, which shows a very wide gap between southern Oriya (on the northeast) and Marathi (on the northwest), and from the fact that the feature appears only in northern dialects of Kannaḍa, that the feature originated in either of the two contiguous languages which show it in all dialects, viz. in Marathi or in Telugu. The only possibility of a decision between these two lies in philological work, which may possibly be successful in establishing a relative chronology. Kashmiri shows a similar phenomenon; this can only be of independent origin.[7]

On the side of morphology and syntax, it may be well to quote Bloch's summary statement (327–8), and then to enlarge on a few details. Bloch indicates that in what follows Dravidian and Indo-Aryan have more traits in common than Munda has with Indo-Aryan.

> Dans le mot, usage constant de la suffixation et absence (en ce qui concerne l'aryen, perte) des préfixes[8] et infixes, lesquels sont courants en munda; dans le groupe, absence (perte) des prépositions et des préverbes comme tels. Dans la flexion, absence (perte) du nombre duel, courant en munda. Dans les noms, double thème, le thème oblique étant susceptible de valeur génitive, et se faisant suivre de mots plus ou moins vides de leur sens propre; pronoms personnels à deux thèmes: celui de nominatif, et celui de régime direct et indirect (un seul thème en munda). Dans le verbe, troisième personnes de forme nominale et variables en genre; présence d'un gérondif (qui manque au munda) tenant un rôle important dans la liaison des phrases et dans la création de locutions composées à valeur stylistique ou grammaticale. ... Quant aux expressions comparables, et par exemple les mots doubles et à écho on en ferait aisément de longues listes dans toutes les familles de langues dravidiennes.

Not all of these will seem to be of equal cogency. Loss of the dual in Sanskrit is paralleled by its loss all over the rest of the Indo-European domain. Loss of infixation, which occurs after all only in a certain verb type, is similarly paral-

[7] Bloch (325 f.) attempts to use the presence of aspirated consonants in the northern Munda languages as somehow responsible for Indo-Aryan aspirated consonants. His argument is both confused and hypothetical; moreover, we know too little of Munda to be sure of the nature and the history of these sounds there. — E. Prokosch, *A comparative Germanic grammar* 39 (1939), based part of his argument about the origin of Sanskrit voiced aspirated stops (and consequently the nature of the Indo-European phonemes involved) on his belief that voiced aspirated stops are found in Tibeto-Burmese, Dravidian, and Munda. This is true for Tibeto-Burmese, and very uncertain in its bearing for Munda (for the reason given above). As regards Dravidian, Prokosch was misled by E. H. Tuttle (*Modern philology* 18.52), who mistakenly assumed that Proto-Dravidian had aspirated stops because of the quite secondary occurrence of such sounds in modern Kannaḍa and Telugu.

[8] Trubetzkoy, *Acta linguistica* 1.84, set down as one of the six characteristics of an Indo-European language: 'Das Wort muss nicht unbedingt mit der Wurzel beginnen. — Es gibt keine indogermanische Sprache ohne Präfixe. ... In den jüngeren indogermanischen Sprachen nimmt die Zahl solcher Präfixe stark zu.' And further (85): 'Eine Sprache, die nicht alle genannten Strukturmerkmale besitzt, darf nicht als indogermanisch gelten.' Did he know the Modern Indo-Aryan languages?

