APEX AND TERMINAL NODES IN THE LINGUIST'S TAXONOMY OF GENETICALLY RELATED LANGUAGES*

C. F. AND F. M. VOEGELIN

1. Terminal nodes: dialects
2. Objective criteria: dialect versus language
3. Subjective criteria: folk taxonomies
4. Internal genetic relationships: key languages versus large sample
5. External genetic relationships: the gap between language families

1. In our recent survey of languages of the world, there were some surprises; for example, we expected to find a coincidence of a single language and single dialect among some enclave communities of small population, but found virtually every separate language was reported to have or rumored to have had more than one geographically distinct dialect. (There were, of course, difficulties in reporting this in the score of fascicles which we published – – e. g. the fact of variety might be noted without labels for the different dialects or the labels might not distinguish between different places in which a given language was spoken.[1])

Our unconfirmed expectation of finding quite a few languages without dialectical variation was probably based on our experience with American Indian languages which nowadays often exhibit dialect leveling. For example, a quarter of a century before we went to study Tübatulabal, there were two distinct dialects spoken on two forks of a river east of its confluence in the Sierra Nevadas; but the North Fork Tübatulabal had moved in with the South Fork Tübatulabal, with subsequent dialect leveling. A quarter of a century after we had studied Tübatulabal the language became extinct. In their folk taxonomy, the last speakers of Tübatulabal distinguished between two styles, which were syntactically different ('strong talk' and 'high language'); it is possible that these

stylistic differences had been dialect differences before the dialects converged. Compare with this the folk taxonomy of modern speakers of Iowa-Oto in Oklahoma; in this language two dialects are said to be distinguished – – dialects named after the formerly distinct tribal societies (Iowa and Oto); they are, in modern practice, mixed. Our informants were a married couple, the husband an Oto and the wife an Iowa. After a week's work, it became apparent that each used dialect stigmata that were at first attributed to the other.[2] Here, again, is an American Indian language doomed to extinction through replacive bilingualism with English in the next quarter of a century (because its present bilingual speakers are all middle aged or older; though some of their children understand Iowa-Oto, they speak only in English).

Our larger world sample indicates that all on-going languages include the kinds of varieties that linguists study in dialect geography, even though many of the geographically isolated dialects are in the process of leveling today, or will be tomorrow – – as soon as their speakers begin to talk not only to each other, but to other speakers from other areas. Those languages with few speakers that appear to be without dialect variety in one generation tend to disappear in the next generation.

In the linguist's taxonomy, dialects of a language are represented as the terminal nodes of a tree diagram:

This diagram is intended to show where linguists come to a halt in studying variety, but certainly not to show the end of all variety in language.

Some linguists have expressed interest in investigating below the terminal node shown on our diagram. But no one has yet devised an adequate notation for transcribing the idiosyncratic features of voice that enable us to identify our friends from hearing them without seeing them. We do not say, when Jack phones us, that we recognize a speaker with a British accent and go on from there to sort one out of several people that we know have come from England. It is not as linguists that we recognize the voices of our friends without hesitation; nonlinguists do just as well. And linguists do not –– or do not yet —adequately transcribe the phonetic features that distinguish different individuals

speaking exactly the same dialect, though such differences have been electronically recorded as *voiceprints*, analogous to *finger prints*; this remains an ideal, not validated by reliable research.

The linguistic usage of 'idiolect' is not synononymous with 'voiceprint'. In one possible usage – – in an unpublished thesis describing one of the Numic languages in California – – Lamb wished to restrict his description of the language to one variety, even though he had recorded several dialects of that language; he said he would include in his report only the idiolect of his main informant, and since she spoke more than one idiolect, he would cite her only when she had spoken in her characteristic idiolect.[3] In this usage, idiolect is a kind of person-identified dialect. rather than a person-distinguishing voiceprint. In another possible usage – – proposed a year before Ferguson published his paper on diglossia – – Hockett identified the linguistic aspect of diglossia in Switzerland as one in which every speaker speaks two idiolects.[4] Whether used to identify the linguistic side of diglossia, or to identify the variety of a language favored by one's informant, 'idiolect' is synonymous with 'dialect'. The total referent range of 'dialect' is quite extensive; an idiolect is a dialect in its narrowest sense.

Interest in dialectology is as old as modern linguistics. Since the 18th century, dialect dictionaries and dialect grammars have attested the varieties of usage that exist within one language. The early interest was sociolinguistic in orientation, insofar as it sought to collect what were then regarded as left-overs or distortions in the speech of common people (including what Ross today calls 'non-U') — in short, whatever was left out from or different than that found in the standard of the same language, in the speech of the educated (including Ross' U-usages).[5]

Subsequently, dialect dictionaries and dialect grammars were compiled to show variety in areas, rather than variety in social classes or levels of education; it then became clear that the standard of a given language, as English, arose from a local dialect, assimilating forms and structure from other local dialects as it developed and gathered authority from being the dialect that was printed, that was the source of unabridged dictionaries (in the English speaking countries), or that was sanctioned by national academies (in the continental countries). The process of developing standards in European languages has been inferred from past century records; but in other parts of the world the development of standards

may be studied as an on-going process when writing is introduced in preliterate societies. The emerging nations sometimes take a local dialect selected by linguists as a point of departure for a standard. Linguistic missionaries devised a writing system for Bauan Fiji and then printed Fiji in this one rather than in other possible Fiji dialects; in consequence, Bauan has become a standard for the Fiji language.

The local dialect, salient in the dialect dictionary and in the dialect grammar, becomes opaque in the dialect atlas of long settled areas. The set of maps in the dialect atlas do not deal with a series of local dialects, one for each map; instead thay deal with the geographic distribution of isoglosses -- some aspect of the sounds, syntactics, or semantics of the language whose variety is being mapped. The dialect atlas gives the most objective report yet devised for a given language's variety; the criteria employed are strictly linguistic. The boundaries obtained from dialect atlas work can then be compared with sociopolitical boundaries of a local dialect or a large provincial area -- e. g. the Deep South in the United States or Yorkshire in England— thereby contributing to the kind of hyphenated linguistics that is known in anthropology as ethnolinguistics. Bloomfield's famous chapter on dialect geography is as much ethnolinguistic as linguistic in orientation.[6]

When maps of linguistic units (isoglosses) are superimposed on maps of sociopolitical units (towns and provincial areas and culture areas and nations), it may be shown that one of the latter units, as a village, is far from homogeneous; for example, north-south isoglosses separate the the western dialect of the village of Kaldenhausen from its eastern dialect. Typically, however, isoglosses do not run through a town; rather, they run between towns, thereby providing an objective boundary for the local dialect.

Most forms mapped by isoglosses diffuse in long settled areas, either singly or in bundles; those that do not diffuse leave their traces in relic forms, characteristically found in isolated enclave settlements and in the peripheries of a large area that is for the most part swamped by the diffusing features.

The goodness of the correlation between sociopolitical areas and areas bounded by isoglosses depends on whether one begins the computation of the correlation with a local dialect in the center of the area, or begins with local dialects that lie on the periphery of the area. The former way of computing gives a good correlation: the central dialect shows isoglosses which permit its

classification -- e. g. as Yorkshire; but local dialects near the periphery of one sociopolitical area will show isoglosses shared by neighboring sociopolitical areas. It is possible that the latter way of setting up isogloss-arrays would permit a new classification of a dialect that even extends to some closely related languages that are easy to classify as separate when a sample is taken from the center of a country; however, when the sample is taken from the adjoining peripheries of two countries, the language barrier has been said to be replaced, seemingly, by isoglosses shared by the languages in the two adjacent countries -- as at the French-Italian and French-Spanish borders and at the Spanish-Portuguese border. The way out of the dilemma, of course, is to accord greater weight to some isoglosses (and especially bundles of isoglosses) that run through or surround an entire village or provincial or national area, and less weight to single isoglosses of narrow distribution. This kind of weighting amounts to saying (after Hockett) that a thicker bundle of isoglosses 'is more significant than a thinner one'. In effect, the heavily weighted isoglosses will then represent the more diffusing features; it is precisely these heavily weighted isoglosses that lead us to expect (after Bloomfield) 'some non-linguistic cultural movement of comparable strength'.

