Afroasiatic Comparative Lexica: Implications for Long (and Medium) Range Language Comparison

Robert R. Ratcliffe

1. Comparing Comparative Lexica

The question of the possibility of identifying and reconstructing language families at historical time depths deeper than those assumed for established families like Indo-European have been hotly debated in recent years, with some arguing that the comparative method simply does not apply beyond a certain point (Afroasiatic often cited as the limiting case) others that there is in principle no reason why the method should have a cut off point, and that deeper relationships can be, and perhaps have been identified.

One way to test the reliability of the comparative method would be to undertake the following experiment. Take a set of languages for which a relationship has been suggested, but for which regular sound correspondences and a reconstructed phonemic system of the proto-language have not yet been established. Furnish two libraries on opposite sides of the world with all the available and relevant information on the languages (dictionaries, grammars, texts). Take two researchers trained in the comparative method, put them in the libraries, keep them in isolation from each other and see what they come up with. If it is a reliable procedure then two trained practicioners of it confronted with the same body of data should come up with broadly similar results-- repeatability of experiments should be expected as in natural science. Unfortunately, as so often in linguistics, ethical considerations prevent us from subjecting real human beings to such an experiment.

The world of comparative linguistics is fortunate, therefore, that something very close to this experiment came to be performed by accident. The year 1995 saw the publication of two Afroasiatic comparative etymological dictionaries, produced in relative isolation from each other: Ehret (1995) henceforth E, and Orel/Stolbova (1995) henceforth O/S. Both works were produced by trained linguists, with knowledge of several of the relevant languages, making a good-faith effort to apply the comparative method. Both works hold strictly to the principle of regular sound correspondence in identifying cognates. And in both cases the reconstructed sound system which the correspondences go back to are consistent with the uniformitarian principle: both the size of the inventory and the types of contrast reconstructed are consistent with what is found in living languages. Moreover, since each team worked (largely) in ignorance of the other's work, a comparison of the two offers a rare opportunity to test the reliability of the comparative method.

In order to apply such a test I undertook a systematic comparison of E with O/S, grouping the proposed cognate sets into the following four categories: agreeing, contradictory, complimentary, and incompatible in principle. Agreeing describes the case where both sources agree in connecting a particular word in one branch with a particular word in at least one other. Contradictory is the case where both sources treat the same word in one branch but connect it with different words in another branch. Complimentary describes both the case where connections are proposed between words not treated by the other source (yet not incompatible with it) and the case where both sources take up a word in one branch but connect it with proposed cognates in different branches. Incompatible in principle describes the case where one source proposes connections between words not treated by the other source, but where the validity of these connections depends upon proposed sound correspondences or assumptions about grammatical structure or semantics not admitted by the other source.

agreeing

branch	A	B	C	D
OS	x		z	
E	x		z	

contradictory

branch	A	B	C	D
OS	x		z	
E	x		b	

complimentary (case 2)

branch	A	B	C	D
OS	x		z	
E	x	y		w

Note that it is theoretically possible for a set to be simultaneously agreeing and contradictory.

agreeing and contradictory

branch	A	B	C	D
OS	x	y	z	
E	x	w	z	

These cases are few and I have counted them as agreeing.

A complication in applying this method is that the sources adopt slightly different, though not incompatible, subclassifications. E maintains a standard six branch division (Semitic, Egyptian, Cushitic, Omotic, Chadic, and Berber), but he ignores Berber. O/S question, but do not necessarily reject, the reality of a Cushitic sub-branch, and present 'Cushitic' data under the headings of six well established sub-families: Beja, Agaw, "East Cushitic," Dahalo, Mogogodo, and Rift. A further difficulty is that E often gives reconstructed forms only without attestation of actual language data.

It is important to emphasize that these categories refer to proposed cognate sets, not the reconstructed etymons which those sets reflect. There are often considerable differences in reconstruction even when the cognate sets are agreeing. For example, one well attested cognate agreed by both sources is a verb "to die"— Arabic reflex *maata* (perfect) *yamuutu* (imperfect). The proposed reconstructions are E 600 maaw, O/S 1751 mawut.

1.1 Incompatible in Principle

The categories of agreeing, contradictory, and complimentary require no further explanation. But the category of incompatible in principle (henceforth ipi) depends upon a judgment on a part of the evaluator (me), hence the reasons for making such a judgment must be made explicit. There are three different causes of incompatibility between the sources: incompatible sound correspondences, incompatible assumptions about root structure, and incompatible assumptions about semantics. The category accounting for the largest number of ipi sets is incompatible sound correspondences. (Note, however, that most contradictory sets are made possible by different assumptions about root structure or semantics.)