leled everywhere. The disuse of verbal prefixes as a living set of morphemes in Modern Indo-Aryan is closely tied up with the general shift of accent to initial syllables (whatever the exact chronology of this may be). The absence of prepositions is striking to an Indo-Europeanist or a speaker of a Western Indo-European language; it should be remembered, however, that in Sanskrit itself (and it inherits this trait from Proto-Indo-European) there is no class of 'prepositions'—the morphemes in question are rather 'adverbs in immediate constituency with nouns', the position being postpositional probably rather more often than prepositional. If these are replaced in Modern Indo-Aryan by noun forms invariably following the oblique form of the head noun, the construction is not too different from that of Sanskrit. Parallel constructions in Dravidian may possibly have helped toward the shift. The two themes of personal pronouns are paralleled by Dravidian phenomena; but Indo-European in general is marked by double stems in the personal pronouns (e.g. English *I*:*m-* [*me, my*]; Sanskrit *aham*:*ma-* [*mām, mayā, mahyam, mat, mama, mayi, me*]). It has been pointed out by others (*LSI* 4.280, whence Bloomfield, *Language* 470) that Modern Indo-Aryan, like Dravidian, adds the same inflexional (case) morphemes to distinct stems for singular and plural, which is unlike general Indo-European inflexional practice. This is convincing and to be interpreted as evidence of borrowing from Dravidian, even though similar structure is seen in Tocharian.

The phenomena pointed out by Bloch for the verb are more impressive. Especially am I impressed by the Sanskrit form which he calls the 'gérondif', which, following Whitney, is usually called in English 'gerund' (otherwise 'absolutive', 'indeclinable participle', 'conjunctive participle', 'adverbial participle').[9] All three major stocks show constructions in which verb stems or nonfinite verb forms are strung together in series which are closed by a finite verb form (or other predicate-ender). This is a prominent feature of Dravidian; it is well known in Munda also[10] and in Indo-Aryan. It is one of the syntactic features of Sanskrit that distinguishes it from other Indo-European languages, even though the actual forms used are relatable to Indo-European morphological material. We must look to the syntax of the non-Indo-European languages of India for the stimulus that brought about this re-use in India of older material. It might be expected that an attempt would be made to find priority between Dravidian and Munda for this type of construction. Such a discussion would be fruitless in the present stage of our knowledge of Munda, and especially moreover since such strings of verb stems or nonfinite forms are a common feature of so many other languages and language families, e.g. Vietnamese, Chinese, Japanese, Korean, Altaic, Finno-Ugric (at least Hungarian); it is Indo-European and Semitic, if I mistake not, that are aberrant in this matter in Eurasia.

[9] This form has neither morphologically nor in syntactic use anything to do with either Latin gerund or gerundive or participle; no really satisfactory term has yet been invented. Slavicists have been equally unhappy in calling Russian forms with a comparable syntactic use 'past gerunds'. These forms in Russian and other Slavic languages are a modern innovation, not found in Old Church Slavic. There is no connection with the Indic phenomena.

[10] Pace Bloch 327. See P. O. Bodding, *Materials for a Santali grammar, II. Mostly morphological* 273 ff. (1929); J. Hoffmann, *Mundari grammar* 189 ff. (1903); G. V. Ramamurti, *A manual of the So:ra: (or Savara) language* 28 §59, 29 §§71–2, 44 §143, §145.

Bloch (328) also mentions that Marathi, Oriya, and Sinhalese have constructions based on a nominalized or adjectivized form of a verb (or rather of a predication ending in a verb) followed by a postposition. I have available only material from Marathi (cf. Bloch, *La formation de la langue marathe* 260 f., §263), such as: *tujhī āī vārlyā-pāsūn* 'since (*pāsūn*) your mother (*tujhī āī*; subject) died (*vārlyā*)'. I have pointed out in *Lg.* 30.484 that there are in Buddhist Hybrid Sanskrit and in Pali constructions like this, in which the first member of a compound is a participle and the second a noun which is not the (syntactic) subject of the participle (e.g. *āgatakāle* 'at the time of [his] having come'). This is all parallel to the pan-Dravidian[11] construction in which a predication ending in an adjectivized or nominalized form of a verb is in attributive construction with a following noun (e.g. Kota *a·m unčvd unyp* 'the thought (*unyp*) which we (*a·m*) have thought (*unčvd*; adjective form of past paradigm)'. Bloch failed to note that Munda also has the construction.[12] There are many parallels to this construction in languages of Asia outside of India. We cannot decide priority as between Dravidian and Munda; we need only note that the Indo-Aryan tentatives in the direction of this construction (in all probability under Dravidian stimulus) are an Indianization, for there is nothing parallel to it elsewhere in Indo-European.