2. The notion of a local dialect is intuitively more appealing than an isogloss or a bundle of isoglosses, however objectively the meandering or cross-cutting isoglosses are delineated in dialect geography. The local dialect seems to stay put, to be the exemplar of the speech of a given language in a single community (hence the coinage 'communilect'); the sum of all local dialects in a given language is often said to represent the horizontal dimension of variety in that language — horizontal dialects being those treated in dialect geography. Vertical dialects lie beyond the scope of dialect geography — in practice if not in principle.

In all non-egalitarian societies -- i. e. in class stratified societies -- there exists a vertical dimension of variety for each language. For example, the heroine of Shaw's Pygmalian (My Fair Lady) is taught a higher vertical dialect as a part of upper class etiquette. Even in egalitarian societies there are attenuated vertical-like differences in speech associated with sex differences, age differences, and ceremonial differences.

The speech of a given language in a single community always includes vertical variety (hence the inappropriateness of 'communilect'). The local dialect of dialect geography amounts to a localized sample selected from the

horizontal variety of a given language. The standard of a language may arise from a local dialect which spreads — necessarily — in the course of and as part of its development. Whether regional or national or international, the standard of a language represents the horizontal dimension – – but also the vertical dimension (wherever upper classes are more at home in the standard than lower classes). Beside such coincidence, there may also exist non-coincidence between the vertical and horizontal dimensions.

The sum of all horizontal dialects may be almost sufficient as a measure of the diversity of a preliterate language lacking a standard. In addition, the sum of all vertical dialects must be included if one is to obtain a measure of the diversity of language with written standards. The apparent lack of objective criteria for measuring diversity within a language — hence the occasional failure to find a clear–cut boundary between language and dialect — is partly due to the possible non-coincidence of the horizontal and vertical dimensions of variety in a given language, and partly to non-coincidence in academic perspectives about the object to be measured – – the language alone, or also the social structures of the societies in which the language is spoken, Nevertheless, the notion of a separate language is just as appealing as the notion of a local dialect.

It seems natural that one language (with all the variety that it contains), as English, should be separated from another language, as Spanish along our Mexican border, by a language barrier. Crossing the border physically is easy for a monolingual — a word or a document to an official who is bilingual is all that is required. Crossing the language barrier at the same place is diffucult for a monolingual; it takes a year of hard work for a monolingual adult to become bilingual, or years of desultory work. (The small number of rapid language learning adults is offset by a larger number of language resistant adults who may learn to read a second language, but fail to talk it or to understand it when spoken, even after years of effort.)

Both (a) degrees of intelligibility and (b) time of learning are relevant to distinguishing separate languages from divergent dialects of one and the same language.

It has been found that for the first variable – – degrees of intelligibility – – a speaker of one variety of a language can comprehend a message in a related language or dialect when three–quarters of the words in the V_2 message are interpretable by the listener who himself speaks only V_1; when fewer words

are interpretable in a third variety, the speaker of V_1 does not comprehend the message spoken in V_3. The test is intended to show objectively that V_1 and V_2 are dialects of one language, while V_1 and V_3 are separate languages. Fieldwork experiments in dialect distance testing among a half dozen American Indian languages measure more than a kind of lexical overlap. Dialect distance testing actually measures the interpretation of syntax in addition (since the words in the message are uttered in sentences rather than in isolation); it does not, however, provide any measure for the time that V_1 and V_2 or V_1 and V_3 have been in contact, nor for the willingness of a V_1 speaker to understand a V_2 speaker, coupled possibly with an unwillingness to understand V_3.[7]

Our second variable – – time of learning – – is also relevant to the dialect-language distinction. The experimental literature on second language learning is enormous; countless teachers in all parts of the world are concerned in one way or another with this variable. The very phrase — second language learning –– shows that it is taken for granted that the speaker of V_1 sets out to learn V_2 as another language – – as though the two varieties were necessarily separate languages rather than, possibly, dialects of the same language.

If V_1 and V_2 are, in fact, dialects of the same language, it can be predicted that there will be no failures: all the V_1 speakers will learn V_2, even though V_2 is quite unintelligible on first contact. The same cannot be said if the varieties are separate languages; when V_2 is really a second language (as they say) some failures in learning can be predicted, despite all the teaching methods that have been devised so far.

For example, what is known in Standard German as *Swiss dialect* (Schweizer Dialekt) is so divergent as to be unintelligible to visitors from Germany; even the self-designation for the dialect is different (Schwyzertüütsch). Children entering school speak nothing but *Swiss dialect*, yet by the time they learn to read and write they will have, without exception, learned Standard German – – fortunately for them,since all their school books are written in Standard German. During the Nazi period many German monolinguals fled to Switzerland where they could understand the regional standard of Swiss German, but not *Swiss dialect*; schools were set up for such adults and there, again, everyone knowing Standard German learned to comprehend the Swiss dialect when spoken.

Not only can every adult learn to comprehend a divergent dialect which is at first contact unintelligible to him, but in some cases he can estimate how long

it will take him to gain such comprehension. Something like a time of learning schedule is known in West Africa where the language divergence of an adjacent village is called a 'two-day dialect', as Hockett says, while the divergence of a more remote village is quantified as a 'one-week dialect'. In native America a partially parallel distinction is left unquantified: the near-by dialects are 'easy to understand', while the more distant ones are 'hard to understand'.

Both for children and for adults, universal success in learning to comprehend another dialect of the same language, however divergent, is dependent on willingness to learn.If a school were set up in Tehran to teach the Luri dialect to speakers of Standard Persian who were forced to go to the school, some unwilling adults might fail; but conversely, if a school were set up in Luristan to teach Standard Persian to adult speakers of the Luri dialect, all would be expected to want to succeed and, in a short period of time, actually succeed. (Wanting to succed in learning a separate language leads to less uniform success.) The Lurs say that the Standard of the Persian capital is a variety of their own language; the speakers of Standard Persian in Tehran say that Luri is a separate language.[8]

Factors of this sort can be accounted for in a hierarchy of ascribed values; the standard of a language has uniformly assured prestige. But conversely, local dialects and regional dialects are not uniformly doomed to be denigrated; it is more honorable to switch from the standard to a local dialct in England or Germany than to do so in France.

It is possible — indeed necessary in sociolinguistics, according to Labov — to begin with the social hierarchy (as well as with the social setting) when investigating on-going change among urban varieties of a given language, as English spoken in New York City.[9] None of Labov's working class and middle class informants pronounced postvocalic –r in a social setting which called for casual speech, while the majority of his upper middle class informants did so most of the time. In social settings which called for different kinds of non-casual speech, all three social classes included some instances of postvocalic –r usage — more conspicuously when reading words or reading sentences than when speaking very carefully but without the guidance of spelling orthography. The upper middle class informants pronounced –r after vowels every time it appeared in writing in the word lists that they were asked to read; but the middle class informants missed out about half the time, and the working class informants succeeded in pronouncing –r after vowels a little more than a quarter

of the time. What might formerly have been ascribed to the local dialect is now in process of change — not randomly so, but following class stratified lines. The impulse toward upward mobility in the stratified society of urban New York finds expression by a kind of unconscious 'etiquette', in the etymological sense of that word (the ticket to get in) – – namely, by the frequency with which one pronounces –r after vowels.

Vertical variety of a language may be investigated by first determining extra-linguistic factors in the stratification of the social structure (working class, middle class and upper middle class). Sociolinguistic questions of interpretation then follow. Does language performance (as pronouncing or not pronouncing –r after vowels) reflect social structure, or does social structure determine language performance? Does perception — or wishful expectation — of one's place in the social stratification influence one's language performance, or does one's language performance serve as a kind of definition of one's actual level or would-be level in a particular stratified society? What are the consequences if language is taken to be identical with social structure or culture?

One consequence seems certain: the distinction between dialect and language — regarded as an important distinction in linguistics, though immediately visible only when sharp language barriers exist — will sink from sight in sociolinguistics.[19]

3. The linguist's taxonomy distinguishes between dialect and separate language — most objectively by isoglosses (*1*, above), but also by consideration of variables which can, in favorable cases, be quantified (*2*, above). We turn next to situations in which speakers say that the variety of a language they speak, as Hindi in India, is one language and that another variety, as Urdu in Pakistan, is another language, while linguists say that Hindi-Urdu is a single language which includes numerous dialects, some more Sanskritized than others.

