We need models to organize our data. Each particular model represents a presupposition about the kinds of interrelationships that obtain among the items being interrelated. The given presupposition should derive from a theory, but always necessarily represents some delineation of relations relevant to some general concern from other relations that are not considered of relevance. The same set of items, of course, may interrelate in other ways, represented in some other manner deriving from some other theory or interests.

I will take a relatively easy and well-understood example as an illustrative case for examining some kinds of problems and payoffs of representational formalism. Here I use language "family trees", as in Figure 1, not for an up-to-date analysis of diachronic linguistics but as a relatively simple example where the formalism problem can be made evident. Similar analytic points could be made in areas such as ethno-biology or kinship. I chose the relatively simple representation of language trees, however, because it is a topic where it is easier to deal with the formalism problem in its own terms and to gain some distance from contested stakes in particular viewpoints.

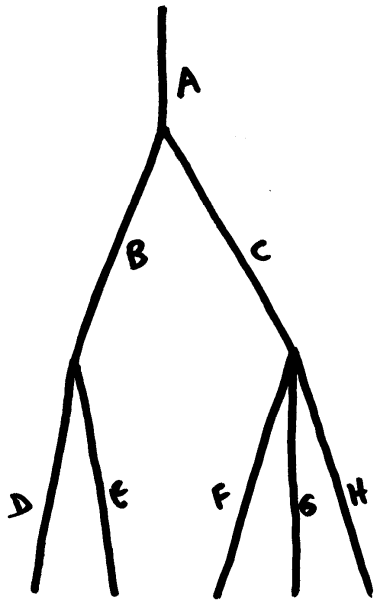


Figure 1: Family Tree with Simple Lines

Traditional discussions both of the ancestral homeland of a language (such as Proto-Indo-European or English)—where it came from—and of when the language split off from its nearest neighbor (such as, for English, from Frisian) are cast in terms of tree diagrams such as that in Figure 2.

Tree (or dendrogram) models have been used by linguists to represent "descent" relations among languages. These models presume that each language has a single parent ("mother"), while each parent may have anywhere from no children ("daughters) to many. Since all languages contain elements from a variety of sources, this combination of model and theory presumes a distinction between elements of the basic system which are present through descent and other elements which are "borrowed" into the system from other sources or which are present via the operation of universal principles (Greenberg's 1968:Chapters VIII and IX on

"universals" including his "implicational universals") or which are present by chance—i.e., where they develop through independent processes in the two languages and where the apparent similarity is only an accidental artifact. A tree diagram is not capable of representing all of the historical processes that one or another theory might specify; it is restricted to processes that produce a one-to-many mapping from source to descendent. A claim, for instance, that some language descends equally from a combination of two or more distinct parent languages cannot be represented in this formalism.