The echo-word construction mentioned by Bloch has been written about in some detail, both by a number of Indian scholars in early circulars of the Linguistic Society of India[13] and by myself in *New Indian antiquary* 1.109–17 (1938). It is generally a construction in which a basic word formulated as CVX is followed by an echo-word in which CV is replaced by a morpheme *gi-* or *u-* or the like (or C is replaced by *m-* or the like), and X echoes the X (or VX echoes the VX) of the basic word. The meaning of the echo-word is 'and the like'; e.g. *puli gili* 'tigers and the like.' There are many variations, though it is notable that nearly all the Dravidian languages have *gi-*. Most notable is the fact that the construction is found in all three families, there being good evidence for Dravidian, fairly good evidence for Indo-Aryan, and good evidence for at least So·ra· in the Munda family (G. V. Ramamurti 150 ff.). The chief So·ra· echo-morpheme is *m-*, which is evidenced also for Brahui, Kolami, Parji, Telugu, Tamil-Malayalam, and for various Indo-Aryan languages such as Ḍogrī. We need more detailed evidence and analysis, but it is clear already that echo-words are a pan-Indic trait and that Indo-Aryan probably received it from non-Indo-Aryan (for it is not Indo-European).

Finally, I would present in detail a matter which has not been noticed before. The phenomenon is of limited areal range, but appears in all three families, having spread from Indo-Aryan, though it is not an Indo-European phenomenon. This is the use of 'classifiers' or 'quantifiers'. In constructions marked by these, when a noun is numerated by means of a numeral or a similar word, the construction contains also one of a smallish class of words or morphemes which

[11] Bloch, *Structure grammaticale des langues dravidiennes* 64–6; Emeneau, *Lg.* 24.321, 1st paragraph.

[12] P. O. Bodding 50–1; J. Hoffmann lv–lvi, 120–1, 201–3; G. V. Ramamurti 49 §168, §§170–3.

[13] Circular 3, 14 May 1928, 7–8; Circular 4, 25 June 1928, 2, 8–10, 13–14, 16.

we can call by either of these terms. The term 'classifier' indicates that there are as many classes of nouns as there are classifiers; the term 'quantifier' indicates that in numeration of nouns there is always specification of the type of unit by which the species indicated by the noun is counted. The units indicated are of various kinds, either measured units of nondiscrete entities (e.g. a quart of liquid, an acre of land) or discrete entities as classed by various criteria (e.g. human vs. animal, animate vs. nonanimate, long and thin vs. flat and thin vs. spherical). Such quantifiers are, to be sure, used in probably all languages; English has *a ton of coal, two acres of land, three head of cattle*, etc. But the languages under discussion at the moment are not those in which only nouns denoting nondiscrete entities and a few others are classified or quantified, but those in which all or nearly all nouns are treated thus. Conspicuous as having such systems are Chinese, Japanese, Korean, Vietnamese, Khmer, Thai, Burmese, and Malay.

The existence of classificatory systems in some of the languages of India has hardly been noted and is, as a matter of fact, difficult to get information on. They have been reported for the Magadhan languages of Modern Indo-Aryan, viz. Bengali, Assamese, Oriya, and in Bihari for some dialects (e.g. Maithili).[14] The systems in these closely related languages involve a half-dozen or more classes. The morpheme order is noun + numeral + classifier, or numeral + classifier + noun; there is no information on different meanings for the two constructions. According to S. K. Chatterji, since all the languages use practically the same morphemes, the modern systems are all descendants of a system that originated in the Magadhan Apabhraṃśa at the end of the Middle Indo-Aryan period.

