Glottochronometry: some critical considerations in general and for sub-Saharan Africa in particular

John Sharman

> "Anthropologists who are not linguists have been largely taken in by glottochronology and should be warned that any conclusions they have come to on this basis are almost certain to be in error."
> – Trager, Current Anthropology, 1962, no.2

> "It is striking how critics of glottochronology continue to rediscover the same criticisms, but not the constructive attempts to deal with them."
> – Hymes, Current Anthropology, 1966, no.4

Introduction

I believe glottochronometry may perhaps yet have much help to give in most parts of Africa: the general idea is clear and reasonable enough.* But I am troubled about the 'African' applicability of the 'standard' test-lists, and values in the (old) formula, both of which (i.e. lists and values) still continue to be enthusiastically used here unaltered and unadapted, in spite of the years of criticism and constructive commentary. I have an uneasy feeling that many non-linguists really do set too much store by the accuracy of the resultant dates, when in fact there is often little justification for their doing so. What follows is a simple superficial attempt to set investigators and the more credulous historians on their guard, and at the same time to point out how much more subtle and sophisticated the study has become since Trager wrote the above-quoted piece of polemics.

Hence I shall be first concerned with general principles, and not so much with detail; later, I shall come to some mathematical considerations which derive directly from the basic assumptions and the word lists themselves. As a preliminary short-cut, I would suggest that it may be useful to read the p39 summary first.

---

* I have discussed elsewhere a possible further step (Research/Staff Seminar, Institute of African Studies/Department of linguistics and African languages, University of Nairobi – October 1972).

Basic Considerations

First, it is necessary to say that not all the early foundations of glottochronometry were solid. The first and second basic assumptions (that certain vocabulary items are less subject to change-in-time than others, and that the retention-rate of the more stable items is (roughly) constant through time must be (roughly) taken for granted in order to make a start at all) although the second assumption "has not been checked for a time-span greater than 2,200 years and this span does not provide adequate evidence for a constant rate of loss over a long period of time" (Gudschinsky, 1956), and in any case, even within the 'basic vocabulary', different items have different retention rates.

"The third basic assumption ... is that the rate of loss of basic vocabulary is approximately the same in all languages" (Gudschinsky, ibid.). But: "This assumption has been tested in thirteen languages* in which there are historical records" (ibid.). One should perhaps say 'only thirteen languages' - languages in which (naturally) there had to be historical records; this in itself differentiates the thirteen sharply from non-recorded languages. Even within this special group, "the results range from a retention of 86.4% to 74.4% per thousand years" (ibid.) (This alone is enough to make a difference of two to one in time units.) The average for the thirteen was 80.5% (Lees, 1953).** Furthermore, "This is not ... conclusive evidence that all languages change at this rate, especially since all but two of the ... languages tested are Indo-European" (ibid., my underlining). Although the 'basic vocabulary' is not supposed to contain overtly 'cultural' items, nevertheless this eco-cultural and linguistic bias is quite certain to give non-universal retention-rates (a) because of the Indo-European preponderance (see especially the fourth assumption below), and (b) because of the very fact of being written: clearly, languages being written down cannot change at the same rate as languages transmitted only by word of mouth. This last interlocks

---

*The long-written languages (Lees, 1953) are: Athenian, Catalan, Chinese, Coptic, Cypriote, English, French, German, Italian, Portuguese, Rumanian, Spanish, Swedish. Studies were later made of Arabic, Japanese and Kannada (Dravidian).

**An 'average' among only 13 languages is really mathematically naughty: among (say) 130, it might make better sense - we might get a reasonable histogram. But where in a field of 13 we have a variation of 74.4% to 86.4%, the term 'average' is itself loaded. There is 'peaking' in retention rate at around 81%, but individual deviations are really much too wide for us simply to be able to impose '81%' on just whatever languages we happen to be investigating. (Compare especially what happens in Bergsland and Vogt (1962)) (See below, p. 31).

with the final proviso, contained in the fourth assumption, which "is a corollary of the third ...; if the percentage of true cognates ... is known for any pair of languages, the length of time since (divergence) can be computed, provided that there are no interfering factors through migrations, conquests or other social contacts which slowed or speeded the divergence" (Swadesh, 1950; Gudschinsky, 1955, 1956) (my underlining: what 'earlier' Indo-European languages would fit such a proviso?) Application of a retention rate derived from only 13 written languages, eleven of them Indo-European, with the last proviso, to (say) a body of hitherto unwritten African languages, known to fall outside the proviso terms, seems a dubious procedure. (Personally, I would be fairly happy to scrap the proviso altogether, on the grounds that most languages have never been for very long in such an isolated state.)

However, let us temporarily assume that we accept a retention rate that lies very roughly between 74% and 87% (these limits may be yet further apart in individual languages within one family, as we shall see); then "a convenient list (of lexical terms) for this purpose is Swadesh's 200 word list ... tentatively tested for percentage retention in languages with written historical records. Later tests may well indicate that a different assortment of words would be more useful ..." (ibid.) "Kroeber (1955) has suggested that a list of 1000 items would be preferable and doubts that deep time-depths can be explored by use of a list as small as 200 items' (ibid.).* However, Swadesh (1955) himself at about the same time evolved a yet shorter 'diagnostic' list of 100 items, together with a further 'supplementary' 100, to which I shall be returning later.

