DOCUMENT RESUME

ED 034 175 AL 002 172

AUTHOR Bright, William
TITLE Hindi Numerals.
INSTITUTION Hawaii Univ., Honolulu. Dept. of Linguistics.
PUB DATE Oct 69
NOTE 20p.; Offprint from "Working Papers in Linguistics," Issue No. 9, Oct 1969

EDRS PRICE EDRS Price MF-$0.25 HC-$1.10
DESCRIPTORS Contrastive Linguistics, *Hindi, *Morphology (Languages), Morphophonemics, *Number Concepts, *Number Systems, *Sanskrit, Structural Analysis

ABSTRACT

In most languages encountered by linguists, the numerals, considered as a paradigmatic set, constitute a morpho-syntactic problem of only moderate complexity. The Indo-Aryan language family of North India, however, presents a curious contrast. The relatively regular numeral system of Sanskrit, as it has developed historically into the modern languages of this group, has undergone striking phonological alteration. Morphemes which had relatively uniform shapes and clear boundaries in the parent language have become fused and difficult to identify. The result is that anyone who learns to count in one of these languages must make a greater learning effort than is usually required for the counting process. The present paper raises some questions concerning these numeral systems, with specific reference to Hindi. Is memorization the only factor involved in the learning and production of the paradigm up to 100? If so, should a grammar simply list these hundred forms with no attempt to state general rules governing their phonological shapes? Or, are the Indo-Aryan numerals in fact governed by rules which are used by the native speaker and may be stated by the linguist? In exploring these questions, a complete set of numerals and a morphological analysis of their paradigm are presented and discussed. (This paper, of a "preliminary nature," constitutes a progress report.) (AMM)

This is an offprint from

WORKING PAPERS IN LINGUISTICS

Department of Linguistics
University of Hawaii
Honolulu 96822

Issue No. 9, October 1969

The notes and articles contained in these Working Papers constitute progress reports on work being carried on in the Department. While they are not intended for public distribution because of their tentative and preliminary status, they are being privately circulated to interested scholars on a limited basis for the purpose of inviting comments and suggestions on the ideas set forth in them. Because these papers are not finished products embodying the final views of their authors, readers are advised that they should not be cited without some allusion to their preliminary nature. They are not to be reproduced without their authors' consent.

HINDI NUMERALS

William Bright

In most languages encountered by linguists, the numerals, considered as a paradigmatic set, constitute a morpho-syntactic problem of only moderate complexity. Thus, a common pattern, not only of Indo-European but in other families as well, involves single morphemes for numbers 'one' through 'ten', and regular rules of co-occurrence to produce numbers 'eleven' through 'ninety-nine' -- with occasional ad-hoc morphophonemic statements applying to the teens and the decades, e.g. English *three* → *thir-* in the environment of *-teen* or *-ty*. The description of such numerals can be incorporated with relative ease into a grammar of the larger, open-ended numeral systems of the language concerned.[1]

In contrast to such patterns, the Indo-Aryan family of North India presents a curious contrast. The relatively regular numeral system of Sanskrit, as it has developed historically into the modern languages of this group, has undergone striking phonological alteration. Morphemes which had relatively uniform shapes and clear boundaries in the parent language have, in this process, become fused and difficult to identify -- a reversal of the general Indo-European trend away from more fusional constructions and toward more agglutinative ones.

Witness Table 1, which gives partial paradigms in Sanskrit (Whitney 1889:177-9) as compared with modern Gujarati (Cardona 1965:84-5).[2] The irregularity visible even in this small sample is in fact, in modern Indo-Aryan generally,[3] maintained as far as '99'. The result is that anyone who learns to count in one of these languages -- whether native speaker or foreign student -- must make a greater learning effort than is usually required for the counting process. Once past '99', the going is easier; there is a monomorphemic word for 'hundred' (e.g. Gujerati sɔ̃), and prediction of higher numberals is then possible (e.g. car sɔ̃ '400').