Less attention has been paid to the fact that Marathi has a meager suggestion of this system with one classifier, viz. *jaṇ* [dzʌɳ] 'person', fem. *jaṇī*, when nouns denoting persons are numerated by numerals higher than four (and optionally for two to four).[15] This is historically related to one of the Magadhan classifying morphemes (Bengali *jan*, etc.), but the word order in the construction is fixed in Marathi (numeral + classifier + noun) as opposed to the variation in the Magadhan languages. This, combined with Marathi's having only one classifier, seems to argue for a certain degree of independence in the development of the systems in the two Indo-Aryan branches, at least since Middle Indic times, though it is not ruled out that the Marathi construction owes its inception to some stimulus ultimately deriving from the full-fledged system of Magadhan.

[14] The best accounts with examples are given by Suniti Kumar Chatterji, *The origin and development of the Bengali language* 777–81 ('enclitic definitives or numeratives'); Banikanta Kakati, *Assamese, its formation and development* 265–9 (1941); (Sir) George Abraham Grierson, *Seven grammars of the dialects and sub-dialects of the Bihárí language*. The matter was noticed by Bloch, *L'indo-aryen du Veda aux temps modernes* 189 (1934); he thinks of substratum influence and refers to Siamese.

[15] The matter is hard to dig out of the Marathi grammars. A summary statement is found in H. M. Lambert, *Marathi language course* 243 (1943). An even more summary statement for Konkani is in S. M. Katre, *The formation of Koṅkaṇī* 117 §236 (1942). Bloch has no notice of this construction in *La formation de la langue marathe*; in his reference to the Magadhan phenomena in *L'indo-aryen* 189, he says that the construction is limited to the eastern, i.e. the Magadhan, group of languages.

For some of the village Hindi dialects show a use of *janā* like that of Marathi. The construction is not described in the grammars of Hindi, which deal essentially with urban forms. The Nepali dictionary also records such a use for Nepali *janā*.[16] Until better descriptive accounts are available for Modern Indo-Aryan languages and dialects, it will be impossible to map the present use of classificatory systems in this section of India, and until a more searching study has been made of the various stages of Middle Indo-Aryan, the history of the systems will remain a matter of speculation.

Classificatory systems have been found also in some of the Dravidian and Munda languages. Of the former, Kolami (Wardha dialect), Parji, the Kui-Kuwi dialects, and Kurukh and Malto have such systems, and in each instance it has been borrowed from Indo-Aryan neighbors.

The Wardha dialect of Kolami classifies persons when they are numerated by the numerals six and over. These numerals are borrowed from Marathi, as well as the Marathi classifier in the form *zen* [dzen], with feminine *zenikul*, i.e. Marathi fem. *jaṇi* plus the Kolami plural suffix *-kul*. This applies also to the numeral five, when the Marathi numeral is optionally used instead of the Kolami numeral. This Kolami dialect is in predominantly Marathi-speaking country. The Adilabad Kolams have not borrowed so many numerals and do not use the Marathi classifier with the Kolami numerals, which are the only ones that have been reported.[17]

In the account of Parji no statement is given about the matter, but the texts provide a few examples that indicate that *jan* is used for persons and that in addition *gŏṭa* is used for certain nouns denoting nonpersons.[18] The neighboring Indo-Aryan language is Halbi, for which our information is not good. The Indo-Aryan classifier represented by Parji *gŏṭa* presumably is found in Halbi (so *LSI* 7, Standard List item 114, would seem to indicate); it certainly is not Marathi, just as certainly is Magadhan, and is also found in Chhattīsgaṛhī dialects of Eastern Hindi (*LSI* 6.215, 225). The account of Parji says that the numerals from six on are borrowed from Halbi; apparently the classifiers are used only with these numerals of Halbi origin.