The 200 was "set up largely in terms of Indo-European languages and cultures ... and has been refined and improved ... by ... application to non-European languages and cultures" (Hoijer, 1956a). The 100 item list (with its supplementaries if necessary) is therefore better for use in glottochronometry generally than is the original 200, although several recent pieces of work are based on the latter. "The goal is to constitute an LCD of the basic vocabulary in the languages of the world" (Hymes, 1960, p.7).

"Greater time depths may be explored ... if the list is filled in with the reconstructed forms of the postulated common parent language of a ... family or stock" (ibid. fr. Swadesh, 1953a). It so happens that we can do this with some accuracy for many items in Common Bantu, but unfortunately we do not yet possess similar lists for Common Nilotic, Common Cushitic and other groups that are of immediate local interest.

---

*This objection later falls away, because 'deep' time depths cannot really ever be explored at all: it looks as if 5000 B.P. is about the meaningful limit, even with best techniques, improved formulae, and use of computers.

The 'old' Formula: Implications and Caveats

Given all the foregoing somewhat shaky underpinning, it was (and sometimes still is, at least 'locally') normal to use the Swadesh refined 100 or the older 200 word lists, together with the formula t = log C / J log r, where J* = (Hattori) 1.411 (not the earlier 2 of Swadesh), and r = 0.86 (Swadesh) for the 100 list, or 0.805 (Lees) for the earlier 200 list. But J is still a dubious factor, in spite of its later four-significant-figure appearance (to use four significant figures is in any case to introduce a spurious degree of accuracy); the formula is naturally very much more sensitive to the values of C and r, which need only change slightly to give us vastly different results.

Because of this high sensitivity, it is very important to be sure of three things about C:- (i) absolute accuracy of morph-phoneme/toneme correspondence for possible cognates (in this respect, I have a feeling that much of the work on non-written languages - both inside and outside Africa - has been more or less 'rough'). Indeed Haudricourt (1966) goes so far as to say, "With only 200 words of a language, it is impossible to get a correct notation or to determine the phonetic laws which permit one to relate each word on the list with a word in another language; for this at least 2,000 words and a great deal of work are required." Possibly true, but only true if all that you have is a 200 word list of 'meanings'; if you already know the relevant sound shifts, the argument vanishes - we could therefore get perfectly good comparisons within Bantu, but nothing like as good for our other sub-Saharan language-groups, where most work has had to depend on 'resemblances', thus increasing the possibilities of error in selecting of cognates and non-cognates;

---

*J is the symbol for a variable constant, once called the 'jiggle-factor': the factor that Lees implies you use 'to make things come right'. I have been calling this 'k' (for 'a variable constant'), but in view of the great risk of introducing confusion (a criticism rather justifiably levelled at van der Merwe (C.A., 1966, no.4)), and especially since k has been used in two important different ways already, it is obviously better to revert to J. (Not J for 'jiggle' but for adjustment-factor' - c.f. Hymes, C.A., 1960, no.1.) Note the flatly contrasting views of Izui (C.A., 1962, no.2): "The number ('2') is mathematically not tinkerable at all. If we change it, the formula will lose the mathematical basis on which it stands", and Hymes (C.A., 1964, no.4): "The number represented (i.e. in his own 1960 paper) a theoretical variable (as Hattori implies in his revised formula)". Pace Izui, it does seem really quite evident that as things stand, J is a genuine variable, "whose appropriateness (is) ... a matter of sociolinguistic assumptions" (Hymes ibid.). Uncritical acceptance of any one value as being valid (for some 'new' area of investigation) will introduce distortion in t. In other words, if we wish to retain '2', we have to select some different r, according to the languages concerned.

(ii) a clear understanding of how (indeed whether) the 'meanings' in the test lists correspond to the supposed equivalent in the languages being tested; and (iii) the relative applicability of the test items to the region tested.

For r, matters are worse: r has so far been based on post facto calculations derived from languages long-enough written down and subsequently 'averaged'. There may indeed have come subsequent support from archaeological evidence in non-written language areas - but at least 5,000 years of C-14 datings are all being revamped by dendrochronology. So an 'overall' r has so far been reliably estimated only for cultures sufficiently old in literacy to make the required answer already known. Even in individual known (written) cases, r may vary between e.g. 67.8 (English/Old German) Swadesh, 1955, and Hymes 1960 - actually Arndt, 1955, 1959; and 97.3 (Icelandic/Old Norse). (Bergsland and Vogt, cited by them in C.A., 1962/3.) These two languages belong to the same basic family, but their retention rates would give differences of one millennium to every sixteen or seventeen, which is absurd: we know that in fact the real datings very much the same.

There remains the third delicate question of the by now apparently automatic 'local' acceptance of the 'old' (if refined) underlying word lists themselves. The choice of resistant or 'hard' items for non-literate cultures remains largely guesswork. It seems we cannot simply be even semi-intuitive; we might then merely assume that items like "who, what", or "all, many", or "foot, hand, knee", or "red, green, yellow", or "new, good" are obvious and unassailable - whereas we know that they are not. Where not guesswork, choice must after all and once again be at least largely based on the resistant words in long-recorded languages, or else post facto on a synchronic study of a large number of languages. For example, Dyen (1964, 1965) made a computer-comparison of common retentions among 89 languages in Austronesia.* Van der Merwe (1966) commented: "An awesome amount of data has been computed through this study..." and again, "it is not merely an opinion, but a demonstrated fact (backed by a mountain of statistics) that the process of morpheme decay can be described universally within specified limits of probability." Since then, Dyen, James and Cole, dividing the list into 9 sub-groups, within which every item is separately weighted, and using yet more sophisticated mathematics, have greatly enhanced the accuracy and extent of the lexicostatistical study of the Austronesian languages previously