Dialect distance as well as learning time between two varieties of one language may be affected by a folk taxonomy. We call the taxonomic unit of the folk an ethnolanguage to distinguish it from the kind of unit (separate language) that is recongnized in the linguist's taxonomy. In the folk taxonomy of Tehran, as mentioned above, Persian is one ethnolanguage and Luri another. In Luristan, however, the local dialect is regarded as being a variety of the same language as that spoken in the capital; the folk taxonomy and the linguist's taxonomy coincide

in Luristan. Dialect distance testing would be expected to confirm the folk taxonomy of Luristan.

The two kinds of taxonomy do not coincide in Tehran where dialect dialect distance testing can neither confirm nor disprove the folk taxonomy. The folk classification of Luri as a separate ethnolanguage makes testing for partial intelligibility quite pointless; a Tehran speaker of Standard Persian would be unwilling to listen sympathetically to Luri speech, unwilling to try to comprehend. A speaker of Standard French is similarly intolerant when within earshot of a modern Provençal speaker (see fn. 7).

Folk taxonomies are also found in which one ethnolanguage is distinguished from another ethnolanguage for other reasons than that speakers of the first are so intolerant of the speech of the second that they will not bother to listen to hear whether they could understand messages in the second. The reason is certainly not this when the speaker of the first ethnolanguage cheerfully admits that he can understand all or most messages in the second, but still insists that the two are separate languages. The reasons for the separate (ethno) language status in such folk taxonomies are now under serious consideration in sociolinguistics. One ethnolanguage (Serbian) may be written in one kind of alphabet (Cyrillic), and the other (Croation) in another kind of alphabet (Latin). These two ethnolanguages look different when written, but Serbo-Croation is one language because dialects of the two ethnolanguages are mutually intelligible when spoken. In Czechoslovakia there are two standards, one quite intelligible to speakers of the other; but these two (Czech and Slovak) are nevertheless regarded as separate languages in the folk taxonomy. There is something very persuasive about this kind of folk taxonomy — so persuasive, indeed, that it has led both linguists and sociolinguists into a contradictory position. Some linguists say that dialects which are mutually intelligible belong to the same language — except when they are classified as separate languages in a folk taxonomy; some sociolinguists argue that the distinction between two mutually intelligible ethnolanguages in the folk taxonomy should — for democratic reasons — be treated as though the distinction were supported by linguistic criteria.[11]

It is when two varieties of a language are quite different – – as in widely separated dialects which are classified as belonging to the same ethnolanguage – – that we have a chance to study the kind of adjustments that the speaker of one dialect (D_1) makes when he encounters a speaker of another dialect (D_2).

After three days or after a week's contact, the D_1 listener may comprehend messages in D_2; but does he then still speak in an unchanged D_1 fashion? It is one thing to comprehend a message in another dialect, and another thing to switch dialects. In Hawaii some speakers of creolized English who have learned to speak Standard English are reluctant to switch back again to creolized English, lest they confuse the two — i. e. leave traces of the creole when attempting to speak Standard English. In Haiti, there appears to be less reluctance to switch between Creole French and Standard French. Haitian Creole enjoys greater prestige than does creolized English in Hawaii; but its structure diverges imore from Standard French than the Hawaiian creolized English diverges from Standard English.

Does the adjustment for different varieties of the same spoken language differ in principle from the adjustments made when the practicioner of one variety of silent language encounters the practioner of another variety? The famous silent language of native America, though generally understood in a multilingual culture area, is less often practiced as a lingua franca than as a a second auxiliary language at home. In the northern Plains, this silent language (hand-talk or sign language, but not classifiable in any taxonomy as a derivative of any spoken language) varies from place to place much as dialects of a spoken language do. But in the dyadic encounter of one hand-talker with a visitor from a distant tribe, immediate and mutual adjustments are made in the hand-talk that is used as a lingua franca. After three hours or a day's contact — rather than after three days or a week's contact, as in the case of distant dialect contact in West Africa – – the hand-talkers communicate freely with one another – – without further adjustments once they have compromised their differences in a single mixed hand-talk dialect. After the visitor departs, readjustment takes place — i. e. the local hand-talk dialect is again resumed on occasions which call for its use.

Something of the same sort appears to happen when our attention is completely engaged in very long conversations with a friend who visits us from another dialect area. Whether or not our friend makes adjustments to our dialict when he speaks, he must (as a hearer) be quick to make adjustments when we speak. If our friend's dialect is classified in the folk taxonomy as being very divergent, we may confine ourselves to making hearer adjustments. and not make any speaker adjustments at all.

In the Hopi dialect taxonomy, Second Mesa is supposed to be quite different

from Third Mesa, yet everyone makes immediate hearer adjustments so that intelligibility is not impeded; but if any man from Third Mesa makes speaker adjustments to the Second Mesa dialect, he will be teased as an *éf-taqa* (literally an 'ef man' for in the Third Mesa dialect the final labial is –p, not –f). In the Osage-Ponca dialect taxonomy, the Osage dialect is supposed to be quite similar to the Ponca dialect ; 'they speak just like we do only faster'. An Osage can afford to make speaker adjustments to Ponca, without danger of being teased for speaking faster (or whatever it is that really distinguishes the two dialects).

These are instances of two different kinds of folk dialect taxonomies. In the first, speaker adjustments are discouraged; in the second, speaker adjustments are almost encouraged. But in both hearer adjustments are taken for granted, as is shown by the fact that full communication is possible between each pair of dialects. These instances also suggest that a folk dialect taxonomy may reflect social distance in addition to or instead of dialect distance.

4. Major isomorphisms and parallel divergences in the syntactic-semantic structure of a given pair of languages are presumed to reflect descent from a single parent language if – – and this is the crucial condition – – the sound systems of the presumptive daughter languages include enough identities (continuities without change from the parental sound system) and sound correspondendences (continuities with changes, as splits and mergers, from the parental sound system) to permit contrasts in the sound system of the parent languages to be reconstructed almost as completely as the contrasts that are found in the daughter languages. When some earlier languages are preserved in writing — as they are in only a few families — then the textual records of the most archaic languages, as Sanskrit, Classical Greek, Latin, Gothic, and Classical Arabic (to mention only the main languages preserved in alphabetic writing) serve as more than sources for additional detail, and give more than etymological lists (e. g. the list showing the several different long and short vowels and dipthongs in Classical Greek that merge in one modern Greek vowel, *i*). The textual records serve as testing frames for the general validity of comparative method reconstructions; independent reconstructions of a parent language (based on nothing more than comparisons of the daughter languages) are quite uniformly confirmed by textual records, whenever they are available. This now gives confidence that in those cases where textual records of some form of a parent language are not available—and this in the general situation—the unaided reconstruction of a parent language by

the comparative method yields (in principle) as reliable a description as the descriptions of the daughter languages themselves.

Hence, the validity of comparative method reconstruction of preliterate languages is no longer debated. The debate has shifted to the internal relationships among languages which are demonstrably members of one language family. The demonstration has in some cases been accomplished by selecting very few daughter languages for comparison out of many daughter languages that are available for comparison. Bloomfild illustrates this kind of key language attestation by showing (after Dempwolff) that if we compare only three languages — Tagalog in the Philippines, Batak in Sumatra, and Javanese in Java — out of the scores of Indonesian or most broadly, Austronesian languages (spoken from Easter Island and Hawaii to Southeast Asia, with an outlier in Madagascar), the three languages are still sufficient to exemplify 'eight normal types of correspondence' and their reconstructions in 'Primitive Indonesian'; the key languages selected for reconstructing Proto Algonquian were Ojibwa, Cree, Menomini, and Fox.[12]

The comparative method has a limitation (*5*, below). However, it is not a limitation of the comparative method but the selection of a few key languages to attest language families that precludes linguistic interpretation of internal relationships in each of these attested language families. In lieu of data for linguistic interpretation, culture area interpretation continued to influence the grouping of Austronesian languages; they were divided into Polynesian, Micronesian, Melanesian, and Indonesian branches until fresh information was obtained from a larger sample to give linguistic evidence for the internal relations of Austronesian languages, largely by Grace and Dyen.[13]

The revolution in the still waters of the Pacific may be appreciated if one compares or analogizes between our single Indo-European language family, and the single Austronesian (Malayo-Polynesian) language family. In the latter, all branches have been questioned, challenged, and rearranged. It is as though we were to decide, in Indo-European, that English was no longer Germanic, as previously supposed, but Italic; that Rumanian was no longer Italic, but Slavic; that Bulgarian was not at all Slavic, and hence not Indo-European, as previously supposed, but Turkic and so on.