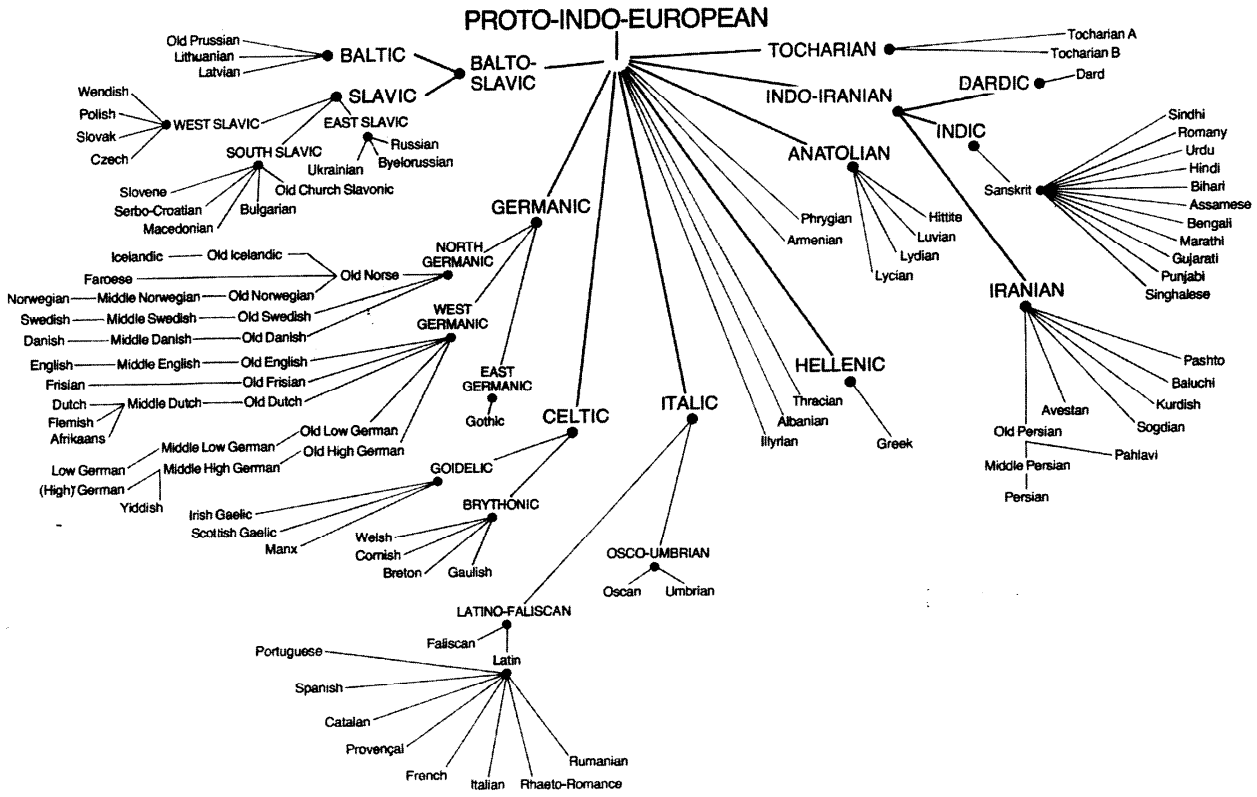


Figure 2: Family Tree of Proto-Indo-European from American Heritage Dictionary (partial reproduction from the inside back cover, Morris 1978)

Interpretations of tree diagrams have typically included an assumption of a smooth transition from minimal dialect differences through increasingly distinctive dialect differences up to dialects that are essentially no longer inter-communicable, and thus which represent different daughter languages of the given mother language. That is, any living language (present or past) entails a community of speakers. Changes in a language arise through parallel innovations by a substantial/sufficient number of members of the relevant speech community. As communities grow, the community of speakers of a given language can become quite large and encompass many local (face-to-face) communities; a given innovation may occur in only one local community (or in a few interacting communities); through this process differing dialects can arise.

Seeing the tree as one of a spread of somewhat diverse speech communities rather than as linear relations among absolute language codes might lead to something like the form of representation shown in Figure 3.

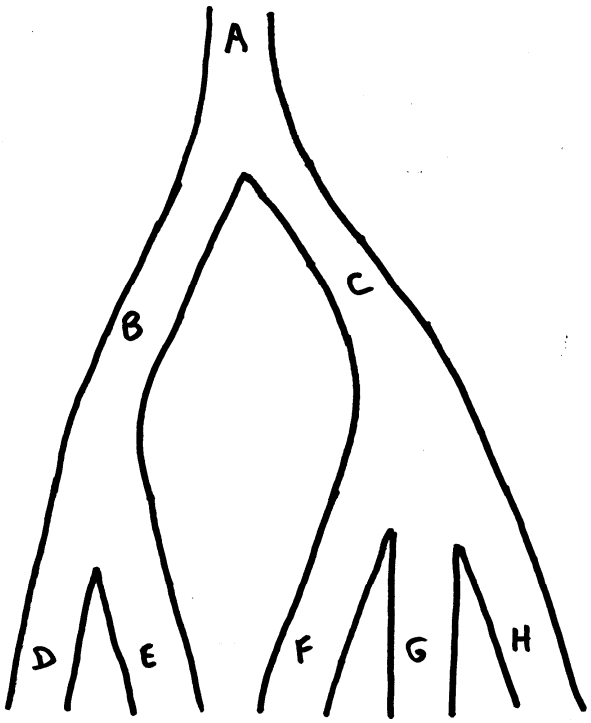


Figure 3: Family Tree with Thick Lines

In terms of such a diagram we might ask ourselves whether the time of divergence question for Languages B and C is to be taken as applying to point 1, point 3, or some point 2 in between, as distinguished in Figure 4. Point 1 is where two dialects first start to diverge These dialects could keep diverging and eventually become Languages B and C (as they do in this drawing), but they also could blend back together into the larger community of Language A, or one could simply die out. Point 3 is the point where they have become clearly distinct languages, and hence not (easily) mutually intelligible[1]. Point 2 represents them when they are somewhat divergent dialects of a single language, and are mutually intelligible.