---

*It was at this stage (actually between 1962 and 1964) that it was 'found' that the test list "could be divided into three equal groups, each exhibiting different viability." Of course, as Chretien (1965) says, "It is obvious that Dyen might have divided the list into fourths or fifths or any convenient number." The practical possibility of dividing it into single items (in the Dyen case, 196 of them) did not arise until the requisite formulae

(Continued...)

examined (see _Language_ 43, 1, 1967). In compiling the "awesome amount of data", and in refuting "beyond a shadow of doubt" the idea that "the probability of replacement of a given lexical item ... defies a universally applicable statement", Dyen was still using the old 200 list.

The feed-in of long-recorded languages is the same kind of operation as that used for the determination of the 'old' version of _r_. There is obviously a built-in cultural skew-tendency in this: with Hoijer, I doubt that we can presently produce a 100 word list so subtly precise as to deal adequately and equally with all our world cultures at once. At the 200 level, we could be well-nigh lost in all manner of complexities. Would not Kroeber's suggested 1000 list be far too unwieldy - or 'uncritical', or even 'mushy'? Are _any_ such lists realistic?

Comparisons between the 100 and 200 test lists and some of the best-known high-frequency lists in English (Hymes, _Current Anthropology_, 1960,1) "(indicate) a strong positive correlation ..." True, but the percentages of the 'first 500' found in the 100-item test list are: Lorge (570 words) 63%, Eaton 64%, Rinsland 51%. These figures are not very different for my result for the most common 500-600 Common Bantu starred form items (69% minus any of 7 queries); and we would expect to find a much stronger affinity than this, _unless_ the 100 item list is more culture-bound than it should be, or, conversely, unless we actually require a _more_ culture-bound list for work in certain parts of Africa.

Hoijer (1953a) says: "(1) (the test list) was originally set up largely in terms of European languages and cultures ... (2) ... it has been refined and improved in quality mainly by its application to non-European languages and cultures ... this (need) was not discovered until attempts had been made to translate the list into a number of divergent languages. It seems likely ... that the difficulties ... will occur again, as the list is ... applied to still other languages. There is, in short, nothing in lexico-statistic theory which enables us, once and for all, to establish a firm test list translatable ... into any language."

---

were introduced (and the computer became more readily available); it had in fact been known from the beginning that the lists could and probably should be so divided (i.e. each item having its own _r_), but noone could then conceive a practical way of doing it; and the old formula was by its very nature incapable of dealing with such a request. It is the continued use of this formula, and the Lees-type illegitimate (and for most unwritten languages, irrelevant) 'average' _r_, which is so misleading today. See below for detail (p.34).

with the final proviso, contained in the fourth assumption, which "is a corollary of the third ...; if the percentage of true cognates ... is known for any pair of languages, the length of time since (divergence) can be computed, provided that there are no interfering factors through migrations, conquests or other social contacts which slowed or speeded the divergence" (Swadesh, 1950; Gudschinsky, 1955, 1956) (my underlining: what 'earlier' Indo-European languages would fit such a proviso?) Application of a retention rate derived from only 13 written languages, eleven of them Indo-European, with the last proviso, to (say) a body of hitherto unwritten African languages, known to fall outside the proviso terms, seems a dubious procedure. (Personally, I would be fairly happy to scrap the proviso altogether, on the grounds that most languages have never been for very long in such an isolated state.)

However, let us temporarily assume that we accept a retention rate that lies very roughly between 74% and 87% (these limits may be yet further apart in individual languages within one family, as we shall see); then "a convenient list (of lexical terms) for this purpose is Swadesh's 200 word list ... tentatively tested for percentage retention in languages with written historical records. Later tests may well indicate that a different assortment of words would be more useful ..." (*ibid.*) "Kroeber (1955) has suggested that a list of 1000 items would be preferable and doubts that deep time-depths can be explored by use of a list as small as 200 items' (ibid.).* However, Swadesh (1955) himself at about the same time evolved a yet shorter 'diagnostic' list of 100 items, together with a further 'supplementary' 100, to which I shall be returning later.

The 200 was "set up largely in terms of Indo-European languages and cultures ... and has been refined and improved ... by ... application to non-European languages and cultures" (Hoijer, 1956a). The 100 item list (with its supplementaries if necessary) is therefore better for use in glottochronometry generally than is the original 200, although several recent pieces of work are based on the latter. "The goal is to constitute an LCD of the basic vocabulary in the languages of the world" (Hymes, 1960, p.7).

"Greater time depths may be explored ... if the list is filled in with the reconstructed forms of the postulated common parent language of a ... family or stock" (*ibid.* fr. Swadesh, 1953a). It so happens that we can do this with some accuracy for many items in Common Bantu, but unfortunately we do not yet possess similar lists for Common Nilotic, Common Cushitic and other groups that are of immediate local interest.

---

*This objection later falls away, because 'deep' time depths cannot really ever be explored at all: it looks as if 5000 B.P. is about the meaningful limit, even with best techniques, improved formulae, and use of computers.