The internal relationships of the Indo-European family — Germanic versus Slavic, and so on — are given in every elementary textbook, and the history of their discovery is given in the most widely read book on the history of linguis-

tics.[14] But the history of the Uralic family affords more insights into the transition from the *prescientific paradigm* to the *comparative method paradigm* then does the history of the establishment of the Indo-European family (cp. *5*, below).

The language family called Uralic (after the Samoyedic branch was added to the Finno-Ugric branch) is as well attested as any other family. The following chart expresses succinctly the modern view of the internal reationships in the Uralic family.

This modern view represents a synthesis of information obtained largely from fieldwork, and successive revisions in reconstructive theorizing that began exactly three centuries ago.[15] As early as 1669, Martin Fogel of Hamburg offered evidence for a Finnish-Hungarian relationship, but his study was never published. G. W. v. Leibnitz tried to secure data from hitherto unknown languages, especially by asking travelers and ambassadors for their collaboration and support. In a letter of 1708 he mentions three widespread 'languages' [language families] in Scythia; the last 'language' [group of languages] mentioned by Leibnitz is 'Fennic' (the language of the Lapps, the Finns and the Hungarians) — a 'language' that 'reaches beyond the Caspian Sea'. On the basis of such sources as Fogel and Leibnitz, it can be said that knowledge of the existence of the Finno-Ugric relationship began in the 17th century, about a century before Jones' announcement of the Indo-European relationship. However, extensive word lists and other linguistic materials became available only in the second half of the eighteenth century.

In 1770, J. Sajnovics established the relationship of Hungarian and Lapp on the basis of lexical and morphological comparisons. He had studied Lapp languages while in Norway for astronomical observations, but did not base his comparisons on what he had elicited from Lapp informants. According to the prevalent ideas of his time, he based his statements on data from written sources. His conclusions were accepted by several outstanding specialists abroad, but rejected in Hungary; the proposed relationship with primitive peoples in the north was not compatible with the Hungarian folk taxonomy.

In 1799 S. Gyarmathi published a book in Göttingen to prove the relationship of Hungarian with the languages of Finnish origin. The reference to Finnish was first of all calculated to be a safe one for Hungarians, since the Finns would have been much more acceptable as relatives than Lapps. Gyarmathi

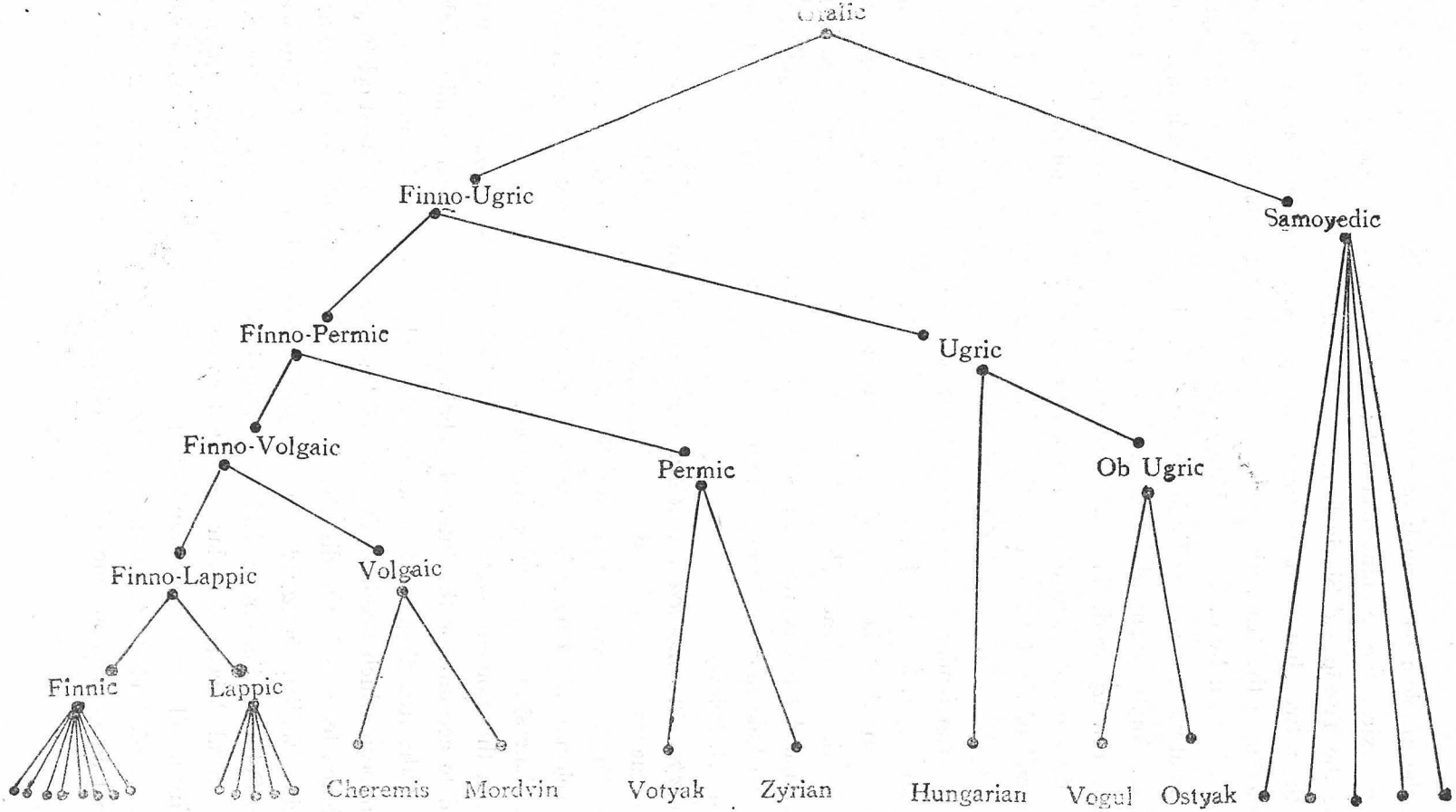


Diagram of Uralic family

(the terminal nodes specify living representatives, except for the names of the Finnic, Lappic and Samoyedic languages which are not listed)

Apex and terminal nodes in the linguist's taxonomy of genetically related languages

dealt with the problem on a much broader basis than did Sajnovics, using all the available lexical and grammatical data; and he was perceptive enough to characterize the Turkic and Slavic forms in Hungarian as borrowings.

Since these two Hungarians of the eighteenth century arrived at conclusions that ran counter to the Hungarian folk taxonomy, their efforts did not stimulate continuation of work by others in Hungary. It was almost a century later before comparative Finno-Ugric linguistics was established with continuity. The established continuity began with the efforts of J. Budenz who was stimulated by nineteenth century work in the Indo-European family; subsequently E. N. Setälä brought the neogrammarian approach to Finno-Ugric linguistics.

By the end of the 19th century the set of methods by which proof of relationship is established were as fully accepted for the Finno-Ugric family as for the Indo-European family. Since then this set of methods has been accepted for comparing languages in other language families especially in native America and in Oceania. For the sake of brevity, we use the phrase ***comparative method paradigm*** for the set of methods in the proof, and find that the crucial condition already mentioned (first sentence in *4*) is most relevant to an understanding of attempts to include more and more languages or groups of languages in one genetic unit (*5*, below).

The ***comparative method paradigm*** is not a game whose object is to squeeze as many languages as possible into one language family. That it is not that sort of game is shown by Dempwolff's and Bloomfield's highly efficient strategy (see above): to select as few languages as possible in order to exemplify 'normal types of correspondences' and their reconstructions.