The Kuwi dialect described by Schulze has a system like that of Parji, with Oriya numerals from three on and the classifiers *zāna* (*z* = *j*) for nouns denoting male persons and *gotta* for all others (corresponding to the Kui-Kuwi gender

[16] Oral communication for Hindi from my pupil Phillip Barker, on the basis of experience in several Hindi-speaking village communities at widely separated places. A few scattering examples turn up in *LSI*, but are not worth recording. It should however be noted that the *LSI* records *jhan* 'people' with *jh-* instead of *j-* for Chhattīsgaṛhī Hindi, and also for Halbi and the Bihari spoken in the Ranchi district (*LSI* 5.2.284). For Nepali, Turner's dictionary s.v. *janā*.

[17] Wardha dialect in M. B. Emeneau, *Kolami, a Dravidian language* §4.63, §8.55 (University of California Publications in Linguistics, Vol. 12; 1955); other dialects reported so far as I know them in §§9.21–3. The Adilabad dialect is in P. Setumadhava Rao, *A grammar of the Kolami language* (1950).

[18] T. Burrow and S. Bhattacharya, *The Parji language* (1953). References for *jan* are 71 sent. 15, 104 sent. 1 and 4, 107 sent. 24, 120 sent. 1 (2 exx.); for *gŏṭa*, 108 sent. 30, 133 sent. 89.

system).[19] The language is spoken at the boundary between Oriya and Telugu. The accounts of other Kui-Kuwi dialects, one of which is to the north in Oriya territory, do not speak of a system of classifiers. However, in Fitzgerald's account of Kuwi there is one phrase with the classifier for persons (126, *tīnijoña māska* 'three girls'). Winfield (37) describes Kui forms for the numerals from three on, which are very like those of Bengali which have the classifier enclitic *-ṭā*; the order too is Magadhan (noun + numeral + classifier). These few examples are sufficient evidence that all the Kui-Kuwi dialects have borrowed the classificatory system of Oriya.

The Kurukh system shows very close similarities to the system of the Magadhan languages, including an option in word order, and a number of classifying morphemes, most, if not all, of which are borrowings from the contiguous Magadhan languages Bihari and Oriya (Chhattīsgaṛhī Hindi is also a neighbor, but does not have so elaborate a classificatory system) (e.g. *jhan, goṭaŋg, ṭhur* [cf. Bengali *-ṭu*]). The classifiers are used not merely with the Indo-Aryan numerals which have been borrowed from four on, but also with the Dravidian numerals two to four.[20]

The Malto system is the most complex of all these.[21] With the numerals from three on, which it has borrowed from the surrounding Bengali or Bihari, it uses a number of classifiers and the order numeral + classifier + noun. The classifiers are in part at least borrowed from a Magadhan system and include *jen* and *goṭa*. In addition there are used *ḍaṉṛa* for long objects (roads, bamboos, articles of clothing; < Modern Indo-Aryan **ḍaṇḍ-* < Sanskrit *daṇḍa-* 'stick'), *paṭa* for 'objects distinguished for surface' (plates, combs, ponds, clouds, bedsteads; < Modern Indo-Aryan **paṭ-*, **paṭṭ-*, **pāṭ-*, cf. Sanskrit *paṭṭaka-*'board'), *kaḍa* for tendril-like things (hair, wire, grass stalks; etymology not yet found), and several others whose etymologies have not yet been found. In addition to this, the Dravidian numerals for one and two, when they enumerate nouns denoting nonpersons, are used with some of the same set of classifiers just listed (*paṭ, kaḍ, ḍaṉṛ*, etc.) plus several others, e.g. *páṉṛ* for round objects (no etymology yet found) and *maq* for animals (Dravidian in origin).[22] In this last construc-

[19] F. V. P. Schulze, *A grammar of the Kuvi language* 100 ff. (1911); A. G. Fitzgerald, *Kuviṅga bassa: The Khond language as spoken by the Paṛjas ... of the Madras Presidency* (1913); W. W. Winfield, *A grammar of the Kui language* (1928).

[20] A. Grignard, *A grammar of the Oraon language* 29 f. (1924).

[21] Ernest Droese, *Introduction to the Malto language* 15–26 (1884).