The 'old' Formula: Implications and Caveats

Given all the foregoing somewhat shaky underpinning, it was (and sometimes still is, at least 'locally') normal to use the Swadesh refined 100 or the older 200 word lists, together with the formula t = log C / J log r, where J* = (Hattori) 1.411 (not the earlier 2 of Swadesh), and r = 0.86 (Swadesh) for the 100 list, or 0.805 (Lees) for the earlier 200 list. But J is still a dubious factor, in spite of its later four-significant-figure appearance (to use four significant figures is in any case to introduce a spurious degree of accuracy); the formula is naturally very much more sensitive to the values of C and r, which need only change slightly to give us vastly different results.

Because of this high sensitivity, it is very important to be sure of three things about C:- (i) absolute accuracy of morph-phoneme/toneme correspondence for possible cognates (in this respect, I have a feeling that much of the work on non-written languages - both inside and outside Africa - has been more or less 'rough'). Indeed Haudricourt (1966) goes so far as to say, "With only 200 words of a language, it is impossible to get a correct notation or to determine the phonetic laws which permit one to relate each word on the list with a word in another language; for this at least 2,000 words and a great deal of work are required." Possibly true, but only true if all that you have is a 200 word list of 'meanings'; if you already know the relevant sound shifts, the argument vanishes - we could therefore get perfectly good comparisons within Bantu, but nothing like as good for our other sub-Saharan language-groups, where most work has had to depend on 'resemblances', thus increasing the possibilities of error in selecting of cognates and non-cognates;

---

*J is the symbol for a variable constant, once called the 'jiggle-factor': the factor that Lees implies you use 'to make things come right'. I have been calling this 'k' (for 'a variable constant'), but in view of the great risk of introducing confusion (a criticism rather justifiably levelled at van der Merwe (C.A., 1966, no.4)), and especially since k has been used in two important different ways already, it is obviously better to revert to J. (Not J for 'jiggle' but for adjustment-factor' - c.f. Hymes, C.A., 1960, no.1.) Note the flatly contrasting views of Izui (C.A., 1962, no.2): "The number ('2') is mathematically not tinkerable at all. If we change it, the formula will lose the mathematical basis on which it stands", and Hymes (C.A., 1964, no.4): "The number represented (i.e. in his own 1960 paper) a theoretical variable (as Hattori implies in his revised formula)". Pace Izui, it does seem really quite evident that as things stand, J is a genuine variable, "whose appropriateness (is) ... a matter of sociolinguistic assumptions" (Hymes ibid.). Uncritical acceptance of any one value as being valid (for some 'new' area of investigation) will introduce distortion in t. In other words, if we wish to retain '2', we have to select some different r, according to the languages concerned.

(ii) a clear understanding of how (indeed whether) the 'meanings' in the test lists correspond to the supposed equivalent in the languages being tested; and (iii) the relative applicability of the test items to the region tested.

For r, matters are worse: r has so far been based on post facto calculations derived from languages long-enough written down and subsequently 'averaged'. There may indeed have come subsequent support from archaeological evidence in non-written language areas - but at least 5,000 years of C-14 datings are all being revamped by dendrochronology. So an 'overall' r has so far been reliably estimated only for cultures sufficiently old in literacy to make the required answer already known. Even in individual known (written) cases, r may vary between e.g. 67.8 (English/Old German) Swadesh, 1955, and Hymes 1960 - actually Arndt, 1955, 1959; and 97.3 (Icelandic/Old Norse). (Bergsland and Vogt, cited by them in C.A., 1962/3.) These two languages belong to the same basic family, but their retention rates would give differences of one millennium to every sixteen or seventeen, which is absurd: we know that in fact the real datings very much the same.

There remains the third delicate question of the by now apparently automatic 'local' acceptance of the 'old' (if refined) underlying word lists themselves. The choice of resistant or 'hard' items for non-literate cultures remains largely guesswork. It seems we cannot simply be even semi-intuitive; we might then merely assume that items like "who, what", or "all, many", or "foot, hand, knee", or "red, green, yellow", or "new, good" are obvious and unassailable - whereas we know that they are not. Where not guesswork, choice must after all and once again be at least largely based on the resistant words in long-recorded languages, or else post facto on a synchronic study of a large number of languages. For example, Dyen (1964, 1965) made a computer-comparison of common retentions among 89 languages in Austronesia.* Van der Merwe (1966) commented: "An awesome amount of data has been computed through this study..." and again, "it is not merely an opinion, but a demonstrated fact (backed by a mountain of statistics) that the process of morpheme decay can be described universally within specified limits of probability." Since then, Dyen, James and Cole, dividing the list into 9 sub-groups, within which every item is separately weighted, and using yet more sophisticated mathematics, have greatly enhanced the accuracy and extent of the lexicostatistical study of the Austronesian languages previously

---

*It was at this stage (actually between 1962 and 1964) that it was 'found' that the test list "could be divided into three equal groups, each exhibiting different viability." Of course, as Chretien (1965) says, "It is obvious that Dyen might have divided the list into fourths or fifths or any convenient number." The practical possibility of dividing it into single items (in the Dyen case, 196 of them) did not arise until the requisite formulae

(Continued...)

examined (see Language 43, 1, 1967). In compiling the "awesome amount of data", and in refuting "beyond a shadow of doubt" the idea that "the probability of replacement of a given lexical item ... defies a universally applicable statement", Dyen was still using the old 200 list.