However, the moment more languages are included—as of course more must be included to account for the internal relationships in a given family — the simplicity of the tree diagram becomes complicated, as by several shared features in Indo-European which overlap otherwise distinct branches. The most celebrated of these shared features is the distribution of a velar consonant (as in *centum*, cognate with English *hundred*) in the western branches of Indo-European (Germanic, Celtic, Italic, Greek, and—but anomalously so—in Tocharian, formerly spoken in central Asia); this velar consonant corresponds to a sibilant consonant (as in *satem*) in all the eastern branches of Indo-European (Albanian, Armenian, Baltic, Slavic, and Indo-Iranian) except Tocharian. This *centum-satem* distribution is not an isolated instance of what Johannes Schmidt tried to account for

by his *wave theory*.

5.0. There is always the possibility that all of the world's language families —hence all of today's languages—may be descended from a single Lingua Humana. This possibility is never doubted; but what is said either about the survival or else the reconstruction of a Lingua Humana may be debated in one 'paradigm', only to be rejected in a following 'paradigm'. For distinguishing between different 'paradigms' having to do with history in the last three centuries of linguistic history, we begin with a *pre-scientific paradigm*, followed by the *comparative method paradigm*, whose limitation has led to the on-going *crisis paradigm* which, in its turn, may lead to what is already an incipient *universal paradigm*—i.e. to a completion of the circle by returning to the Lingua Humana (see introductory footnote*).

5.1. In the *pre-scientific paradigm* the Lingua Humana was thought to survive beyond the Confusion of Tongues. In 1679 Kircher doubted whether this Primitive Language — Adam's 'natural language' — or its underlying base could be reconstructed, but did not doubt the possibility of its survival; he said there was 'strong authority for the belief that the Lingua Humana had not been destroyed, and that it survived in the language of China'. He considered the possibility of a single model for a new universal language, adopting Chinese characters for the description of languages proposed by Leibnitz and others. He imagined that the Lunarian languages of Francis Godwin and Cyrano de Bergerac had a Chinese source — that is, since the Lunarian languages could be produced both by voice and by musical instruments, they were at bottom tonal languages from China.[13]

Other trends in the *pre-scientific paradigm* of 17th and 18th century linguistics led to the development of the comparative method (cp. *4*, above and *5.3* following).

5.2. Accepting the *comparative method paradigm* sometimes goes hand in hand with imposing a limit on askable questions, notably questions about typological similarities between languages that are classisied in different language families. For example, the outstanding Indo-Europeanist of the 20th century, Antoine Meillet, discouraged interest in typology by dismissing it as non-serious, as amusing; one does not address serious questions to non-serious enterprises. Both in the 1924 edition and in the 1952 editiun of *Les langues du Monde*, by a group of linguists

under the direction of Meillet and Cohen, languages are classified in terms of their genetic relationships; typological classifications are excluded. It has been said that this brands Cohen, the surviving editor responsible for the second edition, as failing to keep abreast of developments in modern linguistics. The truth of the matter is, rather, that since typology is as old as the comparative method (indeed, older), it cannot be counted as a modern trend in linguistics.

But typology has never been wholly accepted as a paradigm equal in solidity to the comparative method paradigm. For example, in discussion with fellow Algonquianists, Bloomfield was not reluctant to point out the typological parallels between Algonquian and older Germanic; but he was reluctant to publish these resemblances specifically—i.e. in typological terms—although he did publish a note on the general resemblance of Algonquian structure to Indo-European as a whole. His reluctance to publish the specific typology kept him safely out of danger of being interpreted as in favor of work toward establishing a genetic connection between Germanic and Algonquian. Such a connection, though 'entirely misguided' as Hockett says, was nevertheless urged in four volumes called *The Viking and the Red Man*; Bloomfield did not bother to criticize this kind of work.[17]

The weakness of typology in general is its openness to gross misinterpretation. The comparative method paradigm does not suffer from this weakness.

5.3. In contrast to the *comparative method paradigm*, the strategy of the *crisis paradigm* is to squeeze as many languages as possible into one presumptive genetic unit. In order not to introduce controversial terms at this point of the argument, we will identify attempts to classify as many languages as possible in one genetic unit as the search for external genetic relations. In The Structure of Scientific Revolutions, Kuhn characterizes a *crisis paradigm* as one in which there is persistent worry over anomalies and over the limitations of a prior paradigm. The retention of the velar isogloss by Tocharian (in an area otherwise bounded by the sibilant isogloss) shows that anomalies do occur in results obtained by the comparative method; this is not to be confused with the limitation of the comparative method paradigm. When genetic relations are postulated without heeding the crucial condition of the comparative method (stated in *4*, above, first sentence), they seem to imply that it is possible to go beyond or go without the comparative method paradigm, in one sense or another. The postulated relationships are then debated, of course; such debate often leads to a kind of dilemma,

since most of the world's language families are postulated in this fashion. We are concerned here, however, with the fact that after a language family, as Austronesian, is attested by the comparative method, attention may be given not to internal relationships within the family, but to external relationships, as the postulation that the Austronesian family is related to the Thai family in SoutheastAsia.

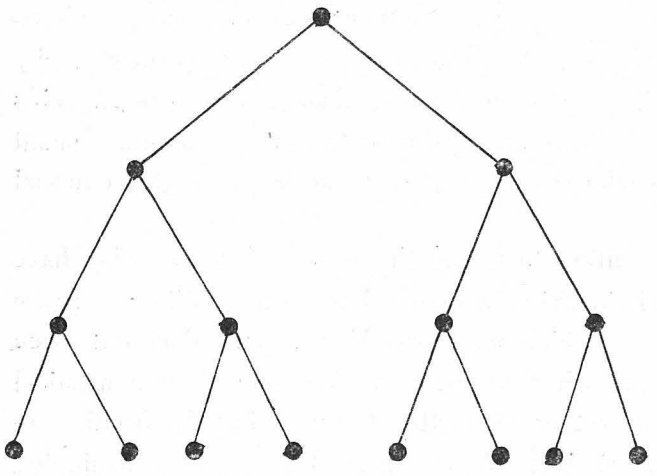
The only way to escape from this kind of postulation in the ongoing *crisis paradigm* is to refrain from searching for and from discussing external genetic relationship; to keep, instead, within the bounds of those language families which have been established by workers meticuously heeding the crucial condition of the comparative method. Linguists who have been known to practice such restraint sometimes become passively fascinated with the occasional first-class linguist who is less restrained. Edward Sapir was invited to Yale after he had brought the crisis paradigm to American linguistics; the crisis paradigm did not at all worry Sapir's Yale colleagues, since they refrained from searching for external genetic relationships and refrained also from lecturing on languages of the world. But since Sapir's day—and especially since the post-war expansion of American horizons — the scope of linguistic interest has extended beyond one language family in Oceania, beyond a few in the Old World and a few in native America. It is today more difficult to escape from the crisis paradigm than it was in Sapir's day; for many of us, it is necessary to orient our students about more than the few better known language families.

How, then, to phrase our report about the lesser known language families and their postulated relationships to one another? Without denying the postulated relationships, it is always possible to neglect the claims that have been made by merely listing such languages in geographical rather than genetic groups. This is a dusty answer, unsatisfying to those who question whether numerous small language families are related — who search or would like to search for external genetic relationships.

Suppose, now, that we are willing to report the work of those who have searched for external genetic relationships between language families; suppose further that the part of the search which we report is the part that has been carried on by sophisticated linguists who are not expecting to turn up a mystical relationship for languages but are rather committed to extending the frontiers of knowledge in terms of the usual kind of relationship. But instead of beginning with a pair of languages that are living representatives of a language family, they begin, or would begin if data were available, with a pair of parent languages — languages already reconstructed by others. The two parent language were

also a single language at some earlier time. The difference between descent to languages still spoken, and descent to a pair of parent languages from a still earlier single language appears to involve nothing more than relative chronology. So the worry aroused by the *crisis paradigm* seems unjustified unless it can be shown that the methodology employed by that paradigm differs in some way from the set of methods by which proof of relationship is established in the comparative method paradigm. If that can be shown, then the worry aroused by the *crisis paradigm* can be restated as a half-acknowledged recognition of the limitation of the comparative method paradigm. It is possible to cite examples which show the unjustified worry type of difficulty, but what we are after is examples that show the limitation of the comparative method.