[22] In this etymology and that given in note 25, the sigilla for the languages are as in *Lg.* 29.339 fn. 1, or as in *Kolami, a Dravidian language*. Ta. *maka, makavu* child, young of animal; *makaṉ* son, man, husband; *makaḷ* daughter, woman, wife; *makkaḷ* children, human beings; *mōṉai* sonny (term of endearment in addressing a child); Ma. *makan*, (vulg.) *mōn* a son; *makaḷ* a daughter; *makkaḷ* children, the young of animals; Ko. *mog* (oblique *mog-*, *mo·t-*) child, wife; *pe-mog* woman (*peṇ* female); *mo·l* daughter; To. *mox* (obl. *mo·t-*) child, son, male, daughter, woman; *toz mox* Toda woman or wife; Ka. *maga* son, male; *magaḷ* daughter; *makkaḷ* children; *magavu, moga, mogu, moguvu* infant; Koḍ. *mo·vĕⁿ* son; *mo·va* daughter; *makka* children; *pom makka* women, wives (*poṇṇī* wife, female); Tu. *mage, mōnu* son; *magaḷu, mōni* daughter; *makkaḷāṭigè* childishness; Te. *maga*, coll. *moga* male; *magāḍu* husband, a man, a male, a hero; *magadi* male of any animal, beast, or bird; *maganālu* wife; Malt. *maqe* boy; *maqi* girl; *maqo* small, little, young; *maqu* a young one (animal); *maqmaqo* small ones (*q* = *x*); Kol. *magvan* husband (so correct *Kolami, a Dravidian language*).

tion many nouns are used as their own classifiers (e.g. *man-ond manu* 'one tree'). A unique feature of this construction for one and two is the order: classifier + numeral + noun, found nowhere else in any of the systems in India and presumably a Malto invention. It is to be noted too that the numeral forms *-ond* 'one' and *-is* 'two' in this construction are found nowhere else but in this construction (cf. *ort* 'one person', *ivr* 'two persons'), and that for 'one' the closely related Kurukh has the forms *ort* 'one person' and *oṇṭā* 'one animal or thing', the latter of which may either precede or follow a noun (contrasting with the single order in Malto). Malto, then, presumably began by borrowing numerals and classifiers from a neighboring Magadhan language, and subsequently elaborated the system independently, chiefly on the basis of inherited material.

For the Munda languages the evidence is less good. Koṛowa[23] certainly uses *jhan* and *gwoṭ/gwoṭaŋ*, borrowings from Indo-Aryan, as well as others, like *boʔ* 'head of cattle' and *hoṛ* 'person', which are not borrowed; *hoṛ* and *jhan* seem to be in free variation. The numerals involved are not borrowings. The order is numeral + classifier + noun. It is fairly clear from the accounts of Santali and Mundari that they use a similar system; it is quite clear that So·ra· does not.

Here then is a large area of India, especially eastern and central India, with this feature. My reconstruction, relying on the fact that some, if not only, Indo-Aryan classifier morphemes are used in all the languages involved and on the further fact that these morphemes are used only with Indo-Aryan numerals in some of the non-Indo-Aryan languages, is that the construction (so far as India is concerned) is originally Indo-Aryan. It spread thence to the other languages as a total construction consisting of numeral + classifier, and then was elaborated in some of the languages with native material, the native numerals, native morphemes as additional classifiers, etc.

The problem of Telugu and Kannaḍa is difficult. In certain dialects of Telugu[24] the numerals from eight to ten are followed by the classifier *mandi* when persons are numerated; e.g. *enimidi mandi manuṣulu* 'eight men'. Up to ten this morpheme is in complementary distribution with the suffix *-guru* found in the forms denoting persons from three to seven (e.g. *mug-guru manuṣulu, nalu-guru manuṣulu*, etc.) and the suffix *-aru* in *idd-aru manuṣulu* 'two men'. I do not know whether *mandi* is to be classed as a free form or as a suffix; presumably complete analysis of the language will answer this question. The morpheme is of Dravidian origin.[25] Taken by itself it might be thought that Telugu had

[23] Information from Phillip Barker. Note the form with *jh-*, which is that found in Kurukh and in the neighboring Hindi dialects.