The feed-in of long-recorded languages is the same kind of operation as that used for the determination of the 'old' version of r. There is obviously a built-in cultural skew-tendency in this: with Hoijer, I doubt that we can presently produce a 100 word list so subtly precise as to deal adequately and equally with all our world cultures at once. At the 200 level, we could be well-nigh lost in all manner of complexities. Would not Kroeber's suggested 1000 list be far too unwieldy - or 'uncritical', or even 'mushy'? Are any such lists realistic?

Comparisons between the 100 and 200 test lists and some of the best-known high-frequency lists in English (Hymes, Current Anthropology, 1960,1) "(indicate) a strong positive correlation ..." True, but the percentages of the 'first 500' found in the 100-item test list are: Lorge (570 words) 63%, Eaton 64%, Rinsland 51%. These figures are not very different for my result for the most common 500-600 Common Bantu starred form items (69% minus any of 7 queries); and we would expect to find a much stronger affinity than this, unless the 100 item list is more culture-bound than it should be, or, conversely, unless we actually require a more culture-bound list for work in certain parts of Africa.

Hoijer (1953a) says: "(1) (the test list) was originally set up largely in terms of European languages and cultures ... (2) ... it has been refined and improved in quality mainly by its application to non-European languages and cultures ... this (need) was not discovered until attempts had been made to translate the list into a number of divergent languages. It seems likely ... that the difficulties ... will occur again, as the list is ... applied to still other languages. There is, in short, nothing in lexico-statistic theory which enables us, once and for all, to establish a firm test list translatable ... into any language."

---

were introduced (and the computer became more readily available); it had in fact been known from the beginning that the lists could and probably should be so divided (i.e. each item having its own r), but noone could then conceive a practical way of doing it; and the old formula was by its very nature incapable of dealing with such a request. It is the continued use of this formula, and the Lees-type illegitimate (and for most unwritten languages, irrelevant) 'average' r, which is so misleading today. See below for detail (p.34).

The discrepancy between the 100 list and the 5/600 CB list certainly suggests that 'the difficulties' have indeed 'occurred again' - and if only we had similar * form lists for other African groups we could probably demonstrate that similar difficulties have occurred elsewhere - but have simply gone unnoticed, or have been obscured, or even ignored in order to 'get results'.

With a formula so sensitive that the striking out or inclusion of a single item in C makes a difference of ± 100 years at a depth of 2000 B.P. (itself subject to standard error), the lists are themselves crucial, and perhaps as difficult to determine for a given cultural/ecological sphere as is the value of r. Also, they themselves naturally affect the determination of r, which is (as we have seen) very sensitive indeed.

Readiness to accept an arbitrary (and constant) cross-cultural number of years for 'one generation' is absurdly naive enough - yet several historians have accepted such a number.* Readiness to accept the apparent authority of a good-looking mathematical formula is all too dangerously close to an active desire to be 'blinded by science' ... when even the scientists themselves should and do (severally and often) admit they are more than just a little doubtful.

A practical suggestion and a query: if the oral-tradition historians and the archaeologists can give us reasonably accurate datings for about the last 500 years 'plus' (including confirmation from cross-correspondences, eclipses, comets; carbon 14, with the recent corrections from tree-ring comparisons...)[1], we might be able to establish plausible dates for the more recent language-splits, and then extrapolate backwards from the first 500 years. Meanwhile, we clearly need to discover what if anything we should do about the basic vocabularies. Do they indeed require readjustment as seems to be suggested by the discrepancy outlined above? How much effect does living in a different culture and environment have on one's hard-core lists? We should talk it all over: simply pressing on regardless is probably not just a waste of time, but misleading to quite a lot of unsuspecting scholars in other fields. (The question of lists applies just as much to lexico-statistics as to glottochronometry.)

---

*Such 'popular' figures applied to the 'English' monarchy (over the last 900 years) are between 33% and 27% too high. All hazards considered, a generation in tropical Africa is likely to have been even less than the 'English' 22 years - say half the usually accepted figure?

[1]Note that any archaeological date-evidence for the last five thousand years which has been based on earlier (i.e. pre-revised) C-14 dating is now to be discounted: beyond this, matters certainly get worse rather than better. Hence any previous 'agreement' between such incorrect C-14 dating and glottochronometry merely disproves

(Continued...)

The practical dangers inherent in the 'older' techniques

Take the 'old' equation: t = log C/J log r.

Obviously, C may easily vary by a few items. Let us say 'only' 5 in 100 that we have collected are uncertain or somehow wrong (by plausible though false 'resemblance', unrecognised borrowings, or inability to pin down truly identical 'meanings'). Let us suppose 50% and 55% as possible limits. Plausible time-depth in round figures is then already between 2800 + 900 and 3200 + 900.

J may conceivably be more or less than Hattori's 1.4(11) for a given ecosphere or cultural area: after all, Swadesh himself (quite rationally) started with 2. Let us take these themselves as example limits.

r can certainly not be determined within the claimed limits for any or all languages: let us take fairly extreme cases of 68% and 97% retention rate for our example (the English/Icelandic figures). After all, we don't yet know quite what would happen if we chose 100 super-hard-core Common Bantu words in order to determine r, but it is pretty certain that we would be up towards the 95% end. What manner of r to use in comparing two Bantu languages using the Swadesh 100 list is to me as yet really an unintelligible question.

