

Proto-Linguistic Variation: A Link between Historical Linguistics and Sociolinguistics

Author(s): Joseph H. Greenberg

Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society (1989), pp. 91-101

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via <http://linguistics.berkeley.edu/bls/>.

The Annual Proceedings of the Berkeley Linguistics Society is published online via [eLanguage](#), the Linguistic Society of America's digital publishing platform.

Proto-linguistic variation: a link between
historical linguistics and sociolinguistics.

Joseph H. Greenberg
Stanford University

This study resumes and expands some of the topics discussed in Greenberg (1978) on the stages of the definite article and Greenberg (1981) which is devoted to two additional examples of Stage III articles, Nilo-Saharan k- and Penutian -s. In the present context, I will be especially concerned with an aspect of the Stage III article which only received incidental treatment in the earlier studies, namely, the fact that the distribution of the forms with and without the article shows a certain kind of randomness in relation to the languages and dialects of the linguistic stock in which it is found. I will also seek to show that this particular kind of randomness of distribution, to be defined and illustrated later, also exists in regard to other grammaticized elements than definite articles in their later stages of development, and indeed is not confined to grammaticized markers, but also extends to variants of morphophonemic origin and perhaps even to those arising from purely phonetic changes. Finally I will suggest that the sociolinguistic and linguistic situations which give rise to such variation are distinct from both the areal and the genetic types to which we are accustomed and thus constitute a third type of phenomenon with both historical and sociolinguistic implications.

In illustrating what is meant here by random variation, I will take as an example the Nilo-Saharan prefix k- which in my first study of languages in that family was called moveable k- (Greenberg 1966). The reasons for considering this the continuation of a Stage I article similar to the English definite article followed by a Stage II article (roughly speaking combining the characteristics of a definite and indefinite article) are set forth in Greenberg (1978, 1981) and are not repeated here. From the point of view of distribution over languages in a particular linguistic stock, what is here called random variation has some, or usually all, of the following four characteristics as illustrated by Nilo-Saharan k-:

1. Random distribution of forms with and without k- over separate branches of Nilo-Saharan, e.g. Maba eri, Songhai kuri, Daza gere 'blood'. Here the languages cited belong to separate branches of Nilo-

Saharan at the deepest level of subclassification. Likewise within the same branch of Nilo-Saharan similar phenomena are found. In Table 1 below some etymologies are to be found which are represented both in Nile Nubian and Bari which belong to separate subgroups of the Eastern Sudanic branch of Nilo-Saharan.

Nile Nubian		Bari	
guar	'ant'	ki-gwur-te	'ant'
ur	'head'	ur-et	'crown'
aru	'rain'	k-are	'river'
gu-mur	'neck'	mur-ut	'neck'

Table 1.

It will be noted that all four possibilities are represented, both languages with reflexes of *k-, 'ant', neither language, 'head/crown', presence in either one without the other, e.g. 'rain/river' with k- in Bari but not in Nubian and 'neck' with k- in Nubian but not in Bari. This represents the usual situation and numerous additional examples could have been given.

2. Such random variation can also be found even among dialects of the same language. In Table 2 examples are adduced from Moru, a language of the Moru-Madi subbranch of the Central Saharan branch of Nilo-Saharan. Regarding the three dialects cited, Tucker (1940:3) states "The three dialects Miza, Kediru and Moroǎndri are so similar as to be almost identical." Yet even at this low genetic level similar variations occur.

Miza	Kediru	Moroǎndri	
k-umu	k-umu	umu	'fly'
k-ini	ini	ini	'skin'
k-ari	k-ari	k-ari	'leprosy'

Table 2.

3. There are even instances of free variation within the same dialect. For instance in the Keliko dialect of Moru kari and ari are given as variants of the same word for 'blood'. Of course, as with numerous statements regarding free variation, one does not know in the absence of sociolinguistic studies the nature of this variation, whether between individuals, social strata, kingroups or geographically determined entities. Nor does one know

whether the variants are differently distributed in accordance with such criteria in lexically distinct items.