A variant of the same analytic problem arises when we speak of the “homeland” of Language B. Do we mean the location of Language A when the dialects that were to become Language B began being differentiated, the location where the dialectical ancestor of Language B was spoken, or the place where the speakers of the distinct Language B were first located? In discussions of the homeland of Proto-Indo-European and its immediate daughters such problems

---

[1] Mutual intelligibility winds up being a messy empirical standard when one gets down to details (certainly I've seen it in conversations between Spanish and Italian speakers). That is, mutual intelligibility can be facilitated or made more difficult by relatively incidental attributes of the languages in question, and can, thus, even be asymmetric. For example, the phonology of Portuguese seems to make it hard for many Spanish speakers to understand (even though it is, by most non-political measures, simply another Spanish dialect), while Portuguese speakers seem to experience no comparable difficulty with Spanish. And, for that matter, the equivalent biological "measure", interbreeding, seems to have some very comparable sources of similar messiness.

do seem to be muddying the waters—especially since Proto-Indo-European speakers appear to have been moving and spreading out during this period. But, of course, such spreading may not represent a confounding phenomenon so much as a common condition of language differentiation.

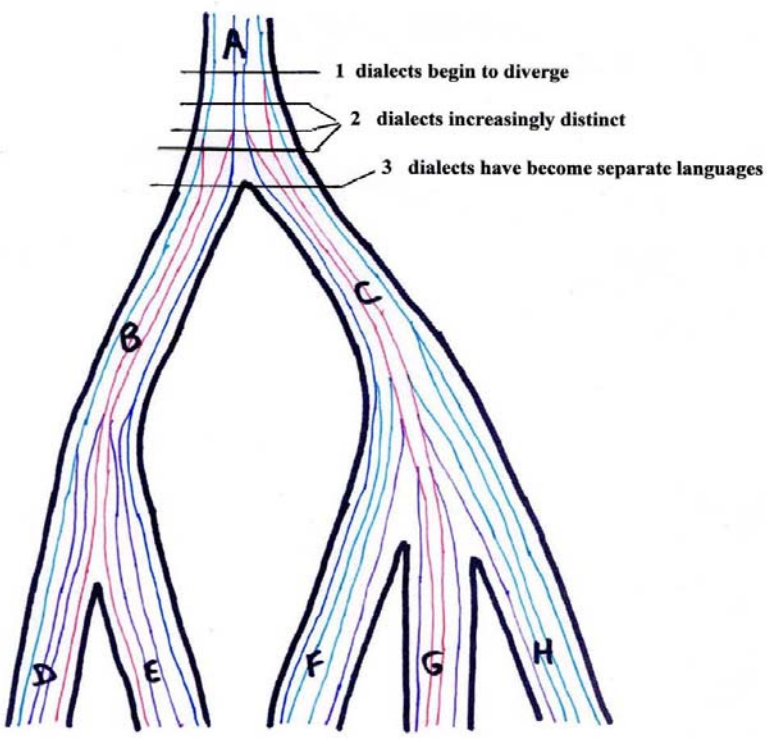


Figure 4: Family Tree with Thick Lines and Time Points 1, 2, 3 Marked (and showing internal dialect lines, where the colors serve to roughly distinguish dialect lines from one another)

This model is often taken also as an approximate representation of the relative time depth at which the various splits (from a single parent language into a set of daughter languages) represented in the tree are considered to have taken place.[2]

There seem to exist at least two kinds of processes of linguistic change which can only be interpreted against a family tree backdrop, but which cannot be adequately represented within such a tree. These are: 1) the creation of pidgins; and 2) the hiving off of a cross-section of a mother-language community (vs. a coherent sub-community) to form a new language community and then language.

---

[2] My colleague Gene Anderson suggests that sometimes 2 closely related but split languages, or 2 divergent dialects of one language "keep in touch and keep borrowing from each other, so that they never diverge more than a certain amount, but never quite fuse either. English and Scots (highland or lowland) is a good example." One might speak of this as "linked drift" or even "co-evolution" (if one doesn't see "evolution" as from 'lower' to 'higher'). Such a process would be another of those leading to a situation where the one-to-many mappings of the tree diagram would be "correct" but where using line length as a representative of linked notions of divergence and time of divergence would not work—because "distance apart" and "length of time of separation" would be either uncoupled or the back and forth adaptive borrowing would significantly undermine the very notion of any clean divergence—if it applied to distinct languages and not just to dialects.

