Festschrift for
Henry Hoenigswald

On the Occasion
of his Seventieth Birthday

Edited by
George Cardona
and
Norman H. Zide

Gunter Narr Verlag Tübingen
Linguistic areas: getting at the grain of history

Peter E. Hook (University of Michigan)

A perennial problem in the study of linguistic areas is the indeterminate nature of the data adduced to link observed effects with inferred causes. The present paper reviews four possible sources of structural parallels in adjacent languages: 1) coincidence, typological harmonics, genetic relationship, and cultural contact. The kind of area which select cultural contact in cause (and which exclude the three other possibilities) is identified and instantiated from field work in South Asia.

Prolonged periods of interaction among speakers of languages belonging to different (or at least not demonstrably related) families may lead to convergence of structure (or other characteristics: see Hook 1982) which is not explainable except by reference to that interaction. This opens up the attractive possibility that we may use the study of linguistic areas (together with other kinds of tacit evidence) as means to inferring prehistorical cultural contact among such groups. As an instance of such inference (controlled by independent historical testimony), the absence of an infinitive in British Romani may be cited:

\begin{align*}
\text{kam are-as te jau khere} & \quad \text{(Sampson 1926:IV\textsuperscript{I}--131)} \\
\text{want-S IMPRF that go-S home-LOC} & \\
\text{hte\text-sinda te de j} & \quad \text{(Sampson 1926:III--191)} \\
\text{rain go-S that give-S} & \quad \text{It's going to rain.}
\end{align*}

This syntactic peculiarity is an indication of protracted residence in the Balkans. We can be fairly sure of this interpretation since all other forms of Indo-Aryan have an infinitive; the lack of one distinguishes Balkan languages from their neighbors; and there are historical records of the Gypsies' sojourn there. In this paper I attempt a preliminary discussion of a basis for the evaluation of convergence phenomena as indicators of cultural contact when there is no explicit historical record.

In his landmark study of South Asia as a linguistic area, Colin Masica (1976) makes an important methodological point: In a linguistic area the parameters along which languages of different families are similar must in fact define that area. That is, to establish a given geographical region as a linguistic area we must show that its languages have features in common which distinguish them as a group from languages outside the area. Thus, for each feature that he discusses, Masica traces its presence (or absence) in ever widening circles until he reaches the line that separates languages having it from those that do not. By following this procedure, Masica discovers a very surprising thing, surprising at least to most Indologists: The syntactic (and semantic) features he selects do not define South Asia as a linguistic area. Rather they show a commonality in structure that includes language spoken
in vast areas of the Soviet Union (the Central Asian Socialist Republics), Korea, and Japan, an area which Masica has more recently termed the 'Indo-Turanian' linguistic area.

What are we to make of this? Linguists (at least those linguists who are Indoologists) are familiar with the idea that Indo-Aryan, Dravidian and Mundari languages may have come to resemble one another as much as they do by virtue of the millennia of close cultural contact for which there is direct historical evidence as well as obvious manifestations in every sphere of life: religion, philosophy, music, cuisine, etc. Are the historical incursions of White Huns and Magyars into India of the same order of magnitude and intensity to explain a degree of linguistic convergence between South Asian and Central Asian languages not much different from that found between Indo-Aryan, Dravidian and Munda in South Asia itself?

In fact, a great deal of attention need not be paid to this question for the same degree of convergence (or similarity) can be found between the Altiace of the seventh or eighth centuries A.D. (see Appendix A) and inscrptional Tamil of the same or even earlier periods (Appendix B). Considering the enormous distance separating the site of the Orkhon inscriptions from Tamilnadu and the absence of historical records of contact between Altiace and Dravidian speakers prior to the sixth century, we must either assume prehistoric contact (or very remote genetic relationship) to be the explanation or consider whether these areas may owe their existence to mechanisms other than these. It is the nature of such 'other mechanisms' that must be understood before we know how to assess the historiographical importance of the Indo-Turanian linguistic area.