4. Finally it should be noted that, usually at a very low genetic level, particular dialects or languages show a distinct or even exclusive preference for one variant over another. Thus in Table 2, Miza has k- forms in all three instances. An examination of the relatively brief comparative vocabulary in Tucker (1940) shows only rare instances in which Miza shows forms without k-.

Although the instances just cited show a kind of functionless choice of variants among related languages and dialects, this need not always be so. Four types of acquisition of new functions can be distinguished, once more illustrated from Nilo-Saharan k-. One of these, however, the second one listed, could not be found in this instance but will be illustrated later.

1. There are sporadic cases in which what would otherwise be homonymous forms are differentiated by the choice of one of the variants as against the other.

An example from dialects of Moru, the Central Sudanic language which is the source of the examples in the foregoing table, will illustrate this possibility. In both Miza and Kediru, the root ari without k- means 'bird', while k-ari, otherwise identical means 'blood'. It is interesting that in the third dialect of Table 2, Moroñdri, the two are differentiated in a quite different fashion. In this dialect ari is the word for 'blood' while ari-va means 'bird'. The suffix -va in this latter form is a diminutive.

2. Sometimes both forms are retained but undergo a secondary semantic differentiation. As noted earlier, I have been unable to find an example of this involving Nilo-Saharan k-. An example in a quite different context is English 'shade' and 'shadow', the first deriving from the nominative singular scadu and the latter from the oblique stem scadw- after the dissolution of the Old English case system.

3. There may be sporadic new grammatical functions. Thus in Kanuri, a language of the Central Saharan branch of Nilo-Saharan, in just one root, the form with k- prefix, k-am is used as a singulative 'person', while the form without k-, am is a collective meaning people. The most closely related language to Kanuri, Daza, uses am as the general root for 'person' and has no form with k-. In Karimojong, a

language of the Eastern Nilotic subbranch of Eastern Sudanic, examples of this kind are more numerous but the formation is still not productive, e.g. (e)-ki-twani 'a single scorpion'; (ni)-twani 'scorpions'. In these forms, the prefixes (e)- 'singular' and (ni)- 'plural' are new second stage articles.

4. However, productive new grammatical functions are sometimes found. Since the k- prefix as a former article deriving from a demonstrative occurs only on nouns, it is reinterpreted as simply a mark of nominality and as such becomes a productive derivational element forming verbal nouns. An instance is Sara, a language of the Central Sudanic branch of Nilo-Saharan. As an example we may cite the verb root usa 'to eat' from which a verbal noun k-usa 'act of eating' is formed.

We see from the above example of reinterpretation, that one possibility is that the formative spreads and acquires new functions. Another possibility is the opposite, which we may call contractive. In such instances the element does not spread and acquire new functions (expansion with resemanticization) but survives in just a few examples in which it has been lexicalized, in the sense, that, from the synchronic point of view, it has been incorporated as an indistinguishable part of a lexeme.

However even in a purely synchronic description, it may be noted that there is, statistically, more than a chance correlation between certain grammatical or semantic classes, and certain sequences of phonemes which thus form a submorphemic entity.

These possibilities, namely of expansive resemanticization and contractive desemanticization can be illustrated from the Chibchan - Paezan suffix -kwa. The original meaning seems to have been 'egg, nut or other similar round object'. It survives in a few instances as a lexical item with this meaning e.g. Terraba (Costa Rica) gwa 'egg', and Cuna (Panama) kwa-(kwa) 'nut'. In this latter case we find -kwa also as a suffix on round objects but also extended to a considerable part of the nominal vocabulary so that it is difficult to assign it a single, definite meaning synchronically.

As an example of expansion with resemanticization we may cite the instance of Millcayac, a Chibchan language of Argentina in which -gue is a productive derivational suffix forming verbal nouns, e.g. cheri 'to give'; cheri-gue 'gift'. Thus it has ended in this case with a derivational function quite comparable to that of Nilo-Saharan k- in spite of its very different

origin.