Language family will be used as an abbreviation for the genetic relationships of languages in a taxonomy whose highest node (alpha, in the diagram) has been established by the comparative method; *phylum* will be used as an abbreviation for the higher genetic relationships of languages in a taxonomy in which there is a node above the alpha node—a node called the 'postulated parent language' because the evidence for proposing it is not sufficient to meet the crucial condition of the comparative method (*4*, above). In short, the *phylum taxonomy* goes beyond the language family taxonomy, and shifts methodology in doing so, but retains the usual interpretation of historical linguistics.



(α) parent languages established by comparative method

(β) parent language of some but not all daughter languages in the family

(γ) living representives (separate languages)

(δ) dialects

Diagram of language family taxonomy

(caveat: there may be more than one beta level, as in the diagram for the Uralic language family in which there are five beta levels)

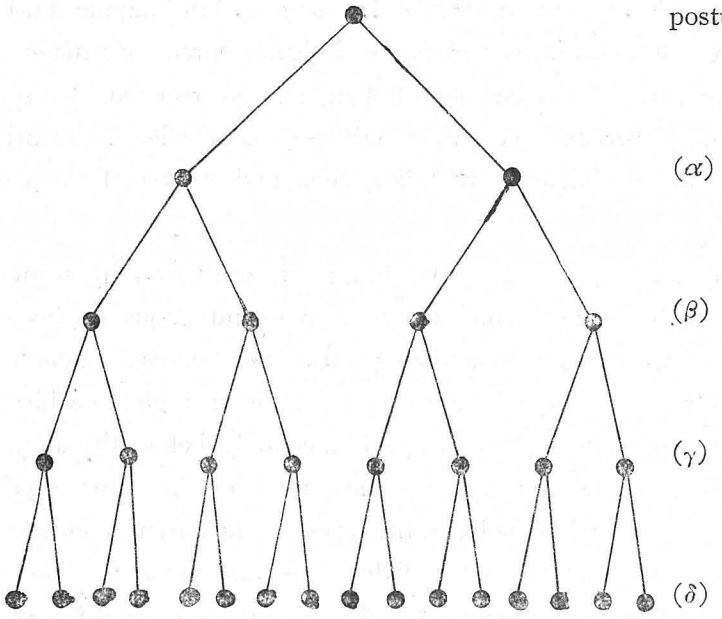


Diagram of language phylum taxonomy

(caveat: there may be more than one postulated parent language level, as would appear in a diagram of a macro-phylum)

The explicatory advantages to be gained from the *language family* taxonomy alone are so great that proponents of the family-phylum distinction sometimes attribute to the distinction a few justifying advantages which, are, in fact, spurious advantages—spurious in the sense that the advantages claimed to be gained by the distinction can be realized by the family taxonomy unencumbered by phylum considerations. Before turning to the one genuine advantage of the family-phylum distinction, we discuss a few of the spurious claims that have been made for it.

1. Some feel that technical terms, such as phylum, are needed to keep one's bearing for higher levels in a multilevel taxonomy. It may be unusual to find a language family taxonomy with seven levels—from apex to terminal nodes — as that shown for the Uralic family, but, in principle, an infinite number of levels can be accommodated in a language family taxonomy. No technical terms at all are needed for the report, if the report is given in the form of a diagram, as in *4*, above; if we were discussing the Uralic family without paper and without chalkboard, we could get along with ordinary English words, saying that

the living representatives of the known Lapp languages are grouped together into a Lappic group; these and the Finnic group of languages are then reconstructed (in principle) into a Finno-Lappic parent language; Finno-Lappic and Volgaic are then reconstructed into an earlier Finno-Volgaic parent language; and so on until the earliest reconstructed parent language is reached (proto Uralic). Returning to living representatives, we would say that Cheremis and Mordvin are reconstructed into a Volgaic parent language; and so on, back to proto Uralic again.

2. Rather than merely keeping one's bearings in a multilevel hierarchy, some urge that the family-phylum distinction makes possible an unambiguous explication of the degree of divergence in a given taxonomy. But the language family taxonomy, alone, can do this just as unambiguously. See, for example, the hierarchical level of Samoyedic in the Uralic family — immediately below the apex node, as is Finno-Ugric, so that these two are at the same level. This says clearly enough that the first and earliest split of the proto Uralic parent language was into Finno-Ugric dialects and Samoyedic dialects — that these are now reconstructed as two different parent languages of the living representatives of Samoyedic languages and Finno-Ugric languages, respectively. Divergence of this sort can be stated without benefit of the family-phylum distinction.

3. Related to 2 is the claim that for the earliest language family split of the sort mentioned, an absolute chronology of five or six millennia can be ascribed, so that in round numbers all splits below five millennia can be included in the language family taxonomy, while those which took place long before that time would belong to phylum linguistics. Would this not permit us to speak of Indo-European as a five millennium language family, while Collinder's postulation of a genetic relationship between Indo-European and Uralic could be categorized as a ten millennium phylum? We hesitate to name a somewhat bitter controversy that not very long ago engaged the attention of every linguist, only noting in passing that for some time back no linguist has taken glottochronology seriously. If the family-phylum distinction had to depend on the support of a theory of a single constant factor in language change, like carbon 14 in physics, then the distinction would of course have to be abandoned.

All of the claims noted (1, 2, and 3, above) are spurious; the one genuine advantage in the family-phylum distinction is that it pinpoints the area in a hierarchy of linguistic descent at which the limitation of the comparative method

is reached.

After the transition in the development of the comparative method paradigm from an emphasis on comparative grammar to an emphasis on sound change — by placing morphology and morphophonemics back stage (in a theatrical sense) coupled with placing sound change and its associated etymologies front stage — the paradigm reached its florescence, and the neogrammarian regularity of sound change was its crowning flower.[18] Among sound changes, it is the splits and mergers from the parental language to the living representatives which we compare that are especially fascinating. Definitely less exciting are the continuities in sound—continuities in which some contrasts in the sound system of the parent language (as *s and *$\check{s}$ in proto Algonquian) are literally continued in the daughter languages (as the preservation of *s and *$\check{s}$ in Fox); or, if changed phonetically, the change in the daughter language does not affect the contrast—so the particular contrast is after all preserved, as when the parental *s appears phonetically in Shawnee as θ: the contrast is still s (now written θ in Shawnee) and $\check{s}$.

What is most interesting in the comparative method paradigm — splits and mergers — can never, in principle, be demonstrated in phylum linguistics which, whenever possible, compares the apex parent language of one language family (as proto Semitic) with that of another language family. When the relevant parent languages (as proto Berber, proto Cushitic and proto Chadic in the Afroasiatic phylum) have not yet been reconstructed, cognates are sought among the living representatives in the language families. Suppose, however, that the apex parent language of two language families (LF_1 and LF_2) are available for comparison. In this situation, an x number of cognates will be relevant to the sound system in the apex reconstruction of one language family (LF_1), and a y number of cognates will be relevant to that of the second language family (LF_2) which is supposed to be externally related to the first. Out of all the x number and y number of cognates, only some will be cognate to each other—a z number; it is only the z number of cognates that are relevant to the sound system in the reconstructed parent language of the phylum. Both in principle and in the known experience of those who work in phylum linguistics, the z number of cognates is never sufficiently large to permit attestation of all the major contrasts in the phylum parent language.[19] In consequence the fragmentary sound system of the phylum parent language is more apt to be invoked as evidence for con-

tinuity in the descendant language than as evidence for splits and mergers, since it is next to impossible to demonstrate sound change of the split and merger kind from any reconstruction that lacks some of the major contrasts in its sound system.