There are two explanations for linguistic areas which do not appeal to contact or to remote genetic relationship: (1) coincident, and (2) typological harmonics. The likelihood of simple coincidence's being the explanation cannot be determined without a better understanding of typological harmonics. That is, unless we know, in general which typological features pattern (or tend to pattern) with which, we cannot judge what the probabilities are of a given constellation of features arising purely by chance. For instance, if we know that the four features (out of the five used in Masica 1976) that show linguistic parallels between South and Central Asia are both independent and roughly equipollent (i.e., equally likely to be present or absent), then the likelihood that a Central Asian linguistic area casually disconnected from South Asia would show the existing convergence simply by chance is one in sixteen. If some or all of those four features tend to pattern together in the world's languages then this rather remote probability of accidental convergence improves. And improves rather markedly. How else are we to explain the presence of languages of typically Indo-Turanian structure in the western cordillera of South America (at the very antipodes of South Asia)? Ecuadorian Quechua, for instance, shares all the syntactic and semantic features with South Asian languages that Orkhon Turkic does (see Appendix C).

The work of Greenberg and others has shown that certain typological features tend to cluster together. If a language has a basic word order in which the object precedes the verb, then it is much more likely to have postpositions than it is to have prepositions, auxiliaries, and verbs that are more likely to follow objects than to precede them, etc. Thus, the agreement among languages along several of these parameters cannot be considered cumulative evidence for establishing a linguistic area, or, to put the matter more precisely, such agreements cannot be considered additive. Rather, a complex formula has to be worked out specifying the weight to be given to each individual parallelism. Recognition of this principle is implicit in the organization of Masica's discussion of South Asia as a linguistic area where word order phenomena are treated together in a single chapter, as is the 'dative subject' construction with the absence of a verb 'to have'. To work out such a formula will require that we determine ratios of incidence for each feature of interest to us in the total stock of the world's languages (or in some controlled sample of them): see Tomlin 1979 and then, assuming them to be independent, compute the expected frequency of each combination of them. Comparing these expected frequencies with those actually observed will tell us how much to discount for typological harmonics, even if there is no satisfactory explanation for those harmonics in linguistic theories presently available to us.

Let us take a very simple example of this: In Appendix II of his pioneering essay "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements", Greenberg classifies a largeish number of languages according to different combinations of basic word order, post- vs. prepositions, adverbic before noun vs. noun before adverbic, etc. The listing of languages and groups of languages he surveys totals approximately thirty-three lines of type. Taking three of these variables individually we find the following ratios of incidence:

A. SOV word order 13 39% SVO word order 14 42%
B. postpositions 17 52% prepositions 16 48%
C. adjectival noun 12 36% noun-adjective 21 64%

(The integers are the number of lines of type. They do not total 33 for variable A because I am not considering verb-first word-order here.) Assuming these variables to be independent of one another, we can calculate expected ratios of incidence for each combination. For simplicity's sake I will find the values for A into B and for A into C (the integer following the figure for percentage is the expected number of lines):

I. AB 20% 7, AB 22% 7, A-B 19% 6, A-B 20% 7
II. AC 14% 5, AC 15% 5, A-C 25% 8, A-C 27% 9

The observed percentages and numbers of lines for each of these combinations are:

I. AB 36% 12, AB 12% 4, A-B 3% 1, A-B 27% 9
II. AC 18% 6, AC 12% 4, A-C 21% 7, A-C 27% 9

Comparison of these two sets of figures shows that the 'basic word order' of a language is much more closely tied to post- vs. preposition than it is to the relative order of adjectives and noun. That is, features A and B tend either to be both present or both to be absent to a stronger degree than do features A and C. Therefore, agreement on the latter between languages is more probative in the demonstration of a linguistic area than is post- vs. preposition and accordingly should be given a greater weight in the calculation (assuming that basic word order has already been given a full vote).

However, there is a limitation on this procedure which may be severe enough to dissuade anyone from undertaking the arduous task of determining separate frequency ratios for a large sampling of syntactic and semantic traits: it is possible that the world is not big enough, that the number of coexisting human languages is too small to give a representative sample of the full potential of man's language faculty and that cultural contacts are too frequent for us ever to know what the inherent typological characteristics of human language
would be if individual languages were allowed to develop undisturbed in pristine isolation.