Regarding the first process, pidginization, I am following Frake (classroom notes from ca. 1964-5) in seeing a pidgin language not as an amalgam of several languages or as the joining of an X "father's" vocabulary with a Y "mother's" grammar, but as a new language constructed by combining vocabulary from an existing language with a created grammar.[3] Frake suggested that such situations arise in conditions (seen for instance in slavery) where a dominant community tightly controls an under-community, where the members of the under-community are drawn from a wide enough array of different language communities to share no common language of their own and where members of the dominant community use their own language for basic commands and instructions. Members of the under-community are thus forced to learn some basic vocabulary of the dominant community but are not exposed to the dominant community language enough for its members to actually learn that language. The result is that the under-community is forced to create its own language out of the vocabulary its members get from the dominant community.[4] My sense is, furthermore, that presumably there exist some other social situations, in addition to slavery, which produce this kind of restricted communication bottleneck[5]. Additionally I would like to suggest that such restricted communication bottlenecks may, in some situations, be a matter of degree rather than all-or-nothing. That is, the new creation may not be totally *de novo*; it seems possible that some elemental parts of morphology and syntax of the superordinate language might have come with the vocabulary—and some structural elements shared across the subordinate ones might also come into the new creation. Greenberg's work on implicational universals (1966, 1968) gives some hint on how a "created" syntax might start out, and then evolve.

We mostly think of pidgins in connection with colonial contact theaters such as the Caribbean or West Africa or Southeast Asia or New Guinea. But, for a long time I have been aware of how strange English is as a Germanic language—particularly syntactically in its loss of most inflections and reliance on a relatively simple word order and lexically in its tremendous borrowing from Romance (particularly, French[6]). The thought occurred that English might be seen as a kind of pidginized Anglo-Saxon. Since I'm not a diachronic linguist, my musings were not of much significance, but then I saw that S. S. Mufwene (2001), then Chair of the Department of Linguistics at the University of Chicago offered a similar observation, as also do Patricia Poussa (1982), Manfred Görlach (1986), and John Hines (1991). (And, now, to stay ahead of the curve, I offer a similar musing—that French is a strange Romance language in much the same way that English is a strange Germanic one!).

---

[3] That is, pidgins are not any kind of mixed language—even if they may, like other languages, be influenced by a variety of languages that their speakers encounter.

I am not getting into the pidgin/creole distinction at this time. One traditional view was that the new language, when spoken by all its users as a second language, was a pidgin, but became a creole when it became the first language of some "native speakers". Another view has creole being more complete, and maybe hewing closer to the source language, if still with much divergent local flavor.

[4] One could imagine that any grammatical elements that most of the subordinate language communities shared might, in some manner, find their way into the newly created language—and that in that sense their new language might not be starting totally from scratch.

[5] Gabriella Rundblad (personal communication), a historical linguist at Kings College, London, suggests the possibility that sometimes immigrant communities come up with a pidgin-like version of the language whose realm they are immigrating into.

[6] My reference to French does NOT mean that I am claiming any kind of "2 parent" origin for English. I only adduce it as evidence for the difficulties that arose at the time in the diachronic transmission of English.

Our representational formalism currently allows no distinction of different kinds of descent relations. One can imagine conventions by which such relations might be represented, but the logical and empirical implications of such changes in the formalism are not yet clear. Methodologically, for diachronic linguistics, this variant of the pidgin problem, if it were accepted, could pose problems for deeper classification, where the kinds of detail that would enable one to distinguish normal inheritance for a pidgin offshoot might not be available.

Pidgins appear to go in the family tree because their vocabulary links them to a language in the tree. The extent of the apparent difference between the pidgin and its source, and thus the degree to which the pidgin appears to belong on the source's family tree, depends in part on how the source's syntax matches up with the created syntax of the pidgin. Whether or not the pidgin *should* go on the source's tree depends on how one wants to use the tree—as a summary of all kinds of descent relations then "yes", or as a summary of the cumulative effects of systematic normal change (linguistic "drift", if you will) then "no". But such a "no", to be meaningful, depends on a way of consistently telling the two apart. Figure 5 offers one way in which such pidginization might be represented graphically.