The second type of development that -kwa undergoes in some Chibchan languages is incorporation into a system of numeral classifiers and indicating small round objects. As noted in Greenberg (1972), the most common source of a general classifier is that for round objects. In some cases, after acquiring this function, the very fact that it can appear with all nouns, even if only in quantifying phrases, makes it once more a possible general marker for nouns. However, given the syntactically limited use of numeral classifiers, it is the development which is less likely to occur, particularly if the original order is numeral-numeral classifier. Under these circumstances it may survive in isolated cases as a mark on a few numerals and can be considered from the synchronic point of view simply as part of the numeral. Still its presence in several numerals, and even more powerfully, the comparative evidence, will lead to the correct historical interpretation.

In the case of Chibchan -kwa, this occurs in some languages. For example in Kagaba, a Chibchan language of Colombia, it is found in just four numerals, mai-gua 'three'; ku-gua 'seven'; abi-gua 'eight'; aita-gua 'nine'. With these we may compare Margua, another Chibchan language of Colombia, which has mai 'three' and avi 'eight'.

In both cases discussed above, Nilo-Saharan k and Chibchan -kwa, we are dealing with grammaticalized elements in which the variation is between presence or absence of the item in question. It could have been further illustrated from instances like petrified honorifics or diminutives.

However, the variational phenomena earlier enumerated are not confined to grammaticalized morphemes which alternate between presence and absence. Similar distributional properties across language and dialects, and in some instances, similar examples of semanticization are to be found in the case of morphophonemic alternants. Since in such cases, the item in question did not previously have a meaning, we may talk about semanticization rather than resemanticization. Moreover, the variants, typically two in number, both have overt phonological expression in contrast to the examples treated above.

An important source of such variation is the breakdown of vowel harmony systems. There are two main types of development. One is through merger of pairs of alternants resulting typically in so-called neutral vowels as in the instance of Mongolian i which

functions both as back and front vowel. The usual assumption is that this is the result of a change by which its former back partner y becomes i, thus eliminating the alternation. It is easy to see how further mergers may finally destroy a vowel harmony system completely and this has indeed happened in Kerek and Aliutor, both dialects or perhaps separate languages closely related to Koryak, a Chukotian language. In Koryak itself as well as in Chukchee and the more distantly related Kamchadal, a system of high-low harmony still functions. Another course of events, however, gives rise to the kind of variability that we have already seen in the case of grammaticalized elements. An example is that of the East Mongolian languages, all of which are spoken in China. It is clear that Proto-Mongol in regard to vowel harmony was essentially like Classical Mongolian and the present day Western Mongolian languages such as Khalka, Buriat and Kalmuck. The back-front harmony of these languages is stem-driven in that the vowel of the stem remains constant, and the derivational and inflectional affixes, which all follow the stem, vary in regard to backness or frontness depending on the stem.

In the Eastern Mongolian languages, the stem vowels remain basically in their inherited form, but the vowels of affixes tend to have a single variant, each one being an independent case. Thus in a particular language, the front variant can be found in one affix, but the back in another, or they may be in free variation, a variation which thus has nothing to do with the vowel of the stem. The breakdown of harmony may here be attributed to contact with Chinese. In Table 3, a number of inflectional morphemes which originally had the vowel a with stems in back vowels and e in stems with front vowels is shown for the four Eastern Mongolian languages Baoan, Dagur, Dunsian and Monguor.

	Baoan	Dagur	Dunsian	Monguor
Plural	le	--	la	--
Ablative	se	se	se	dza
Locative	re	aare	--	--
Instrumental	gale	gala	--	--
Comitative	--	--	le	la
Causative	ge	gaa, gee	ga	ga, ge
Past Participle	sang	sen	sen	dzan

Table 3.