In short, all languages may belong to a single linguistic area, namely, the Earth.

Such 'boss mots' aside, there is another reason for not undertaking mammoth typological studies of the world's total stock of languages (interesting as such studies might be in themselves): From the point of view of demonstrating linguistic areas they are simply not necessary. Rather if we turn to detailed cross-dialect, fine-grain comparisons we can show a pattern of convergence that will conclusively put to rest the two-headed bogey of coincidence and typological harmonics and which will also distinguish contact phenomena from common rections associated with remote genetic relationships (eg, North-Central). If a statistical study of a given feature of syntax or semantics shows a regular correlation with distance, and, especially, if this correlation is independent of genetic groupings and subgroupings, then we have a prima-facie case for contact as cause.

Recently I have conducted field research on the geographical distribution of aspiscopal contrasts in South Asian languages. I happened to include a pair of complex sentences on the questionnaire which show a regular progression in the frequency ratios of subordinate-main as opposed to main-subordinate clause orders as one proceeds from north to south within western Indo-Aryan (Sindhi, Kutchi, Gujarati, Marathi). (See figure.) Subordinate-main order is typical of Dravidian, main-subordinate, of Pramesh. Since we are no longer dealing with a simple binary (yes-no, present-absent) value for a feature, but with a finely modulated correlation of frequency with distance, coincidence is ruled out. Ascribing the facts to typological harmonics simply puts the operative factor at one or two remote and still does not explain the regularity in the geographic distribution of that factor. Similarly, remote genetically based groupings and subgroupings do not vary continuously with distance; ie, the Stammbaum model would not predict that Marathi would share more syntactic and semantic features with Kannada and Tamul than the North Dravidian language Brahui does.

In effect, fine-grain correlations of this kind can be explained only by positing chains of bidialectal and/or bi-lingual speakers along which change in some feature of syntax or semantics is transmitted much in the manner distortions in a message are propagated and elaborated in the party game 'Telephone'. Such correlations promise to help explain some of the regularity of sound correspondences did for the development of historical-comparative linguistics. The difficulties in establishing such correlations have to do with the availability of relevant information. Before Labor linguists were rarely interested in gathering data on the relative frequency of competing forms.

Assuming that such data become available in the future, we may discover that the fine-grain processes and social interactions that lead to the clines in the frequencies of competing linguistic forms and structures may be relatively short-lived while their global effects may persist for centuries after them. For example, the Indo-Turkic linguistic area uncovered by Masica may be a kind of fossil, a trace of an area once knotted together by chains of bidialectal and bilingual speakers, but now subject, in its various parts (and from diverse directions), to new influences that have yet to obscure the large-scale parallels in structure established in prehistoric times. An ambitious survey of competing patterns in a number of linguistic dimensions carried out on a village-to-village scale along selected geographic arcs in South Asia and Central Asia is needed to answer this question.
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOC</td>
<td>accusative</td>
</tr>
<tr>
<td>Adj</td>
<td>adjective</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
</tr>
<tr>
<td>CB</td>
<td>calculative</td>
</tr>
<tr>
<td>CP</td>
<td>geographic participle</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>EMPH</td>
<td>emphatic</td>
</tr>
<tr>
<td>ERG</td>
<td>ergative</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>HWT</td>
<td>hortative</td>
</tr>
<tr>
<td>IMPER</td>
<td>imperative</td>
</tr>
<tr>
<td>IMPF</td>
<td>imperfect</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>INSTRU</td>
<td>instrumental</td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>N</td>
<td>noun</td>
</tr>
</tbody>
</table>

Notes

* This paper derives from one presented at the thirty-fifth annual meeting of the Association for Asian Studies (San Francisco, 1983) as part of a panel on South Asia as a linguistic area. The field work on which it is based was carried out in India and Pakistan starting in 1958 on research trips supported in varying degree by the Henry M. Huntington School of Graduate Studies, University of Michigan; the American Institute of Indian Studies; and the United States Education Foundation of India and Pakistan (Fullbright-Hays Faculty Research Abroad). Their financial and organizational assistance is gratefully acknowledged as is the help and cooperation of colleagues at the University of Delhi, the Central Hindi Institute (Agra), the Northern Regional Language Centre (Patna), the University of South Gujarat (Surat), the Oriental College at the University of the Punjab (Lahore), the Institute of Sindology (Kemoreh), the Pakistan Information Department and many other institutions and individuals in both countries. I am particularly indebted to Miss Kathleen Chaudhry of Lahore for her constant help, introductions and hospitality and to Dr. H.K. Guz of Suez for providing much needed data from Western Gujarat.