However, taking every limit to the limit, and assuming the error margin of 5 items at the 50% level, either:

$$\text{(i) } t = \frac{\log 55}{2 \log 68} = \text{about 770 years B.P.}$$

$$\text{or (ii) } t = \frac{\log 50}{1.4 \log 97} = \text{about 16,500 years B.P.}$$

Even taking the limits actually found in the 13 tested long-literate languages, and using Hattori J for both, we have:

$$\text{(iii) } t = \frac{\log 55}{1.4 \log 74.4} = \text{1600 years B.P.}$$

$$\text{(iv) } t = \frac{\log 50}{1.4 \log 86.4} = \text{3400 years B.P.}$$

---

the accuracy of the glottochronometry. Just as r is not after all constant in time, so it has now been conclusively demonstrated that C-14 has not been constant 'recently'. The decay-rate is constant, but quantities have varied: this was only to be expected, because of variations in the intensity of incoming radiation. Dendochronology, using the long-lived bristle cone, can give us accurate numbers of calendar years up to about 5000 B.P.

Then finally, take the 'mid-path' figures. Let us say we have 52.5% cognates and use the conventional r (80.5%):

$$\text{(v)} \quad t = \frac{\log 52.5}{1.4 \log 80.5} = 2120 \text{ B.P.}$$

This is the result that would have been obtained by e.g. Blount and Curley, or Heine, using the 200 list, the usually accepted r for that list and the 'old' formula with Hattori's J. Using the same number of cognates, Hattori's J and the so-far-established outer limits of r we get either about 1,200 B.P. or about 17,000 B.P. Granted that 2,100 B.P. is much more 'likely' than either of the other two, the trouble is that we simply do not know what order of r we should be using for e.g. Nilotic/Bantu/Cushitic/Chadic ...

Even these results are of course subject to the quite normal 'standard error'. For a probability of 7/10 (the usual) we have a standard error = the square root of C (1-C)/n, where n is the number of pairs actually compared and C is the percent cognate, as before. For our hypothetical examples, the standard error could well work out as about 600 and 900 years respectively. Thus we could quite reasonably arrive at results with rough centres at 1600 B.P. and 3400 B.P., and rough outer edges at 1000 B.P. and 4300 B.P. A ratio of over 1 to 4, with a certainty of 7 in 10 ..... If we tried to make our 'certainty' 9 in 10 the 'limits' would naturally be far wider, with the centres the same as before. And this without really bringing into consideration the possible lack of applicability of the Swadesh lists to the African scene.* We do not yet know what significance to attach to the discrepancies between diachrony and synchrony for lists for a given language 'group' ('family' or whatever).

One thing is certain - using the 'old' formula (and values) the genetic relationship in time can only be pushed two or three thousand years back anyway. After that, we are deceiving ourselves. What I wish to make clear is the present lack of certainty about even these few thousand years: ratios of 1 to 2 are to be expected, 1 to 3 very probable, and 1 to 4 quite possible, even at a greatest depth of only 3000 years. As we have seen, more extreme cases can give differences of ten or fifteen to one.

'Fine-structure' of r

This was at first regarded as being determined only by various sub-groups in the test lists. These sub-groups are part-grammatical, part-lexical:

---

*Can we be non-cultural? Do we then contradict the entire theory? - maybe some local items need to be ex/included (cf. 'sand' in Austronesian). Properly, mathematics can only be invoked after we know what lists we are working on, because we ask different questions, and get different formulae, as we change the character of our lists. Is there a non-cultural list, and if so, should we be using it anyway?

e.g. pronominals, demonstratives; numerals; body-parts, colours, natural phenomena, verbs of perception, motion, and many more. It has been recognised from the beginning of the study that some cultures have retained or replaced items more or less according to categories.* One might well conjecture that even a given single language would do rather the same. Again, retention or replacement may be affected even by sheer specialised conservatism or lack of it. "Successful' impact on a given language or group of languages by another culture (which could affect more than one category at a time) should be partly detectable at morphophonological level - that is, where replacements are borrowings (not all replacements need necessarily be simple borrowings). "... it would be ideal if several such sets (of largely independent vocabulary items,) could be found, one with high rate (of change) for measuring shallow depths, and one with low rate for ... deep depths" (Gleason, 1959). Gleason also pointed out that "every lexical item at every given time has a certain probability of change (which) is variable ..." But very early on, Lees (1953) and Swadesh (1955) said that the list is not homogeneous: "Through a given time period, as the more resistant classes survive the less resistant, their concentration in the sample will increase ..." and "... it is not to be imagined nor is it suggested by the persistence scores that (the list's) component items are of uniform stability..."

There followed what seems to me a most extraordinary period in research history. Very early in the development of the study, r was held to be an 'average' within 13 long-recorded languages. But (Gleason, 1955 quoted by Hymes, 1960, requoted by van der Merwe, 1966): "Averages cannot be used in an exponential equation-the retention rate risesconstantly in time" (C.A., 1960); and Chretien (1962): "Since averages cannot be used in an exponential equation (both functions) (using r) are invalid, because they are both exponential equationsand use an average rate of retention" (quoted by van der Merwe, 1966). In any case, "If Gleason is right, r is not a constant but a slowly changing variable according to time depth (and hence) cannot be used as a constant by which to determine ... another variable." (Chretien, Language, Vol.38, 1, 1962).