As can be seen we have a kind of cross-linguistic and internal language variability in principle basically similar to that which was encountered in regard to Nilo-Saharan k- and Chibchan -kwa even though the variants are of phonological rather than morphological origin.

A very similar situation obtains in regard to the so-called Iranicized Uzbek dialects in which, presumably under Iranian influence, the Turkic vowel harmony system, which except for the existence of distinct high back and front vowels has essentially the same structure as Mongolian, has broken down in affixes with results that are basically similar to those found in Eastern Mongolian languages. In neither of these instances is there any semanticization of variant forms. That morphophonemic variants can be utilized to express grammatical distinctions is, of course, well known from the example of German umlaut in which noun pluralization, the expression of the subjunctive in verbs and that of the comparative and superlative of adjectives are expressed by umlauted vowels although generally as a subsidiary mark along with affixes of the usual kind.

An interesting example of the breakdown of a vowel harmony system based on height, in a system which still functions in a few limited aspects of the grammar, e.g. the numeral classifier system, and the third person singular prefixed pronominal object of the verb, occurs in Gilyak.

The Gilyak system of vowel harmony is shown in Table 4.

High	i	y	u
Low	e	a	o

Table 4.

If we compare the two main dialect areas, that of the Amur basin in Siberia, and that of Northern Sakhalin, we see along with numerous instances in which both dialects have generalized the low variant, or both have retained the high variant, a considerable number of cases such as those in Table 5 in which one dialect has chosen the low and the other the high variant.

Amur	Northern Sakhalin	
yl	al	'mouth'
park	pyrk	'only'
mut	mot	'pillow'
nik	nek	'recently'

Table 5.

In general the Amur dialect prefers the high variant and Northern Sakhalin the low variant but there are a fair number of exceptions. In addition there are instances in which both variants are found in both dialects but have differentiated their meanings, e.g. vi-, 'go, walk'; ve- 'run (of animals)'; lax 'cloud', lyx 'rain'. In one instance we even have an incipient grammaticalization, nog 'to be fragrant' (intr.) and nugnug 'to smell' (tr.). In all the forms just cited the two dialect areas agree.

In addition to morphological elements and morphophonemic alternants, it seems likely that variants developing out of word sandhi can give rise to a similar pattern of cross-linguistic and intralinguistic variation. A well-known instance is the so-called s-moveable of Indo-European which appears preceding roots in unvoiced stops, r, l, m and n. An example is Latin tegere 'to cover' as against Greek stégos 'roof' which also has a variant tégos in the same language.

Can a similar pattern of distribution result from variants produced by sound change? The usual pattern is either clearly genetic or areal. Thus the Indo-Aryan change of *e > a, merging with original *a which is found only in this branch of Indo-European is clearly genetic. A classic instance of an early areal feature is found in the case of "incomplete satemization."

The fronting of front velars to sibilants found in the satem branches of Indo-European, namely Indo-Aryan, Balto-Slavic and Albanian is only complete in Indo-Aryan and probably Armenian. In Balto-Slavic there are instances of velar reflexes in particular forms, in regard to which the languages differ from each other and even show variation within the same language. Thus corresponding to Sanskrit sru- 'to hear' we find Old Church Slavic slu-ti as expected but in Baltic Lithuanian klausy-ti, Lettish klausi-t and Albanian guhëm 'I hear'.

Corresponding to Sanskrit ásman 'stone' we find forms with velars in both Baltic and Slavic: Old Church Slavic kamy, Lithuanian akmuo and Latvian akmens. However, in Lithuanian there is the doublet asmuo with the differentiated meaning 'edge, blade'. In the word for 'dog', Sanskrit śvan corresponds to Lithuanian szuo as expected but within Latvian we find the surprising variation suns 'dog' but kuna 'bitch'. Sanskrit śmasru 'beard' is connected etymologically to Lithuanian smakra, Latvian smakrs 'chin' and Albanian mjakrë 'beard'.