Cognates can be likened to colored stones of various shapes and sizes and colors that have fave fallen out of one old mosaic, and recast—as many times as needed—to fit into several modern mosaics. If time has faded some of the stones, their hues may be restored semantically; the pattern of the old mosaic may also be restored, so long as too many of the stones have not been lost. But restoration of the old mosaic is more difficult after the modern mosaics are tampered with—after the stones that once fitted into a particular modern mosaic are removed and recast and exchanged and reset into all the other modern mosaics in a language area dominated by borrowings.[20]

The best example of such a language area is that in which the Ural-Altaic languages are found. The gaps between the following appear to be neither great nor clear-cut: (1) Uralic family; (2) Tungus family, including the Manchu language; (3) Mongol family, including Khalkha Mongol, the Standard language of the Mongolian People's Republic, but also spoken as a lingua franca through most of the Mongol-speaking territory; (4) Turkic family, including modern Turkish and a half dozen other languages that are unbelievably similar to one another today, possibly in consequence of earlier more differentiated languages remaining in contact during the last several centuries. These four language families have remained typologically constant for a long time, as can be seen from written records. It was in fact precisely this typological uniformity that suggested an 1836 postulation (by W. Schott) of genetic affiliation of all four families into what we would call today a language phylum; two years later F. J. Wiedemann brought together more than a dozen typological features shared by all the families then included in the Ural-Altaic phylum (since then Korean has been added, also for typological reasons): (1) vowel harmony, (2) agglutination, (3) personal possessive suffixes, (4) extensive verbal derivation, (5) postpositions without prepositions, (6) attribute-head order, (7) singular noun compatible with numeral quantifier while plural noun is incompatible, (8) comparative constructed with ablative case, (9) conjugated negation, (10) use of interrogative particles, (11) verbal constructions without conjunctions, and (12) case without gender. Much valuable field work has been done (e.g. by M. A. Castren) since

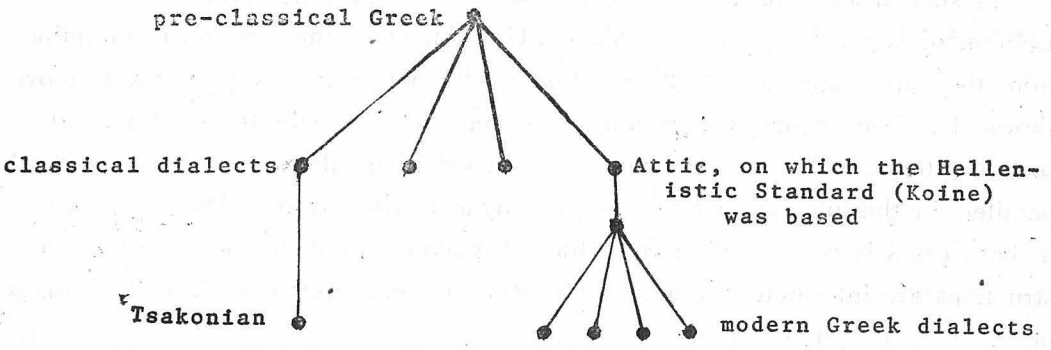
the Ural-Altaic phylum was postulated, well over a century ago, but typological considerations such as those cited here still dominate discussions of the external relationships of language families in the Ural-Altaic phylum (cp. fn. 15).

In such discussions the typological parallels of the four (now five, with the addition of Korean) families included in Ural-Altaic are interpreted as continuities from the parent language of the phylum. The Afroasiatic example given above shows that from cognates of *z* number—a relatively low number—a fragmentary sound system can be reconstructed; and in addition, there are also structural parallels in this phylum, though not as many as in Ural-Altaic. The interpretation in both cases is quite similar; from the Afroasiatic parent language, also, reconstructions are interpreted in terms of continuities rather than of sound changes of the split and merger kind.

What is least interesting in the *comparative method paradigm*—continuity of contrasts within a sound system (as **s* versus **š* remaining as θ versus š in Shawnee), or continuity of pattern in other processes—turns out to be most favored in interpretations which go beyond the reconstructed language family. Partly disguised evidence of such continuity is sometimes discussed as continuum or drift. We are not surprised when Sapir distinguishes, as linguists usually do, between sound changes (mergers—'phonetic leveling'—and splits) and continuities (a 'sort of shifting about without loss of pattern'); but we would be surprised—if we did not understand the continuum-drift development of Sapir's later work beyond the limit of the comparative method paradigm—to see that as early as 1921 he says, as linguists usually do not say, that continuities are more important for historical linguistics than sound change of the split and merger kind.[21] The following paragraph gives our interpretation of how important the continuity-continuum-drift notion is in the development of the phylum perspective by its most articulate proponent.

For Sapir, drift is found at all levels in the taxonomy of language relationships, from the terminal nodes (dialects) to the apex (postulated parent language of a phylum). Drift is the tacitly directing norm that holds the individual's variations within the confines of the dialect; if the individual insists on going beyond this limit, his eccentric variations are cancelled out by the concensus of dialect usage (p. 158). Despite this, dialects do differentiate, and thereby make it possible to think of language as floating 'down time in a current of it own making' (p. 160). Leveling only temporarily contracts differentiated dialects,

which are bound to differentiate again (p. 162).



Drift is unidirectional variation not always visible in the multidirectional flux of present time; hence its actual direction can be best appreciated in historical spans of time (p. 166). This unidirectional drift is irresistible; though some of us may be reluctant to accept it, it continues inexorably (p. 166f.), while we hesitate (p. 173). It is difficult to quantify the velocity of the drifting currents in a language because they run at two levels, simultaneously, but the current on the surface runs faster than the deep current (p. 183-4). After the dispersion of dialects, the faster drifting surface features differentiate the former dialects into separate languages, while the depth features bear witness to the continuity of the more fundamental drift (p. 184): e.g. the repeated occurrences of the same sound shift in daughter languages that have long since been so distantly disconnected that the parallel sound shifts cannot be interpreted as a consequence of contact interinfluence (p. 184). But the on-set of this kind of drift may be sequential (p. 184-5): e.g. the on-set of plural marking umlaut appeared in the history of English (8th century) before it appeared in the history of German (not before the 11th century); hence, examples illustrating umlaut cannot be reconstructed with umlaut for proto Germanic. Continuity of contrast over millennia, despite change in phonetic detail—i.e. retention of a particular pattern by compensatory phonetic adjustments — that appears in one language (and its descendants) may be incompatible with the drift in another (p. 196). Why?

For reasons which cannot be understood until we are ready to study 'the intuitional bases of speech', and until we relinquish the notion that sounds, syntactics, and semantics are each self-sufficient or 'mutually irrelevant' (p. 196-7). Direction of drift in one line of descent may be slightly deflected when the descendant of a proto language is swamped by borrowings (p. 206). Or the drift may have had its on-set before contact with the donor language—as was the drift of English toward a more analytic structure before its contact with French (p. 206 f.). Language differs from culture, but not because the latter diffuses with the greatest of ease while the former does not; rather because diffusion (borrowing) changes nothing more than the surface features of language, not its depth features—the donor language contributes nothing more than 'superficial additions on the morphological kernel' of the borrowing language (219-20). It is in respect to its depth features that language is massively resistant to change through impact with a donor language; it is easier to kill off a language than to change the direction of its autonomous drift (p. 220).

5.4. The comparative method paradigm represents as unshakable an advance in nineteenth century science as does Darwin's theory of evolution. In that century the beginnings of phylum linguistics seem more like Lamarckian evolution —persuasive but premature. In the present century, phylum linguistics goes as far as it can with the comparative method paradigm, and then attempts to go beyond the limitation of that paradigm. Yet all postulations above the reconstructions attested by the comparative method are open to doubt, to contradictions, and to unending revisions.

Yet the interest in phylum linguistics continues, in this century, and not merely because it suggests a still earlier parent language in the prehistory of possibly related language families. In addition, it has something to say about the nature of language; the notion of drift lies behind modern interest in phylum linguistics. In a future history of science perspective, it may well appear that phylum linguistics had engaged the attention of linguists as a transitional enterprise between the *comparative method paradigm* and the now incipient *universal paradigm* which is primarily concerned with the nature of Lingua Humana.