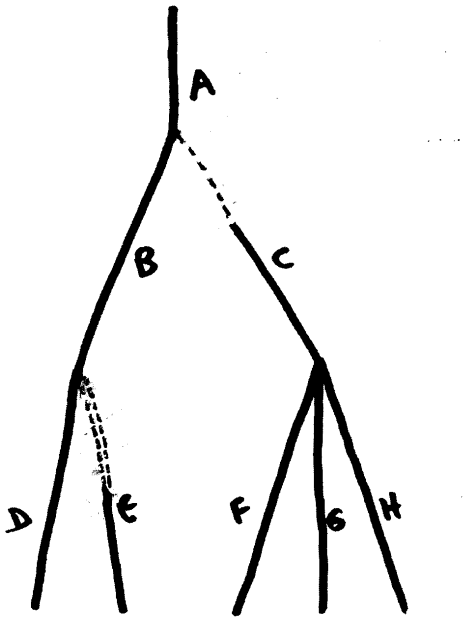


Figure 5: Family Tree with Pidgin Lines representing different degrees of pidginization

Key: In this arbitrary example, single dashes, ------, signal relatively little pidginization, while double dashes, ==== , signal greater pidginization, where the issue concerns how much of the parent language grammar survived the social process that produced the pidgin.

Turning now to the second process, cross-sectional hiving off, Iceland was founded by a collection of Vikings from all over Norway, rather than by settlers from only one or two source communities.[7] This history meant that, initially, there existed no distinct Icelandic dialect, and

---

[7] This paper in general, including this particular example, is intended as a delineation of problems involving formal diagrammatic representations of theoretical claims. While I draw this example from Hymes (1960), who draws on

that for many years something like momentum kept changes in Icelandic parallel to those in Norwegian. Since Iceland was much smaller, though,[8] it did have more interaction among descendants of various Norwegian dialect communities that obtained in Norway, leading gradually to a distinctive Icelandic dialect; the physical separation of Icelandic from Norwegian also meant that normal processes of linguistic drift also gradually moved the two communities apart. But the founding situation meant that it took much longer for Icelandic to emerge as a separate language than it would have had it been founded by members of an already defined speech community. This slowness or delay also explains the glottochronological "finding" that Icelandic has apparently just now split, even though we know from history that the split is 1000 years old. The case is not—as it sometimes is presented—a disconfirmation of glottochronology, but it is does offer a strong warning of the dependence of glottochronological dates on the sociolinguistic conditions at the split—especially for really shallow time depths.[9] The cross-sectional founding of a new community creates a situation in which it takes a while for that new local community to stabilize itself and to begin showing any innovations not present also in the source language. This situation is quite unlike the normal process by which local dialects gradually become ever more distinct (unless something re-connects them to some national stream). In this kind of situation, the (what one might label as a) "drifting apart clock" starts well before the final split, and presumably in some number of cases local communities begin to drift off but then get folded back into the larger community (meaning that the clock then has to restart)—see Figure 6. Thus, communities that hive off already have a fair amount of linguistic distinctiveness, even if not enough to be considered separate languages. But in the cross-sectional (or the Iceland-type) founding situation, however, there is no history of sub-language distinctiveness, and so the drifting apart clock does not start until the new community has been established—see Figure 7. Thus, this sort of situation would seem to explain why glottochronology shows Icelandic as only now differentiating from Norwegian—in spite of a well-attested thousand years of separation!

==========

Bergsland (1958), my emphasis is not on substantive claims in these or other examples. For illustrative purposes I have made use of an abstracted and simplified version of the Icelandic situation—including presumed historical relations among the relevant language communities and linguistic claims and interpretations regarding the effects of those relations on apparent retention rates, given the known ca. a thousand years actual separation.

[8] Also, it seems that most of the settlement was clustered around the capitol—which also facilitated increased interaction among speakers of diverse dialects.

[9] We also know that the resolving power of glottochronology's picture gets to "grainy" to be distinguishable form chance when time depths get moderately deep—making it a tool of limited usefulness.