[24] Information from Bh. Krishnamurti.

[25] A. H. Arden, *A progressive grammar of the Telugu language* 88 §236 (4th ed., 1937), does not make the matter at all clear. Etymology of *mandi*: Ta. *maṉṟu* hall of assembly, court of justice, cow-stall, herd of cows, raised platform under a tree for village meetings, juncture of four roads; *maṉṟam* hall, assembly, court, meeting place under a tree, open space, cow-shed, long street; *mantai* (not in old literature) flock, herd, common pasture, open space in middle of a village; Ma. *maṉṉam, maṉṉu* a place of judgment or discussion; Ko. *manḍ* Toda *mund*; burning place for dry funeral; *mandm* meeting; To. *moḍ* locus of tribal activity, including village with dairy, dairy apart from village, funeral place; patrilineal clan; Ka. *mande, mandi* flock of sheep or goats, herd of cattle or buffaloes; open place (in the jungle

developed this construction under the stimulus of Oriya; without a complete mapping of dialects and philological work to determine chronology one cannot be sure about this possibility. Another problem is why the use of *mandi* should start only with eight. Perhaps the statement of it as in complementary distribution with *-aru* and *-guru* is the answer; these suffixes do not occur from eight to ten and perhaps *mandi*, as it were, makes up for their absence. But such an explanation is uncomfortably teleological and ad hoc, and the fact that there seems to be some free variation above ten is disturbing. It is hardly possible to dissociate the Telugu facts from those of Kannaḍa. About the latter we are told[26] that in the modern language and occasionally in the medieval one, *mandi* is added to the cardinal numerals to replace special forms for numerating persons. It is uncertain at what point the series starts; one authority says at four, a better one gives the form for three as an example. The special political interrelation between Telugu and Kannaḍa in the medieval period (Vijayanagar kingdom, 1336–1565 A.D.) allowed extensive borrowing between them in both directions. If the etymological discussion in note 25 is correct, Telugu in all probability owes the construction and the form to Kannaḍa. It is not impossible that the Kannaḍa construction might be a calque from Marathi.

We must note too that Tamil, presumably the colloquial, uses *pēr* 'name' as classifier in numerating persons from two on; e.g. *nālu pēr tiruṭar* 'four thieves.'[27]

I am at a loss to say more about these Tamil, Kannaḍa, and Telugu examples. But certainly we must not ascribe the use of classifiers to Proto-Dravidian.