These and other examples show some degree of randomness in the distribution of the palatal and sibilant but on the whole exhibit an areal pattern which, as noted above, suggests an eastern origin for the sound change with imperfect propagation westward among the Indo-European dialects.

An instance which does seem to show the type of distribution seen above for morphological and morphophonemic variants are the reflexes of the reconstructed syllabic r of Proto-Indo-European, in Greek ra or ar and Germanic ru or ur. Brugmann in the second edition of his comparative grammar (1897:4) notes that there is no satisfactory solution to this alternation. In his later summary grammar (1902:131) he says "Probably a PIE difference in pronunciation is the reason." This variation in regard to Greek is illustrated here by a few examples. Athenian kardía 'heart' corresponds to Ionian and Homeric kradía. Within the Athenian dialect itself we find kárta 'very' but kratús 'strong' and the variant past passive participles dartós, dratós 'flayed'. In Pindar who wrote in the Theban Aeolic dialect we even find thrasús kardía 'bold heart' in which the reflexes ra and ar are found in the same phrase. The choice was probably because of the meter but that both variants were available to him is significant.

An interesting case of secondary semantic differentiation is the existence of the two forms thrásos and thársos for 'courage' in Athenian and the Standard Koine. Aristotle (Eudemian Ethics 1234b.12), after his discussion of the golden mean as lying between a particular vice and its corresponding virtue, distinguishes another case, namely that the excess of a good quality is likewise condemned and is to be contrasted with its possession to a moderate and fitting degree. His example is thársos 'courage' vs. thrásos 'foolhardiness'. In a similar vein Ammonianus, a lexicographer of the second century A.D., says that thársos is said of human beings, that is 'reasonable courage', as against thrásos, the unreasoning courage of animals.

I believe that we are to envisage the third alternative as distinct from the genetic and the areal but describable from the sociolinguistic point of view as follows. The protolinguistic community showed variations reflecting changes which were just in progress as it began to break up. They were distributed idiosyncratically across small groups and even individuals. Within each group that later became a separate language there was a specific

distribution of subject, of course, to later analogic changes and often with an inherited preference for one variant or the other. We may compare these to the 'founder' groups of population geneticists. No particular subgroup represents a perfect sample of the original population in language just as in regard to gene frequencies.

The ensuing results take the form of a random distribution, continuing often for a surprising length of time, that we have found in the examples illustrated in this study. Such ongoing changes in the proto-language only affect a small part of the total linguistic structures involved but it may be present in any aspect of the language.

Examples such as those discussed are then not amenable to either genetic or areal explanation. They present a problem to the comparatist seeking to reconstruct a total and uniform ancestral linguistic system. If, however, we take seriously the facts about linguistic variation observed and studied by sociolinguists, we will not find such phenomena surprising. It is the thesis of this paper that the kind of random distribution of competing forms discussed here and which could be illustrated by many more examples, is reasonably accounted for by the sociolinguistic factors just mentioned.

References

- Brugmann, Karl. 1897. Grundriss einer vergleichenden Grammatik der indogermanischen Sprachen. 2nd ed. Vol. 1. Strassburg: Truebner.
1902. Kurze vergleichende Grammatik der indogermanischen Sprachen. Vol. 1. Strassburg: Truebner.
- Greenberg, Joseph H. 1966. The Languages of Africa. Bloomington: University of Indiana Press.
1972. Numeral classifiers and substantival number: problems in the genesis of a linguistic type. Stanford: Working Papers in Language Universals:9.
1978. How does a language acquire gender markers? in J. H. Greenberg ed. Universals of Human Language 3:47-82. Stanford: Stanford University Press.

- Greenberg, Joseph G. 1981. Nilo-Saharan movable k-
as a Stage III article (with a Penutian
parallel). Journal of African Languages and
Linguistics 3:105-112.
- Tucker, Archibald N. 1940. The eastern Sudanic
languages. London, New York: Oxford University
Press.