1 The first South Asians to reach the shores of England arrived well before the East India Company was founded, at the latest by 1550 according to Vansy-Fitgerald (1944:20). Their northwestern migrations began sometime before 250 B.C. (Turner 1927:27).

2 See Handbook (1930) and Solta (1958). The Baltic languages that lack the infinitive include:

   Modern Greek: θέλω να λάβω 'I want to go.'

   Bulgarian: произвеждах 'I produce' (Bulgarian 1935:236)

   Romanian: uite = cf. il viitor este 'look at the future' (Castras 1957:123)

   Albanian: mi nga tealaj 'I want to write.' (Solta 1980:212)

   Macedonian: се волим да си речение 'I want that it be true' (Lam 1952:285)

   Georgian-Bosnian: ziva ne na mi se / we wanted that we bring S.' (Solta 1980:214)

3 For an Arabic report of Gypsies among the Byzantines in the ninth century A.D. see Vansy-Fitgerald (1944:6ff.).
Appendix A: Orkhan Turkic

Orkhan Turkic, as known to us from inscriptions first discovered 250 years ago in what is today central Mongolia, shares with Dravidian the four syntactic and semantic parallelisms that Masica 1976 uses to define the Indo-Turanian linguistic area: (1) word order, (2) derived causative verbs, (3) conjunctive participles and (4) explicative compound verbs. Examples of these (unless otherwise indicated, from the eastern or southern faces of the Kül Tigin inscription) are:

(1) word order:

a. tufaq ay ilig in tur-dli
   governor ACC hand-INSTRU captured
   'He captured the governor by (his own) hand.'
   (E 38) SOV

b. tufik qara qamy rum horun
   Turk black entire people
   '...the Turkish common people...
   (E 8) AdjN

c. ini m kul tigin birji
   brother my Kül prince with
   '...with my brother, Prince Kül...'
   (E 26) Postposition

d. ozkan yli da ay yiqi ayaq ammis
   Ozkan mtn-from better at all not in
   'There is no better (place) than the O. Mountains.'
   (S 4) SMAdj

e. umay ek izi anam
   Umay like mother-my
   'My mother (who is) like the Goddess Umay...'
   (E 31) Rel/PN

(2) derived causatives (and transitive):

a. ol-de olar-kill-kil-come killir-bring
b. ar-be darim-make darum-make durem-make have made
c. eme adina il ni ban yurur-ti um
   them (DAT) wonder'til tomb build CAUS-POST-1SG
   'I had them build an extraordinary mausoleum.'

(3) The affix of the conjunctive participle (CP) in Orkhan Turkic has a number of forms. Chief among these are a vowel (-a, -i, -u, -e) and an ending -p or paru/pim. Like their counterparts in Dravidian (and unlike English having fed) the Orkhan conjunctive participles show great syntactic freedom occurring in imperatives and as well as in the narrative declarative where analogous constructions occur in English and most languages:

a. yuranq qetin kil -qa yang a elati
   armed whence come CP disperse CP sent
   'From where did the armed (ones) come and scatter (you) ?'
   (E 23)

b. hem kure-bi em
   this (ACC) see-CP know IMPER
   'See this (inscription) and learn (from it) !'
   (S 12)

The same-agent constraint (Hook 1976) applies (although such agents do not have to be the same in structure):

c. yof(q) avar et yi-te yadryr yali
   tusunw ay -tur-um
   upwards homes lead-CP on-foot trees climb CP climb-made-ING
   'They had them climb upwards on foot leading their horses and clinging to the trees.'
   (Tonsykouk I N 3)

(Here, it is clear from context that the leader of horses and the clinger to trees is not the 'I' of the account.)