*This aspect of individual 'language-reaction' within a given list is well exemplified by Athenian and Chinese: the former retains none of the five 'colour-words', the latter retains them all. For the same words, "Romance languages show ... retention about one-third of the time, Indo-European generally just about half the time and Athapaskan just three-fourths of the time" (Hymes, 1960, from Swadesh (1955). After further discussion of 'semantic groups' and "important consequences for the adaptation of the test list to particular cases", Hymes states "different but equally valid lists for each of different language families are theoretically possible", and adds an obvious but vital caveat: to the effect that "a completely persistent list would provide no dating at all"; but a list of low persistence would be very useful for recent divergences, and useless for remote times.

Recognition of the non-homogeneity of (any) test lists had come almost at the very beginning, and statements about this and allied difficulties continued to be made for more than a decade thereafter - but noone did anything really effective about it until 1966. And even after that, the argument raged on - worse still, a lot of investigators have continued until today to ignore the delicacy of r, and the advances that had earlier been made precisely in order to deal with that delicacy. To take some of the 'high-lights' only:

Lounsbury (1961) writes of "the invalid assumption of equal average viabilities for all items ..." Because attrition first affects the most vulnerable parts of the list, and leaves an increasingly resistant residue, he says, "the result is a gradual deceleration in the rate of attrition" - that is, r increases as the list becomes more resistant. But curiously enough, he does not then adjust for a changing rate: he simply alters the average rate from 80.5% per millenium to 75.3%, which naturally gives 'better' results at greater depths (around 4 or 5,000 B.P.) but is not so good for about the first 2000 B.P.

Joos (1964) also has: "The current mathematical theory can be adequate only if the numbers... are equally likely to vanish in time. But ... their retention rates are probably spread out ..." He then gives an eight-group skewed distribution pattern but makes no further step. What is clearly needed is a more sensitive type of expression; this is provided by the normal decay formula used in all natural situations of this kind. It takes the shape of a summation over all the different groups involved in the decay-process; or, in glottochronometry, the rate of decay of any list is proportional to the sum of the decay rates of all its items (which are themselves subject to the aforesaid decelerations through time). Van der Merwe (1966) introduced such an adjusted formula (see below, p.38), which neatly deals with all the more immediate objections and difficulties.

Jacobsen (1966) repeats "(van der Merwe) has recognised that (1) the ... words ... on the ... lists have ... different retention; (2) r of a given list increases with time-depth and makes it ... clear that these concepts ... were previously stated ... by Lounsbury, Gleason, Chretien, Joos, and Dyen" (and, as I said, by Swadesh and Lees as far back as 1952 and 1953).

Milke (1966) makes a more general point: "...different languages exhibit widely differing retention rates, even if the same test list is used and time-depths ... are comparable." (my underlining). Cf. Gudschinsky (1966), who perhaps rather tentatively queries: "Is it not more reasonable to suppose that every word in every language has its own retention rate?" Naturally - what is surprising is that it took so long to see it.

Trager (1966) quite rightly re-states (as others before him) that vocabularies are determined by morphemics (cf. the tri-consonantal-root self-reinforcement in Semitic as opposed to the phonologically restricted list of monosyllables of Chinese, by syntax, by semology, and most importantly, by all the rest of the culture and its history. But I cannot see this in any way invalidates "the idea that the process of morpheme decay can be described mathematically" (van der Merwe). Trager (also correctly) says that van der Merwe's conclusions mean that "every word of any list has its own viability" - a much more crucial matter. Trager thought this made nonsense of the whole study, but ...

By using a van der Merwe-like formula* we can make the number of groups what we like - the really important point is that by

---

*<u>Van der Merwe's adjusted formula</u> (typing T for 'tau')

$$y = \frac{1}{x} \sum_{i=1}^{i=x} \exp -t/Ti$$

Where y = C = common retentions

x = number of subdivisions of list

T_i = rates of retention of subdivisions (cf. also Whitman, <u>ibid</u>.)

This formula, re-expressed (using Swadesh's symbols) by Jacobsen (1966) is:

$$C = \frac{1}{x} \sum_{i=1}^{i=x} r_i^t$$

I am greatly indebted to W. Driedger (personal communication, 1972) for the following:

$$p = \frac{1}{n} \sum_{i=1}^{m} n_i \exp -t/Ti$$

where p = proportion of cognates remaining at

t = time elapsed since P was 100%

T = time constant = $\frac{1}{\log_e r}$

n_1, n_2 = number in group 1, group 2 ...

N = total of $n_1 + n_2 + \ldots$

m = total number of groups

Note: Here <u>each item</u> may be regarded as a 'group', each with its own <u>r</u>.

continuing to use the 'old' formula with its implication of homogeneity, we are certainly distorting results, more especially at the greater depths. The difficulty is that the use of the more recent formulae inevitably requires computer-examination of date - you simply cannot do the whole job 'by hand'. However, we may still nevertheless agree with Swadesh (1966) that "one can go on using the present approximations", as long as we are fully aware that they are only approximations: many users and 'believers' are not so aware. (Even in the more modern approaches, we are using 'approximations', but now in a deliberate, conscious (mathematical) sense; that, is we use mathematical 'approximations' within the process, so that the end results may become much less approximate in the ordinary sense.)