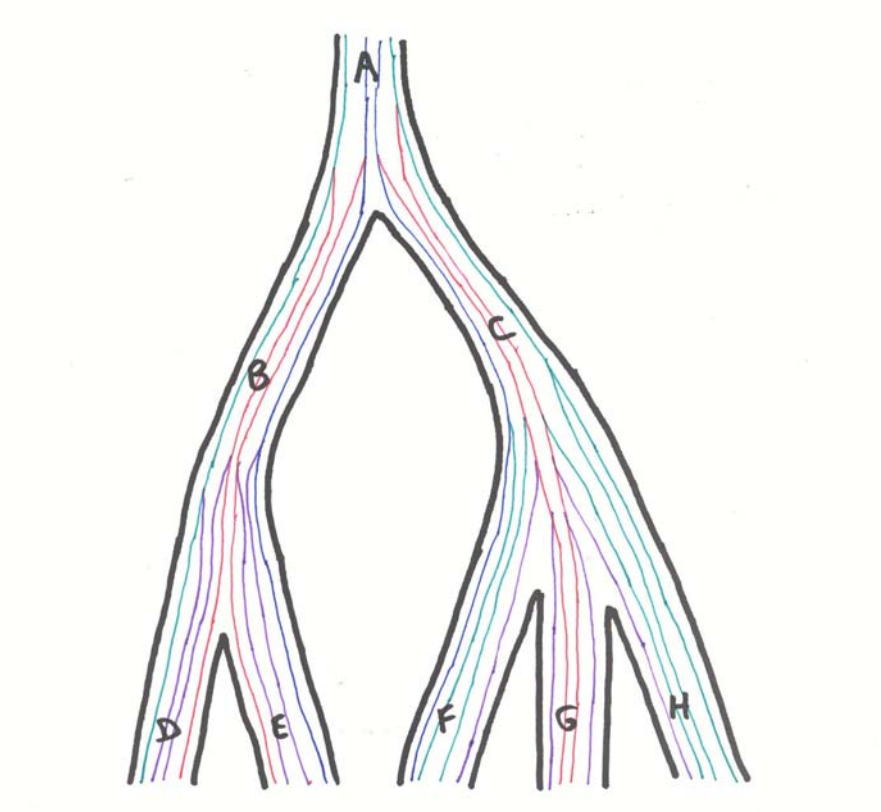


Figure 6: Family Tree with Internal Dialect lines

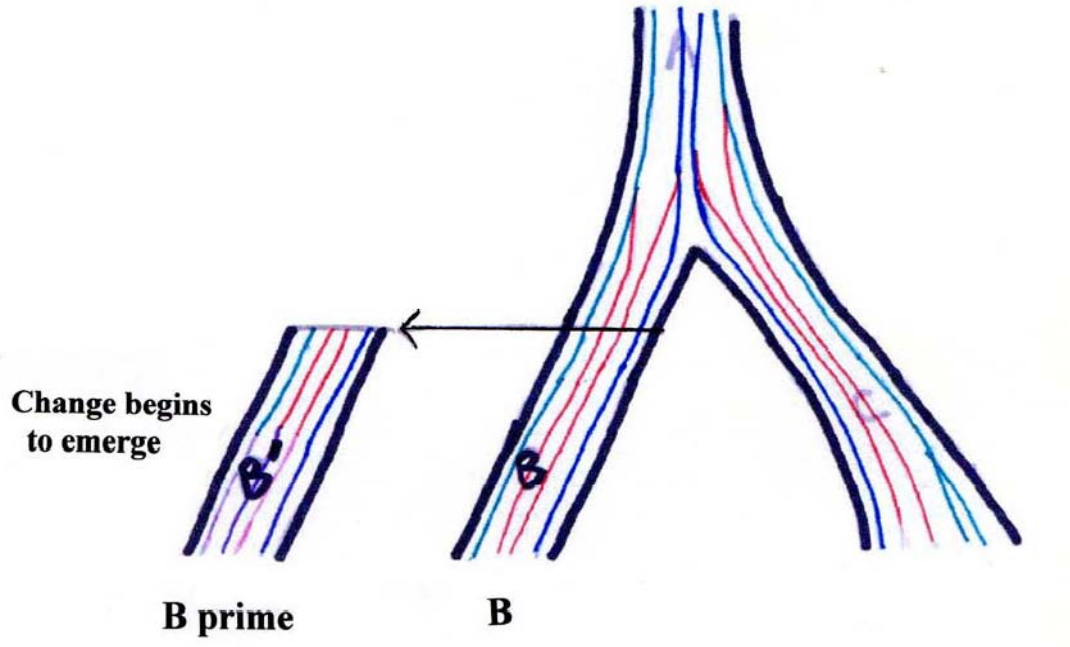


Figure 7: Family Tree with Dialect Lines and Cross-Sectional Split

In both of the kinds of special cases shown in Figures 5 and 7 a tree diagram (such as that in Figure 1) does not misrepresent the basic one-to-many structure (and so both situations could be represented within it), but it does mislead in two ways. First, as such diagrams are normally read, each node is considered to be the historical source of two or more normally related daughter languages. All nodes in the tree are treated as uniformly dominating similar kinds of branchings. Thus, the relations implied by nodes are homogeneous—all are considered to be of the same sort—relative to the kind and degree of relatedness. One could perhaps fix this problem by introducing multiple kinds of branchings from nodes—at the cost of greatly complicating the interpretation of the resulting tree. Second, as the diagrams are normally read, the length of a line between nodes represents the degree of separation—roughly jointly in terms of time and amount of movement; that is, the lines are presumed to represent a smooth historical process. There do exist ways of complicating the tree diagrams to take account of this kind of information—as indeed I have offered in Figures 5 and 7—but such conventions are not part of the conventional format for such diagrams.