TABLE I

	Sanskrit	Gujarati		Sanskrit	Gujarati
40	catvaariśat	calis	50	pancaaśat	pɔ̃cas
41	eka-catvaariśat	ektalis	51	eka-pancaaśat	ekavɔ̃n
42	dvaa-catvaariśat	betalis	52	dvaa-pancaaśat	bavɔ̃n
43	trayaś-catvaariśat	tetalis	53	trayah-pancaaśat	trepan
44	catuś-catvaariśat	cumalis	54	catuh-pancaaśat	cɔ̃pɔ̃n
45	panca-catvaariśat	pistalis	55	panca-pancaaśat	pɔ̃̃cavɔ̃n

Questions which are of some linguistic interest may be asked with regard to these systems. Is memorization the only factor involved in the learning and production of the paradigm up to 'one hundred'? If so, should a grammar, for the sake of psychological realism, simply list these hundred forms (as, in

fact, practical grammars do), with no attempt to state general rules governing their phonological shapes? To put the matter in other terms, should we regard all the forms from '11' to '99' as suppletive? Then we would simply say that the string of morphemes ONE-FIVE-DECADE yields ekavən 'fifty-one' in Gujarati, just as we might say that English GO-PAST yields went. Or, to take an opposite view, are the Indo-Aryan numerals in fact governed by rules, irregular to be sure, but having SOME degree of generality, which are used by the native speaker and may be stated by the linguist?

The present paper will explore these questions with specific reference to Hindi, in the following steps: a complete set of numerals from one to a hundred will be presented; a morphological analysis of this paradigm will be attempted; and finally, the value of the analysis will be discussed. However, there is one difficulty at the start: namely, that many published sources give alternative forms for the Hindi numerals -- and, indeed, virtually every source gives a slightly different set. For example, '67' is given variously as satsaTh (Harter 1960), sarsaTh (Kellog 1938), and saRsaTh (Sharma 1958).[4] The present description is based, to begin with, on the usage of a single informant on a single occasion: Miss Manjari Agrawal, a native Hindi speaker from Delhi, was asked to count to a hundred at a 'normal' speed, and the results were tape-recorded and transcribed.

Subsequent discussion with Miss Agrawal revealed alternative forms in her usage -- though not as much free variation as the published sources suggest. The attested variations will be taken into consideration at a later stage in this discussion; but first, let us consider the tape-recorded forms and their analysis (Table 2).[5]

TABLE 2

1 eek	21 ikkiis	41 ikyaaliis	61 iksaTh	81 ikyaasii
2 doo	22 baaiis	42 byaaliis	62 baasaTh	82 byaasii
3 tiin	23 teeiis	43 tẽẽtaaliis	63 treesaTh	83 teeraasii
4 caar	24 caubiis	44 cãũtaaliis	64 causaTh	84 cauraasii
5 pããc	25 pacciis	45 pãĩtaaliis	65 pãĩsaTh	85 picyaasii
6 chee	26 chabbiis	46 cheeyaaliis	66 chyaasaTh	86 chyaasii
7 saat	27 sataaiis	47 sãĩtaaliis	67 saRsaTh	87 sataasii
8 aaTh	28 aThaaiis	48 aRtaaliis	68 aRsaTh	88 aThaasii
9 nau	29 untiis	49 unancaas	69 unahtar	89 unaanbee
10 das	30 tiis	50 pacaas	70 sattar	90 nabbee
11 gyaarah	31 ikattiis	51 ikyaavan	71 ikahtar	91 ikyaanbee
12 baarah	32 battiis	52 baavan	72 bahattar	92 byaanbee
13 teerah	33 tãĩtiis	53 treepan	73 teehattar	93 teeraanbee
14 caudah	34 cãũtiis	54 cauvan	74 cauhattar	94 cauraanbee
15 pandrah	35 pãĩtiis	55 pacpan	75 pichyattar	95 picyaanbee
16 soolah	36 chattiis	56 chappan	76 chyattar	96 chyaanbee
17 satrah	37 sãĩtiis	57 sataavan	77 satattar	97 sataanbee
18 aThaarah	38 aRtiis	58 aThaavan	78 aThattar	98 aThaanbee
19 unniis	39 untaaliis	59 unsaTh	79 unaasii	99 ninyaanbee
20 biis	40 caaliis	60 saaTh	80 assii	100 sau

The question of what morphemes are present in these forms, and in what combinations they occur, is the simpler part of our analysis. There are thirteen minimum meaningful elements, falling into several classes. Each of the following morphemes constitutes a distributional class in itself:

TEN, phonologically <u>das</u> in isolation

MINUS ONE, with the usual shape <u>un-</u>

DECADE, with no uniform phonological shape

HUNDRED, invariably <u>sau</u>

The remaining morphemes belong to the general class of UNITS, and are further subject to two intersecting sub-classifications, as follows (phonological shapes given are those occurring in isolation):[6]

	$UNIT_a$	$UNIT_b$
ONE <u>eek</u>	+	-
TWO <u>doo</u>	+	+
THRHE <u>tiin</u>	+	+
FOUR <u>caar</u>	+	+
FIVE <u>pāǰc</u>	+	+
SIX <u>chee</u>	+	+
SEVEN <u>saat</u>	+	+
EIGHT <u>aaTh</u>	+	+
NINE <u>nau</u>	-	-

The given numeral paradigm may then be generated by the following rules. The numberals produced by each part of the rule are indicated in the column at the right.[7]

NUMERAL →	UNIT	'1-9'
	NUMERAL COMPLEX	'10-99'
	HUNDRED	'100'

NUMERAL COMPLEX →	($UNIT_a$) — TEN		'10-18'
	($UNIT_a$ / MINUS ONE)	— DECADE COMPLEX	'20, 30, ...80' '21-28, ...81-88' '19, 29, ...79'
	(UNIT / MINUS ONE)	— NINETY	'90' '91-99' '89'

DECADE COMPLEX →	$UNIT_b$ — DECADE	'20, 30, ...80'
NINETY →	NINE — DECADE	'90'

The above rules do not, however, account for the phonological shapes of our data, where the major complexities lie. We may ask first, then: is it possible to regard these forms as composed of successive phonological strings? In two cases, the answer is clearly positive: MINUS ONE and HUNDRED appear invaribly as un- and sau, respectively. However, to go to the extreme, the DECADE morpheme, though correlated with a string -iis in biis '20', tiis '30', caaliis '40', clearly cannot be assigned any non-arbitrary phonological shape in pacaas '50',

saaTh '60', and the rest. We must, then, consider the combinations of UNIT and DECADE as suppletive in shape, generated by rules such as TWO — DECADE → biis and SIX — DECADE → saaTh.

Combinations of UNITS with TEN, and with the DECADE COMPLEXES as suppletively produced, must then be accounted for. All of these forms are, in fact, divisible into successive morphemes, if one allows enough allomorphic variation. It is clear that one would prefer those morpheme cuts which yield the lowest allomorph count. Some choices are easy: saaTh '60' appears in the allomorph -saaTh after all UNITS. Other choices are difficult, as in the following forms:

ikyaaliis	'41'	ikyaavan	'51'
byaaliis	'42'	baavan	'52'

Here, apparently, there are common elements iky- '1', -aaliis '40', -aavan '50'; only for the morpheme TWO would we have to recognize allomorphs by- and b-. But looking further in the data, we find:

cãũtaaliis	'44'	cauvan	'54'

These forms suggest that perhaps the element meaning '50' is -van instead of -aavan, in which case we must recognize a morph ikyaa- 'one' in ikyaavan '51'. We can, in fact, recognize the same ikyaa- in ikyaaliis '41', with a general morphophonemic rule, aa + aa → aa, applying to its combination

with -aaliis '40' (which is independently identifiable in cheeyaaliis '46'). In a few cases, however, only an arbitrary morpheme cut can be made; e.g., in unancaas '49', unan- would be a unique alternant of un- 'minus one', but -ancaas is a similarly unique alternant of pacaas '50'.

After a good deal of trial-and-error, one can arrive at a description of minimum complexity. The one which follows contains two types of rules: (1) those which state the allomorphs of particular morphemes, and (2) those which state morphophonemic replacements operating throughout the system. Some of the latter rules apply through the entire language; others are more limited.