1.1.1.Incompatible Sound Correspondences

If a cognate set depends upon a sound correspondence contradictory with the sound correspondences proposed in the other source it is ipi. The consonant systems reconstructed by the two sources are superficially fairly similar.[1] But the correspondence sets which the reconstructions reflect are sometimes contradictory. A system of 42 consonants is reconstructed by E and one of 33 by O/S. The 12 consonants reconstructed

[1] This is not to say that all points of agreeing reconstruction are non-controversial. From a Semiticist point of view the reconstruction of affricates, as precursors of the Semitic interdental fricatives, and of p and f as distinctive segments raises eyebrows.

by E but not by O/S are three nasals (palatal, velar, and labio-velar ɲ , ŋ , ŋʷ), three sibilants (z, s, and ʃ), a labial ejective (p'), and a labio-velar series of 3 stops (kʷ , gʷ , k'ʷ) and two fricatives (xʷ , ɣʷ). The three consonants reconstructed by O/S but not by E are an ejective dental affricate (ts') and two uvular stops, unvoiced and ejective (q, q'). This leaves a set of 30 identical or nearly identical reconstructed consonants. For 12 of these there is more or less complete agreement in the proposed correspondence sets:

b	t d			ʔ
			ħ ʕ	h
m				
	r	l		
w		j		

If some leeway is allowed the total number of more-or-less agreeing sets can be augmented as follows. On both analyses a p/f distinction is reconstructed on the basis of Egyptian and Chadic with merger assumed in Semitic and (for OS) Berber. OS assume a set of mergers and splits in Cushitic and Omotic which perhaps prevent complete agreement [2 more]. The correspondence sets for the velar stops are also the same (essentially no sound change assumed for any branch) except for E's labio-velars [3 more]. The dental nasal /n/ corresponds to /n/ in all branches in both approaches. E's 5 nasals are based on Cushitic and, supported somewhat by Chadic. The palatal, velar, and labio-velar nasals are assumed to have merged with /n/ in Egyptian and Semitic [1 more]. E's three labio-velar stops and three back nasals are not incompatible with OS's reconstruction. One can adopt OS's reconstruction and assume that the extra segments reconstructed by E are later splits. This gives a compatible set of 18 reconstructed segments(corresponding with 24 segments in E).

p b	t d		k g k'		ʔ
			[kʷ gʷ k'ʷ]		
f				ħ ʕ	h
m	n	[ɲ	ŋ ŋʷ]		
	r	l			
w		j			

That still leaves 18 proposed correspondence sets (= proposed reconstructions) in E and 15 in OS which are contradictory.

E: p', s, ɳ, tʃ, z, dz, tʃ, d□, t', s', tʃ', tl', ɬ , dl, ɣ , ɣ ʷ , x, xʷ

OS: s, ts, dz, tʃ, d□, t', ts', tʃ', ɬ , tɬ , tɬ ', R, χ, q, q'

Entries reconstructed as beginning with these segments account for a total number of 362 entries in E or roughly 36% of the total: p' (17), s (32),ʃ (24),ts (18), z (27), dz (8), tʃ (6), d□ (19), t' (23), s' (18), tʃ' (22), tl' (33), ɬ (31), dl (26), ɣ (24), ɣ ʷ (9), x (16), xʷ (9)=362/1011= 36% of total.

The problematic segments in OS account for 670 entries or 25% of the total: s (168), ts (41), dz (45), tʃ (5), d□ (31), t' (48), ts' (49), tʃ' (21), ɬ (29), tɬ (61), tɬ ' (12), R (24), χ (92), q (23), q'(21)=670 /2672 = 25% of total

Note that besides /s/ and /x/ the average number of supporting entries for each correspondence in OS is below average. (avg. 81)

Entries beginning with these segments have been a priori excluded as ipi and account for the largest number of ipi entries (362/619=58%). The second largest number of ipi entries consists of those entries which have one of these consonants elsewhere in the root (126 entries or 20%).

1.1.2 Incompatible assumptions about root structure.

Both E and O/S allow two-to-three matches, that is cases where a word with two consonants in one language is proposed as cognate with a word with three consonants in another. However, the justification for this is different in each case. O/S allow for the loss of certain consonants, notably sonorants and gutturals, in all environments, and allow for a small number of consonants to be treated as fossilized prefixes. E takes the more radical view that all third consonants are fossilized suffixes. This allows him further to recognize two-out-of-three matches, cases where the first two consonants in a three consonant word correspond with the first two consonants in a three-consonant word in another language, although the third consonants do not correspond.

permitted by E and OS (3-to3 and 2-to-3)

ABC	ABC
ABC	AB

permitted by OS, but not by E (2-to-3)

ABC	ABC
BC	AC

permitted by E but not by OS (2-out-of-3)
ABC
ABD

I have regarded proposed correspondences which rely on the second two types of matches as ipi. This accounts for 115 cases or 19%