(4) Explicator compound verbs in Orkhan Turkic, as in Dravidian, are formally indistinguishable from conjunctive participle plus finite verb constructions (cf. 3a):

a. tufik irar azapa ir in aq-bar bard-i
   (E 34)
   great (names) few men-INSTRU free-CP go -PST-3SG
   'The great Ikan ran away with a few (of his) men.'

The explicators are homophonous with lexical verbs:

b. bodun yar en sub-in s-baq tabri-yu-I
   people land-near water-their leave-CP China toward went
   '...people left their land and went and went toward China.'
   (Bilgii Kagan inscription, E 35)

c. qayun-in yirini aq-mi
   kagan their lose 'P-leave-PST
   '...they lost their kagan...'
   (E 7)

The alternation of the explicator with its absence (one of the two criteria for the classificatory 'compound verb'; see Hook 1977:336) is found in Orkhan Turkic under similar conditions:

d. kül tigin qan yil aq-yi yegurmi-ká ul-di
   Kül Prince Sheep year-LOC seven twenty-LOC fly-PST
   'Prince Kül died in the Year of the Sheep on the 17th day.'
   (KT NE)

Here the tone is matter-of-fact. Contrast this with:

e. e aq-har-d yax
   fly-CP go -PST -2-P1
   (KT SE)
   where the writer, Prince Kül's nephew, expresses his sense of loss (The sentence follows directly on: 'You used to nourish (the people) better than your beloved children and descendants.')

(5) Another feature common to Orkhan Turkic and Dravidian of the time is the use of the participle of the verb 'say (re-tell) as a complementizer with verbs of speaking and thinking and to introduce clauses of purpose (cf. Kuiper 1967):

a. olaray ti yam te-mis
   stay-2PL say-ing say-2PL
   '"Stay!" he said.'
   (Tonsykouk I N 10)

b. ukeu ti yam na ká qenqar bii
   many say-ing what DAT fear -PRS we
   'Why should we be afraid (of there being) many (of them) ?' (Tonsykouk IL W 4)

c. anuq id mit ti yam siiq
   (Bilgii Kagan E 25)
   caravan send-not say-ing campaign PST-1SG
   'I went to war (against them) because they did not send the (tribute) caravan.'
Appendix B: Inscriptional Tamil

The earliest Tamil prose of appreciable length is that found some fifty years ago in an inscription of some forty lines that was unearthed in Palliṇṭhukkůlvōli, a small village in Tanjore. The text, recording a grant of land, is considered a date before that of 550 A.D. (Zeisler 1964:6). From it we may illustrate the same syntactic and semantic features as have been exemplified in Appendix A.

(1) word order:

- a vijayam-pini-kurunakaram ... nilaṁ kṣopu-tu ... (nīnaṁ) teacher DAT-EMPH land give PST-we

To the same teacher V. . . we gave land. . .

- b tan nāi ... your district...

- c puruṣā-nilkutu-n ucca ... great four-smaller T-interior-OBL

. . . with at the great four boundaries. . .

- d (no comparatives available in the inscription)

- e kārya-muy pāḷi calī -graze portion

. . . the part (where) the calves graze. . .

(2) derived causatives:

While no clear instances are available from the Palliṇṭhukkůlvōli inscription, they may be found in earlier inscriptions (from 200 B.C.) found in caves:

- a ari yakṣo cāpatariṇī koṇī ti caus -CAUS-PST-3SG

'Shir Yaksha Canatarant . . . had (this cavern) cut.'

Compare this with the non-derived transitive in:

- b veḷḷai niyamāṅit ukkō... (nīnaṁ) citizens (s) cut

'The people of Velvai cut (the cavern)'

(3) conjunctive participles:

- a nāṭṭā ... irunukkam kas-tu roha-tu... (33-34)

members of nāṭṭā too royal-order see CP nameṛkiṇī

'And the members of the district assembly having seen the royal proclamation and having shown their deference...'