The allomorphic rules are as follows:

ONE eek	→	gyaa-	/ ___	TEN
		ik-	/ ___	'20, 60, 70'
		ikaX-	/ ___	'30'[8]
		ikyaa-	/ ___	'40, 50, 80, 90'[9]
TWO doo	→	baa-	/ ___	{TEN '20, 50, 60'}
		baX-	/ ___	'30, 70'
		by-	/ ___	'40, 80, 90' (all cases before aa)
THREE tiin	→	tee-	/ ___	{TEN '20, 40, 70'}
		tai-	/ ___	'30'
		tree-	/ ___	'50, 60'
		teer-	/ ___	'80, 90' (all cases before aa)
FOUR caar	→	cau-	/ ___	{TEN '20, 30, 40, 50, 60, 70'}
		caur-	/ ___	'80, 90' (all cases before aa)

FIVE pääc →	pand-	/	__	TEN
	pac-	/	__	'20, 50'
	päi-	/	__	'30, 40, 60'
	picy-	/	__	'70, 80, 90'
SIX chee →	soo-	/	__	TEN
	chaX-	/	__	'20, 30, 50'
	chyaa-	/	__	'60'
	chy-	/	__	'70, 80, 90'
SEVEN saat →	sat-	/	__	{TEN / '70, 80, 90'}
	sataa-	/	__	'20, 50'
	sai-	/	__	'30, 40'
	saR-	/	__	'60'
EIGHT aaTh →	aThaa-	/	__	{TEN / '20, 50'}
	aR-	/	__	'30, 40, 60'
	aTh-	/	__]	'70, 80, 90'
NINE nau →	niny-	/	__	'90'[10]
TEN das →	-rah	/	'1, 2, 3, 5, 7, 8'	__
	-dah	/	'4'	__
	-lah	/	'6'	__

biis 'twenty' ⟷ -Xiis / {MINUS ONE / '1, 2, 3, 5, 7, 8'} —

Mtiis 'thirty' has no allomorphic alternation.[11]

caaliis 'forty' → -Mtaaliis / {MINUS ONE / '3, 4, 5, 7, 8'} —

-aaliis / '1, 2, 6' —

pacaas 'fifty' → -ancaas / MINUS ONE —

-van / '1, 2, 4, 7, 8' —

-pan / '3, 5, 6' —

saaTh 'sixty' → -saTh / {MINUS ONE / UNIT} —

sattar 'seventy' → -ahtar / {MINUS ONE / '1'} —

-hattar / '2, 3, 4, 5, 6, 7, 8' —

assii 'eighty' → -aasii / {MINUS ONE / UNIT} —

nabbee 'ninety' → -aanbee / {MINUS ONE / UNIT} —

The morphophonemic rules, with examples, are:

(1) X + C → CC (otherwise X → ∅)

ik- '1' + -Xiis '20' → ikkiis '21'

chaX- '6' + biis '20' → chabbiis '26'

(2) VV + M → ṼṼ (otherwise M → ∅)

cau- '4' + -Mtiis '30' → cãũtiis '34'

(3) aa + aa → aa[12]

ikyaa- '$_1$' + -aasii '80' → ikyaasii '81'

(4) ee + aa → eeyaa[13]

chee '6' + -aaliis '40' → cheeyaaliis '46'

(5) Cy + h → Chy

picy- '5' + -hattar '70' → pichyattar '75'

(6) Ch + h → Ch

aTh- '8' + -hattar '70' → aThattar '78'

If a somewhat different set of forms had been used, e.g. those presented in any particular textbook of Hindi, the details of this description would of course be different; but the general outlines, and the approximate degree of complexity, would be the same. Having arrived at this point, we are still faced with these questions: Is such an analysis meaningful, or worth doing? Does it correspond to any reality in the competence of the Hindi speaker who uses these forms?

We may refer, in this connection, to the discussion of related issues by Garvin and by Voegelin 1962. The former author, presenting a complex problem in the morphological analysis of Palauan, raises this question: How does the linguist, on the practical level, make a non-arbitrary choice between alternative analyses, each of which achieves some sort of economy? His answer is: one chooses that solution which 'yields results of the kind that allow the analyst to suspect the presence of general conditions in the language rather than merely particular conditions pertaining to the sample at hand' (Garvin 1961:68). Commenting on Garvin's article, Voegelin (1962:47) makes the point that the dictionary of a language,