A claim or finding that it was possible for languages to have two or more parents—whatever one meant by "parent" and however one defined it—would be logically incompatible with the tree model and thus would necessitate some other form of model. An attempt to summarize, say, borrowing relations among language would seem to necessitate some kind of network, as suggested in two different ways in Figure 8; perhaps some future systematist will be able to come up with a way of showing network relations among nodes of trees (both intra and inter any given tree)[10] ?

The special cases, while not contradicting the tree representation, do, of course, impinge on any attempt to use the tree as a homogeneous representation of consistent past relations—as, for example, a tool for recovering earlier forms of ancestral language; how much such cases do impinge depends on knowing how rare or frequent they are, and this seems still uncertain. The move by some linguists to treat English as a pidginized Germanic language illustrates that the question is being pursued—and raises the possibility of pidginization perhaps being a matter of degree rather than an all-or-nothing phenomenon that I spoke of above.

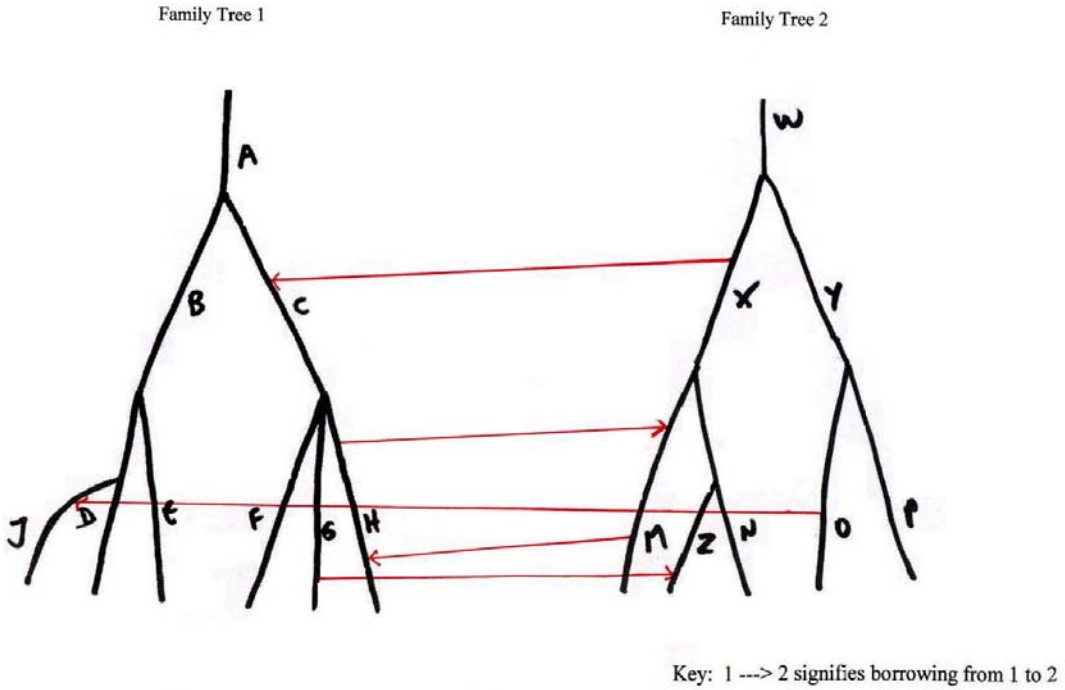
What I have been concerned with specifically in this paper are the kinds of formal representation of relations entailed by theoretical approaches to language change. These theory-oriented concerns explain both my reasons for exploring alternative representations and my reticence in pushing them too far. In that context I have considered how some variant nuances might interact with the representation.