FOOTNOTES TO 'APEX AND TERMINAL NODES . . . '

* This is a reconsideration and revision of a paper delivered at a fortnight's conference on the History of Linguistics, held at Burg Wartenstein in August, 1964, under the auspices of the Wenner-Gren Foundation. We wish to thank the Foundation's Director, Lita Osmundsen, and

our colleagues at the conference for the criticism, stimulation and pleasure that remain in our memory together with a host of unresolved questions. These questions became salient again in two papers on Uralic dialects and Altaic dialects delivered at the Ethnolinguistic Seminar in October of 1967 by Alo Raun and John Krueger, respectively; we are indebted to these Indiana University colleagues for the idea of returning to the question, and to Max Zorn for debating the question with us in the light of divergent viewpoints expressed in an exchange of letters with four linguists, Warren Cowgill, Kenneth Hale, Charles Hockett, and Dell Hymes. Of the unresolved questions that remained after the 1964 Conference, there is one that is a 'crisis' question (in Kuhn's terminology); this is the one that a few of us, as mentioned, have recently discussed in correspondence. In the last section of this paper (5), we state the question in terms of Kuhn's successive 'paradigms' (but leave the question unresolved): Thomas S. Kuhn, The Structure of Scientific Revolutions *in* International Encyclopedia of Unified Science, II. 2 (University of Chicago Press, 1962).

1. C. F. and F. M. Voegelin, Languages of the World, Sino-Tibetan Fascicles One through Five, AL 6, no. 3 (1964), AL 7, no's 3, 4, 5, 6, (1965); Indo-Pacific Fascicles One through Eight, AL 6, no's. 4, 7, 9 (1964), AL 7 no's. 2, 9 (1965), AL 8, no's 2, 3, 4 (1966); Native America Fascicles One and Two, AL 6, no. 6 (1964), AL 7,no. 7 (1965); Indo-European Fascicle One, AL 7, no. 8 (1965); African Fascicle One, AL 6, no. 5(1964); Boreo-Oriental Fascicle One, AL 7, no. 1 (1965); Ibero-Caucasian and Pidgin-Creole Fascicle One, AL 6, no. 8 (1964); Index to Languages of the World, A-L, and M-Z, AL 8, no's. 6 and 7 (1966).

2. C. F. Voegelin, A Problem in Morpheme Alternates and Their Distribution, Lg. 23.245-54 (1947); Tübatulabal Grammar and Tübatulabal Texts, UCPAAE 34.55-246 (1935); Working Dictionary of Tubatulabal, IJAL 24.221-8 (1958); and Morris Swadesh, A Problem in Phonological Alternation, Lg. 15.1-10 (1939).

3. Sydney M. Lamb, Grammar of Mono, unpublished Ph. D. dissertation, University of California, Berkeley.

4. Charles A. Ferguson, Diglossia, Word 15.325-40 (1959), reprinted in Dell Hymes, Language in Culture and Society (Harper and Row, 1964); Charles F. Hockett, A Coures in Modern Linguistics (Macmillan, 1958), p. 321; subsequent phrases in this paper that are ascribed to Hockett are also taken from the Course.

5. Alan S. C. Ross is a twentieth century linguist who has reawakened an eighteenth century Zeitgeist with his Linguistic Class Indicators in Present-Day English, Neuphilologische Mitteilungen 55.20-56 (1954).

6. Leonard Bloomfield, Language (Henry Holt, 1933), Chapter 19, Dialect Geography, pp. 321-45; a subsequent phrase in this paper ascribed to Bloomfield is also taken from this chapter of Language.

7. For bibliography of dialect distance testing, see Hockett, (*op. cit.*, end of chapter 38). For a West African situation in which it would be pointless to attempt dialect distance testing, see Hans Wolff, Intelligibility and Inter-Ethnic Attitudes, AL 1:3. 34-41 (1959), reprinted in Dell Hymes, *op. cit.* Upon reading Hans Wolff, we were reminded of our Parisian friend who spent his summers in the South of France as a child. When asked whether he could understand the Provencal spoken there, he was surprised at the question. A Parisian does not pay attention to what a patois speaker says; it is the patois speaker who has to worry about how much Standard

French he can understand.

8. A. Banayan (personal communication, 1966, based on Banayan's fieldwork in Luristan).

9. William Labov, The Social Stratification of English in New York City(CAL, 1966).

10. For example, the dialect-language distinction has come to have no significance in Fishman's typology which adopts only the social structure side of Ferguson's diglossia situation, and ignores the superimposed dialect or linguistic distinction which Ferguson proposed; that is to say, bilingualism is merged with the social structure aspects of a kind of diglossia that is no longer identifiable as it once was by Hockett — as a social situation in which virtually every adult commands two idiolects in one language. See Joshua A. Fishman, Bilingualism With and Without Diglossia; Diglossia With and Without Bilingualism, Journal of Social Issues 22:2.29-38(1967).

11. E. g., V. Krupa, Review of Languages of the World: Indo-Pacific Fascicle One, Linguistics 25.132 (1967); Heinz Kloss, 'Abstand Languages' and 'Ausbau Languages', AL 7:7.29-41(1967).

12. Leonard Bloomfield, Language, p. 310; the four Algonquian languages that Bloomfield used as key languages for reconstruction (p. 359 f.) are also the ones he uses for his discussion of features of grammar, as the obviative (p. 241), and personal-definite pronouns (p. 256 f.), and tolerance of words of all lengths (p.395 f.).

13. The internal relationships are summarized by Isidore Dyen, A Lexicostatistical Classification of the Austronesian Languages, Indiana University Publications in Anthropology and Linguistics 19 (1965); Dyen utilized a large lexicostatistic sample (gathered by many field workers, notably George Grace) to formulate an interesting network of internal relationships which served as a guide for grouping Austronesian languages in our Indo-Pacific Fascicles One through Four. For the influence of culture area interpretation on the interpretation of the internal relationships in a half dozen of the better known language families of native America — and, since the work of Lounsbury, Chafe, and Postal, we can now add Iroquoian — see C. F. Voegelin, Influence of Area on American Indian Linguistics, Word 1.54-8 (1945), reprinted in Dell Hgmes, *op. cit.*

14. Holger Pedersen, The Discovery of Language (Linguistic Science in the Nineteenth Century), translated by John Webster Spargo, Harvard University Press, 1931; Indiana University Press, 1962.

15 We are indebted to our colleague Alo Raun for supplying us with the information which we cite here on the history of Uralic and of Uralic studies (personal communication, and our collaboration with Alo Raun and David Francis, Boreo-Oriental Fascicle One, AL 7, no. 1, 1965),

16. For seventeenth and eighteenth century references to the Lingua Humana (in the context of languages of other planets, ideal languages with 'rational' grammar, and imaginary languages of imaginary voyages), see Edward D. Seeber's review of Paul Cornelius, Languages in Seventeenth- and Early Eighteenth-Century Imaginary Voyages, Modern Philology 65.73-5 (1967).

17. Leonard Bloomfield (personal communication); Charles F. Hockett, Implications of Bloomfield's Algonquian Studies, Lg. 24. 117-31 (1948), reprinted in Dell Hymes, *op. cit.*

18. For Indo-European, this transition is usually associated with Bopp for the earlier emphasis and Schleicher for the latter; a modern revision of the latter is given by Henry M. Hoenigswald, Language Change and Linguistic Reconstruction (University of Chicago Press, 1960), and the history of the transition is given in that author's working paper for the 1964 Burg Wartenstein conference.

19. Those who work in Afroasiatic expect that the ≈ number will before long be shown to be sufficiently large; though this expectation may be fulfilled, the principle would not be affected. We should then simply confess to an error, and say that Afroasiatic has been shown to be a family (with Semitic, Egyptian, Berber, Cushitic and Chadic main branches) rather than a phylum. See Carleton T. Hodge, Afroasiatic, to appear in the Current Trends in Linguistics volume which is to go to press (Mouton and Co.) early in 1968.

20. Our application of the mosaic metaphor differs deliberately from that given by Paul Thieme, p. 595 of Dell Hymes, *op. cit.*

21. This was said eight years before Edward Sapir published his revolutionary paper, Central and North American Languages, Encyclopaedia Britannica (14th ed.) 5.138–41 (1929); reprinted in Selected Writings of Edward Sapir in Language, Culture, and Personality, David Mandelbaum, ed., University of California Press (1949). The reference eight years earlier is to p. 195 in Edward Sapir, Language (N. Y. 1921). All page references in the next paragraph are to this book.