rather than the grammar, is the proper 'repository for irregularities' such as our rules of Hindi allomorphy. But these authors are focussing on the question of alternative segmentations of morphological data, and of resulting degrees of generality and economy in linguistic statement; they are not explicitly raising the issue of psychological validity which I wish to consider here. From my point of view, the question is this: Is it meaningful to make ANY segmentation of the Hindi numerals into specified phonological strings? Or is it more realistic to simply list the entire set of forms, with the implication that native speakers produce them entirely from memory, rather than by application of rules? If the answer is to be obtained purely on the basis of economy in description, than we might reason as follows: In the description given above, there are 184 items which represent phonological shapes or specific environments. If, on the other hand, we give a simple list of the phonological shapes of forms -- in effect, an ad-hoc rule for each form -- then, of course, there would be just 100 items, with a clear advantage in economy. But we have no guarantee that economy in rules is a simple or unique reflection of psychological reality.

I would like to suggest that the great variability in the phonological shapes of the Hindi numerals -- a characteristic as striking as their morphological complexity -- may bear on the question which has been posed. Some degree of free

of free variation, of the type of English [fykənàmiks ~ ékənàmiks], may, of course, be expected to characterize languages in general. Hindi, as a language which has numerous geographical and social dialects, and which is not afflicted with a very long-established or rigidified literary standard, will show a certain amount of such free variation, even in the most homogeneous style. It is surprising, however, to find such an extreme range of variation in the numeral system, going much beyond that found in other types of words. It is clear that if these numeral forms were learned simply by memorization, and produced out of memory without any other conflicting factors, then the multiplicity of alternate forms could never come into existence. This appears to indicate that factors other than memory ARE to be considered in the description of the Hindi numerals.

One factor seems to be the influence of adjacent forms in the sequence of counting (cf. fn. 4). Consider again the allomorphs of SEVEN saat in the numeral '67', variously pronounced as satsaTh, sarsaTh, and saRsaTh. Of these pronunciations, the first seems to be historically the most conservative, with preservation of t. The second shows a weakening of t to r, a process which has parallels elsewhere. The third, however, shows a retroflex articulation, which has no historical justification if we consider this word as an isolate. Looking at it in the counting sequence, however, we observe that

saRsaTh '67' may well have arisen on the model aRsaTh '68', where the retroflex is historically 'right'. Can we say, then, that the speaker who uses saRsaTh is doing so in conformity to a RULE, perhaps one of 'distant assimilation of retroflexes'? Such a statement does not seem profitable, since in fact this rule would have as little generality as the alternative ad-hoc statements that saat → saR- before '60', or simply that '7' + '60' → saRsaTh.

There are, however, examples of other kinds. The numeral '63' is given as treesaTh or tirsaTh by most published sources, but as teesaTh by Harter. A simple assimilation to neighboring baasaTh '62' or following causaTh '64' cannot be postulated in this case. Instead, we may suppose that Harter's informant was using a rule which has already been recognized, namely THREE tiin → tee-, but that he applied it in an expanded set of environments -- not only before TEN and before '20', '40', '70', but before '60' as well. The production of such forms as teesaTh '63' may thus be understood not in terms of a new rule, but rather in terms of the rules given above, with some reshuffling in the statement of environments.

As college students of elementary Hindi can readily attest, these numerals are hard to learn, and they cannot be that much easier for the native speaker. In either case, much must be memorized. Some things are easy to remember, e.g. that saaTh '60' becomes -saTh after all elements; because this is so easy,

not a single source shows any variation in the shape of this -saTh. But it is hard to remember that THREE tiin becomes tee- in certain environments, tai- in others, tir- in still others. When memory becomes confused, a rule like tiin → tee- may be applied in a novel environment. But such changes cannot be explained unless we accept that allomorphic rules DO exist for the production of these forms.

My conclusion is this: Where linguistic data are extremely complex, the simplest description, measured by economy of symbols, may be a simple list. But such a list is not the only possible psychological reality. Where partial similarities of meaning and phonological shape exist between forms, they can be perceived by the native speaker as well as by the linguist. These patterns, messy as they may be, can be formulated as implicit rules in the head of the speaker, just as they can be formulated as explicit rules by the linguist. The amount of variation which can be observed in the Hindi numerals is in part a manifestation of those implicit rules, and of their unusual complexity.

Footnotes

[1]Such grammars have recently been published for Dutch by Van Katwijk 1965, for English by Brainerd 1966, and for Chinese by Kelkar 1966.

[2]Transcriptions are normalized for comparability with that used here for Hindi.