"(Swadesh, 1952; Lees, 1953, and Dyen, 1962, 1965) make the assumption that the retention ... rate ... is the same for each item. This ... is badly in need of modification ... to produce a realistic model. We now permit different retention rates for ... different meanings. Cognation between words ... of a lower r scores more strongly for closeness of relationship than cognation between words ... of a higher r" (Dyen, James and Cole, 1967). And, using 81% r per millennium, they derive a "time unit of t and r" of 1069 years, "i.e. our time units are roughly millenia" (ibid). This means that Lees' 'average' for the 13 original (historically documented) languages turns out to fit remarkably well with the findings resulting from their mathematically much more sophisticated lexico-statistical treatment of 89 Austronesian languages. It does not follow that we can therefore assume this same value for r is generally acceptable. But the techniques now deployed (from van der Merwe and the others onwards), together with the immense possibilities opened up by computerized comparisons, do offer far greater hope for obtaining better datings. The establishment of an accurate r (the form tau, 'time-unit' in the later formulae is still essentially an r, even though applied to separate items instead of to an entire list), remains at the heart of the matter.

Summary

Whether we use the simple (and inaccurate)'hand' formula t = log C/J log r, or a more accurate formula, item-weighting and a computer;

(1) C depends on:

(a) the list used is the list as suitable as possible for the culture to which it is being applied;

---

*Please note that these figures mean that figures elicited by using the 200 list and the 'conventional' average of 80.5 (or 81)% retention could be either eight and a half times too deep, or nearly twice too shallow. Thus a calculated date of say 200 B.P could be anywhere between 230 B.P. and 3700 B.P. (200 B.P. would simply be somewhere between 23 B.P. and 370 B.P.)

(Footnote to next page...)

(b) the accuracy of recovery of true cognates: we must know all about the possible sound-shifts and governing laws, and not judge merely by 'resemblances'. (Much contemporary work on non-Bantu Africa has necessarily been of the latter type.)

So, the first question of all is still "what items should be on the list?" Then, how we define the criteria that decide between cognate and non-cognate?

(2) r has been shown to vary between very wide limits, unless as an (illegal) 'average' for an assumed (but impossibly) homogeneous list. The variation in results for t may easily be two to one, and five or ten to one are not at all odd: even 16+ to 1 is possible.* The conventional values - 80.5 (200 list), 86.0 (100 list) are based on 13 long-recorded languages only, of reliable time-depth 2,200 years only. In any event, the only 'proper' way of treating the list is by applying a separate r, varying separately in time for each item, to each item (which takes us back to the construction of the lists from which we derive C, in its turn affected by and affecting J).

(3) J in the older formula is a variable, probably mostly dependent on the interaction or infra-action of the culture(s)/languages concerned: it has to take into consideration dregs, drift and contact. Maintenance or resumption of contact after a split will slow divergence down. Drift tends to keep change going along the same lines, if the "balance of internal forces for changes" was more or less the same to begin with. But where the test list approximates the proto-language, drift and contact can be pretty well neglected. The 'conventional' lists do not approximate proto-Bantu, and probably not other African language groups either. Dregs are those items that are more persistent than others, the proportion of which will increase with time B.P. If for lack of personnel, time, money and access to computers, we still have to use the 'hand' formula, what do we put for J?

True, J and r can be taken care of completely if we can establish r_i for every item in the list, but do we in fact need this or will sub-groups be adequate? In any case, and most important of all, what do we put for r (or T)? Just supposing we do treat our list as homogeneous, how great a distortion in results can we expect (or tolerate)?

*See previous page for footnote

References

Arndt 1955 Unpublished Ph.D. dissertation, University of N.Carolina.

1959 Language, 35.

Bergsland & Vogt	1962	Current Anthropology, 3.
Blount & Curley	1970	Journal of African Languages, 1.
Chretien	1962	Language, 38,1.
Driedger	1972	Personal communication.
Dyen	1962	Language, 38, 1:3.
	1965	International Journal of American Linguistics, Memoir 19
	1966	Current Anthropology, 4
Dyen, James & Cole	1967	Language, 43, 1.
Eaton	1966	quoted by van der Merwe, Current Anthropology, 4.
Gleason	1955	Work book in Descriptive Linguistics.
Gudschinsky	1955	International Journal of American Linguistics, 22.
	1966	Word, 12.
Guthrie	1971	"Comparative Bantu", Vols. 1-4.
Hattori	1953	Gengo Kenkyu 22, 23.
Haudricourt	1966	Current Anthropology, 4
Heine	1971	Mila, II, 2
Hoijer	1953	
	1956	Language, 32.
Hymes	1960	Current Anthropology, 1.
	1964	Current Anthropology, 4.
	1966	Current Anthropology, 4.
Izui	1962	Current Anthropology, 2.
Jacobsen	1966	Current Anthropology, 4
Joos	1964	Proceedings of 9th International Congress of Linguistics.

Kroeber	1955	International Journal of America Linguistics, 21.
Lees	1953	Language, 29.
Lorge	1966	quoted by van der Merwe, Current Anthropology, 4.
Meeussen	1969	"Bantu Lexical Reconstructions"
van der Merwe	1966	Current Anthropology, 4.
Milke	1966	Current Anthropology, 4.
Rinsland	1966	quoted by van der Merwe, Current Anthropology, 4.
Swadesh	1950	International Journal of American Linguistics, 16
	1952	Proceedings of American Philosophical Society.
	1953	International Journal of American Linguistics, 19.
	1955	International Journal of American Linguistics, 21.
	1966	Current Anthropology, 4.
Savage	1966	quoted by Dyen in "Comment". Current Anthropology, 4.
Teeter	1963	Language, 39.
Thorndike	1966	quoted by van der Merwe, Current Anthropology, 4.
Trager	1962	Current Anthropology, 2.
Whitman	1966	Current Anthropology, 4.