(4) As in Orkhton Turkic the morphology of the explicatory compound verb construction is indistinguishable from that of conjunctive participle plus finite verb:

- a api -yolai caey -tu korte-tu viru -uka... speak-T/leaf make CP give -CAUS leave-HORT

. . . prepare the proclamation and issue it.'

Note the use here as in Orkhton Turkic of a verb homophonous with the verb for 'leave' as explicator.

(5) The conjunctive participle of the verb en 'say' is used as a sentential complementizer (see Kuiper 1967)

- a tāṅkak ... am ... aviyolai cease -tu korte-tu viru -uka ... en ... you-PF.too speak-leaf make CP give CP leave-HORT-T say CP

nāṭṭā -kata i-rūnumāṅit ucca ... members DAT T-royal-order leave-INF

'The king has given an order to the members of the assembly that they too should. . . prepare a proclamation and issue it.'

Appendix C: Marathi and Quechua

Marathi is one of the Indo-Aryan languages most strongly Dravidianized in its syntactic and semantic structures. The examples adduced here show structural parallels between it and Quechua, the modern descendant of the state language of the pre-Columbian empire of the Incas. Convergence through cultural contact may in this instance be safely ruled out.

(1) word order:

- a mi tyāli patiye jī de  , -a SOV

taka pay-man kaluki-ta kha jī -a

'll give the money-ACC give-PRES-1SG

b sundar stri samak varni

pretty woman

- b sāvatā  sīrā Postposition

tytā  pā pāmi-yū gā pā takmē

'father-POSS-GEN friend-POSS-GEN farm-POSS

'My father’s friend’s farm...'

- d miḷaḥkukan hubh Postposition

toca phe inacumā...

'touch somebody amarā ...

'more from smart

'wiser than me. . .'

- e huṇnu i di lela par jī pāzhv Reji-PNP

juan cu-acha quilla-ta cachai

Juan-ERG give-PST letter -ACC send

'Send the letter that Juan gave.'

Sentential objects, in both Marathi and Quechua, often follow their verbs (see sentence 5-a below).

(2) derived causatives (and transitives):

- a mar -mār -mār -nār -warbā -warbā -nār -warbā -nār

die kill -CAUS

'have someone kill'
(3) Conjunctive participles:
- a ge'vi o hi ri-eca ni pish go-PST.1sg also 'Yam I went, too.'
- b kay-mam-pi wak-wam-pi this way also that way also
  'Even if he gets angry, I won't go.'
  *(I cannot find this pattern in Marathi.)*
- c p'ila-ri e pi-pish mara ri sha-cha nag ye-an hi kay uppog
  anger-give-on also no go-will-no
  'Even if you get angry, it's no use!'
- d Quechua: pi 'who?', p'ipi 'somebody'; ima 'what', ime-pi 'something'
  Marathi: kon 'who?', koni (kon + hi) 'somebody', koni hi 'anyone'
  ka'ha 'where', ka'ha hi 'anywhere'
  kiy 'what', k'iti 'some', something'
  *(absent in Quechua?)*
- e p'iti-hi gele five-also went
  'All five went.'

References


Emeneus, M. B. 1966. *India as a linguistic area*, *Language 32*: 1-16.


Lunt, Horace Gray, 1952. *A grammar of the Macedonian literary language* (Skopje: Družna kom.)

Markham, Clements R. 1972. *Contributions toward a grammar and dictionary of Quechua, the language of the Incas of Peru (reprint of 1864 edition)* (Oswestry: Bibliothég)

Marina, Colin P. 1993. South Asian languages: typological coincidence or areal convergence? Presented at the 35th annual meeting of the Association for Asian Studies.


Sampson, John, 1968. The dialect of the Gypsies of Wales, being the older form of British Romani preserved in the speech of the clan of Abram Wood (Oxford: Clarendon Press).


Tucano Matias, H., 1953. El español en el Ecuador: Revistas de filosofía española, anno no. 61 (Madrid).


Vanpaz, Honorato, 1940. Repaso sobre nuestro lenguaje usual (Quito: Editorial Ecuadoriana).

Verey-Fitzgerald, Brian S., 1944. Gypsies of Britain: an introduction to their history (London: Chapman and Hall Ltd.).