[3]A glance at some grammars of modern Indo-Aryan languages indicates that such systems occur at least in Bengali, Marathi, and Punjabi, as well as Hindi and Gujerati. They do not occur in the Dravidian languages, nor in Singhalese -- which, though Indo-Aryan, is spoken far to the south, in Ceylon.

[4]The notation used here for Hindi words is that of Harter 1960.

[5]In general, these forms fall within the range of variation which is attested in textbooks of Hindi. A few, however, may be novel. For '41', ikyaaliis, instead of the commonly reported iktaaliis, may have arisen by analogy with following byaaliis '42' (one is reminded of the well-known contamination of numerals in other languages, e.g. PIE *kwetwó:res, *pénkwe > Latin quattuor, quinque, English four, five -- to cite Bloomfield 1933:422-3.) For '44', cãjtaaliis, instead of unnasalized cautaaliis, seems to reflect influence of neighboring teetaaliis '43' and päftaaliis '45'. For '49', unancaas stands alongside more standard uncaas; cf. Punjabi uninjaa (Hares 1929). The forms unahtar '69' and ikahtar '71' diverge from standard

unhattar and ikhattar through a modified application of the rule which drops medial a's in the prototypes *unahattar and *ikahattar. For '75', pichyattar, by comparision with standard pachattar, shows inserted y on the model of following chyattar '76', as well as the effect on the first vowel of the palatals which adjoin it. For '85' and '95', picyaasii and picyaanbee, alongside more standard pacaasii and pacaanbee, again show y on the model of chyaasii '86' and chyaanbee '96'. For '89', unaanbee shows the element un- 'minus one', which is normal in '19, 29, 39, ... 79' -- though the textbooks give navaasii; Punjabi influence is again possible. The numbers '91' through '99' show an element -nbee '90', where most sources give -nvee.

[6]The minus signs indicate that, although an element belongs to the general class of UNITS, it does not belong to the particular sub-class.

[7]The dash is used here as a concatenation symbol.

[8]The morphophonemic symbol X is here introduced, with the meaning 'doubling of adjacent single consonant' (specified by morphophonemic rule 1, below).

[9]It is here understood that the element on the left-hand side of the rule will remain unchanged if no environment on the right-hand side is applicable; thus eek in isolation remains eek.

[10]This morpheme does not, of course, occur before any other decade complex.

[11]The morphophonemic symbol M means 'nasalization of preceding vowel' (by rule 2 below).

[12]This rule is valid for Sanskrit, and thus for many learned compounds borrowed into Hindi, e.g. dayaanand 'bliss of grace' (used as a personal name)<dayaa 'grace' + aanand 'bliss'. In Hindi verbal morphology, however, a y is inserted in such sequences, e.g. aayaa 'he came' < aa- come' + -aa '3sg. masc. preterit'.

[13]The y is non-contrastive in this position, in the language as a whole, and could in fact be omitted from phonemic notation.

REFERENCES

Bloomfield, Leonard. 1933. Language. New York: Holt.

Brainerd, B. 1966. Grammars for number names. Foundations of Language 1.109-33.

Cardona, George. 1965. A Gujarati reference grammar. Philadelphia, University of Pennsylvania Press.

Garvin, Paul L. 1961. Evaluation procedure in linguistic analysis. Studies in Linguistics 15.62-9.

Hares, W.P. 1929. An English-Punjabi dictionary. N.p.

Harter, J. Martin, et al. 1960. Hindi basic course. Washington, Center for Applied Linguistics.

Kelkar, Ashok R. 1966. A note on standard Chinese numerals. Indian Linguistics 26.196-202.

Kellogg, S.H. 1938. A grammar of the Hindi language. 3rd ed. London, Routledge & Kegan Paul.

Sharma, Aryendra. 1958. A basic grammar of modern Hiudi. Delhi, Ministry of Education.

Van Katwijk, A. 1965. A grammar of Dutch number names. Foundations of Language 1.51-8.

Voegelin, C.F. 1962. The generative criterion for deciding alternatives in descriptive linguistics. Studies in Linguistics 16.42-8.

Whitney, William Dwight. 1889. Sanskrit grammar. 2nd ed. Cambridge, Mass., Harvard University Press.