The use of classifiers can be added to those other linguistic traits previously

---

or near a village) where a flock or herd stands, a pen, fold; persons, people; Koḍ. *mandï* village green; Tu. *mandi*, *mandè* people, persons; Te. *manda* flock, herd, drove, pack, (B also) place where flocks or herds are kept outside a village, hamlet inhabited by herdsmen; *mandi* a number, crowd, or collection of persons; retinue, infantry; Kol. (SR) *mandī* men; (Kin.) *mandi* man. If all these words are to be put together, the Te. forms require some phonological explanation. The correct phonetic correspondence is: PrDr *$\underline{n}\underline{r}$: Ta. $\underline{n}\underline{r}$, Ma. $\underline{n}\underline{n}$, Ko. $\underline{d}$ (the forms in this etymology will not be discussed here), To. $\underline{d}$, Ka. Koḍ. *nd*, Tu. *ṇḍ* (and other developments), Te. *ṇḍ*, Kol. *nd*, Pa. *nd*, (NE dialect) *nḍ*, Oll. *nḍ* (see further Emeneau, *Kolami, a Dravidian language* §10.25). A simple example for Te. is the word for belleric myrobalan (Terminalia bellerica Roxb.): Ta. *tāñṟi*, Ma. *tāñña*, Ka. *tāṟi* (this is a form without nasal), Tu. *dāṇḍi*, Te. *tāḍi*, *tāṇḍra*, Pa. *dēndi*. However, there a few noteworthy exceptions to the statement, including the word under discussion and the following two: 'pig,' Ta. *paṉṟi*, Ma. *paṉṉi*, Ko. *paj* (*-$\underline{n}\underline{r}i$ > *j*), To. *pody* (probably a borrowing from Badaga, a Ka. dialect), Ka. Koḍ. *pandi*, Tu. *pañji*, Te. *pandi*, Go. *paddī*, Pa. *pend*, (NE) *peṇḍ*, Oll. *paṇḍ*, Kui *paji*; Ta. *kaṉṟu* young of various animals (calf, colt, etc.), sapling, Ma. *kaṉṉu* (oblique *kaṟṟu*) young of cattle (especially buffalo calf), young plantain trees round the mother plant, *kaṟṟa* boy, calf, To. *koṛ* female buffalo calf below one year, Ka. *kaṟu*, *kaṟa*, *kaṟuvu* calf, *kanda* young child, *kandu* calf, young plantain trees round the mother plant, Tu. *kañji* calf, Te. *kandu*, *kanduvu* infant, Pa. *kar* sapling, Kur. *khadd* child, young animal or plant, Malt. *qade* son. There seems to be no possibility of a contextual explanation for the Te. irregularity. Influence from Ka. might be invoked in all these cases, from Ta. also in the latter two.

[26] F. Kittel, *A grammar of the Kannaḍa language* 258 (1903); Harold Spencer, *A Kanarese grammar* 94 (1914).

[27] A. H. Arden, *A progressive grammar of common Tamil* 118 (4th ed. 1934); Julien Vinson, *Manuel de la langue tamoule* 102 (1903).

discussed, which establish India as one linguistic area[28] for historical study. The evidence is at least as clear-cut as any that has been advanced in the establishment of a linguistic area in any part of the world, and in fact a good deal more so than much that has been offered. It is to be hoped that it will not be neglected henceforth when the question is raised whether linguistic features, especially those of morphology and syntax, can diffuse across genetic boundaries. Some of the features presented here are, it seems to me, as 'profound' as we could wish to find (if we must attempt to apply Sapir's value criteria). Certainly the end result of the borrowings is that the languages of the two families, Indo-Aryan and Dravidian, seem in many respects more akin to one another than Indo-Aryan does to the other Indo-European languages.[29]

In another place I adumbrate an attempt to include the linguistic area India in the larger linguistic area of East, Southeast, and South Asia. The evidence so far found concerns the use of classifiers and makes it at least possible that this trait reached the Indo-Aryan languages of the Magadhan area from Southeast Asia; but the demonstration of this is not as clear as that of the relationships within India and need not be given here to obscure the clear outlines of the matter discussed in this paper.

[28] This term 'linguistic area' may be defined as meaning an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the families. It is perhaps not quite satisfactory as a technical term, though it has the virtue of having been used previously in this sense by H. V. Velten as a translation of Trubetzkoy's 'Sprachbund', *Pacific Northwest quarterly* 34.271-92 (1943). (My attention was called to this by Sebeok, *IJAL* 10.214 [1944], and Voegelin, *Word* 1.58 [1945]). Among the disadvantages of the term is the lack of an adjective and the impossibility of using the reverse phrase 'areal linguistics', since this is preempted by the Italian neolinguistic school in another sense. Perhaps however it will do for the moment, until some more ingenious scholar invents better terminology.

[29] We must not, however, neglect Bloch's final remark and his reasons therefor (330): 'Ainsi donc, si profondes qu'aient été les influences locales, elles n'ont pas conduit l'aryen de l'Inde ... à se différencier fortement des autres langues indo-européennes.'