I have not directly considered any formal theory, nor have I addressed the more general cases of other non-linguistic processes that are represented with tree diagrams or of such processes that are not model-able with tree diagrams. The formal representation problem, however, is not restricted to historical linguistics and is basic to any modeling exercise. Problems involving the basic structural difference between tree diagrams and multi-dimensional scaling pictures were discussed in some detail in Kronenfeld (1985). The seriousness and importance of the problem was illustrated there with an extended example from an eminent genetic systematist; in this regard see also Fix (2000). A formal representation should *not* be inconsistent with the

---

[10] I'm told that there is some movement by linguists concerned with borrowing to combine tree models with wave models. These should be interesting. Since I am unfamiliar with these models, I cannot speak to how formal or precise their representations are intended to be, nor to how clearly they relate to a formal theory (or, at least, to formal theoretical propositions).

theory on which it is based, and, ideally, it should be implied by the theory and interpretable in terms of it. At the same time, I suggest, using explicit formal representations of posited relationships and taking the representations seriously can provide a means for probing a theory and possibly developing new theory.



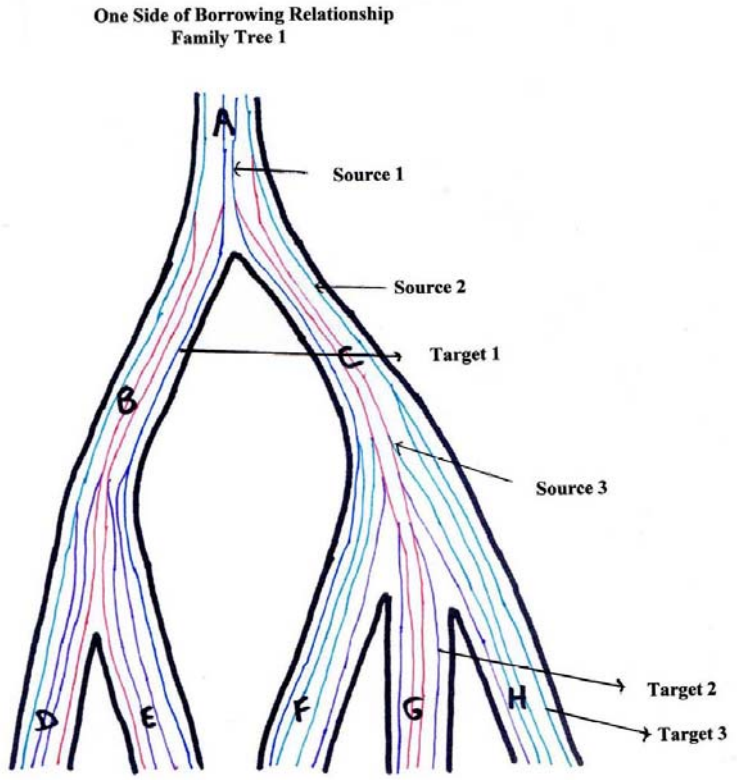


Figure 8: Two Forms of Family Tree with Borrowing Lines Included

References

Bergsland, Knut. 1958 Is Lexico-statistic dating valid? *Proceedings of the Thirty-Second International Congress of Americanists, Copenhagen 1956*. Copenhagen: Munksgaard. Pp. 654-675.

Fix, Alan G. 2000 Genes. Language, and Ethnic Groups: Reconstructing Orang Asli Prehistory. *Indo-Pacific Prehistory Association Bulletin* 19:11-16

Görlach, Manfred. 1986 Middle English – a creole? *Linguistics Across Historical and Geographical Boundaries*. Eds. Dieter Kastovsky and Aleksander Szwedek. Berlin: Mouton de Gruyter. Pp. 329-344.

Greenberg, Joseph H. 1968 *Anthropological Linguistics: An Introduction*. New York: Random House.

———— 1966 *Language Universals with Special Reference to Feature Hierarchies.* The Hague: Mouton (Janua Linguarum No. 59) .

Hines, John. 1991 Scandinavian English: A Creole In Context. *In*, P. Sture Ureland and George Broderick, pp. 403-427. Language Contact in the British Isles. *Proc. of the Eighth International Symposium on Language Contact in Europe,* Douglas, Isle of Man. Tübingen: Niemeyer.

Hymes, D. H. 1960 Lexicostatistics So Far. *Current Anthropology* 1:3-44.

Kronenfeld, David B. 1985 Numerical Taxonomy: Old Techniques and New Assumptions. *Current Anthropology* 26:21-41.

Morris, William (editor). 1978 (College Edition) *The American Heritage Dictionary of the English Language*, Boston: Houghton Mifflin Company.

Mufwene, Salikoko S. 2001 *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.

Poussa, Patricia. 1982 The Evolution of Early Standard English: The Creolization Hypothesis. *Studia Anglica Posnaniensia* 14:69-85.