1.1.3 Ipi or contradictory because of semantic assumptions

This includes two types of problems, first for a word with many meanings or many translation equivalents, the sources may choose to take different ones as basic and search other languages accordingly. For example, for the Arabic root b-k-r, O/S take the Arabic bakr "young camel" and connect it with a Berber form meaning "lamb" or "kid." E cites Arabic bakar "morning" and connects it with an Egyptian "to be bright" Cushitic "kindle" Chadic "roast," Omotic "star" (E 11 bâk "shine" O/S196 bakVr "young animal"). Ignoring the semantics these entries are complimentary. But I have made a judgment that in this case the semantics are too far apart for one source simply to incorporate the other's proposal. Hence I judge these proposals ipi.

The second problem, not unconnected with the first, is that the authors have different assumptions about what sort of vocabulary to reconstruct for the Proto-Language. As this example also illustrates, E has tendency to reconstruct verbs wherever possible, while OS prefer concrete nouns.

I have used this category sparingly in deciding ipi, only judging 15 examples (2.5%) as ipi on this criterion. But it is clear that many of the contradictory sets also result from different semantic assumptions, meaning that the category could be and probably should be more widely applied.

2. Measuring the discrepancy

If the comparative method is reliable at the depth of Afroasiatic, and if the practitioners have applied it accurately, then the expectation is that the bulk of the entries should be agreeing or complimentary. From this point of view it is surprising and indeed disturbing that the two dictionaries are radically different. Specifically I was able to find only 59 agreeing sets, as against 158 contradictory sets. Of E's 1011 entries, 619 are incompatible in principle with O/S,while only 175 are complimentary. (There remain another 1000 entries in O/S which are either complimentary or ipi with E, but which I have not yet evaluated).

Thus less than six percent of the cognate sets proposed by E are also proposed by O/S. Slightly more than another 17% are complimentary. From the reverse perspective, since O/S have roughly two and a half times as many entries as E slightly more than 2 percent of the cognate sets proposed by O/S are found in E. Die-hard opponents of long-distance comparison may gleefully leap to the conclusion that the method is 94 to 98% inaccurate even at medium depths, but such a conclusion would be premature. Still the fact remains that two sets of scholars have been able to reconstruct mutually unrecognizable proto-languages, and this demands an explanation. If O/S's assumptions about sound correspondences, semantic correspondences, and root structure are correct, then E has hundreds of spurious cognates. Contrariwise if E's assumptions are correct, then O/S must be full of spurious cognates. Of course both sets of assumptions could be flawed to some degree, in which case both sources would be full of spurious cognates.

3. Margin of Error

A traditional, and by no means incorrect way to try to reconcile this discrepancy, would be to evaluate the reconstructive assumptions in each source, or even each proposed cognate set, on substantive grounds. In the present state of Afroasiatic studies, however, there are simply too many unknowns, too much room for subjectivity, for such an approach to deliver a decisive verdict. Of course since no one has direct access to proto-Afroasiatic, there is no way to determine absolutely whether a given proposed cognate set is spurious or not. Moreover this approach misses the larger point. If it is possible for trained and knowledgeable scholars to produce hundreds of plausible, but necessarily spurious cognates (in one case or the other), what does that tell us about the pitfalls of the comparative method at deep time-depths.

It is clear that the comparative method has a certain margin of error, that it will inevitably turn up a certain number of accidental matches, which do not reflect shared history but are a result of chance. (Depending upon the criteria adopted by a given researcher, a match may be defined as a pair of semantically equivalent words starting with the same segment, or starting with segments hypothesized to correspond, or containing two or three or more hypothetically corresponding segments in the same position, etc., with a different margin of error in each case.) What may be less clear is that this margin of error is a direct consequence of how a researcher goes about making a comparison, and that it can be characterized mathematically. If we know everything about how a researcher has conducted a comparison, it is possible to say that in that case on average x number of chance matches are expected.

The problem can be formulated in terms of probability theory. The average expected number of chance occurrences of an event is the probability of the event (P) times the number of trials (T). (The probability of a flipped coin coming up heads is 50% or .5. If the coin is flipped 100 times the expected number of chance occurrences of heads is 100 x .5= 50.) The probability of a match, for example the probability that semantically equivalent words in different languages will start with the same consonant or with consonants hypothesized to correspond, is a function of the relative frequency of the segments compared (Ringe 1992). The number of trials in a comparative study is the number of comparisons made. In conducting a comparison, there are various assumptions and procedures which a researcher may adopt, which are not inherently illegitimate, but which nonetheless lead to increasing the probability of a match or to increasing the effective number of comparisons made, and hence to increasing the margin of error. What I want to do now is to consider what these are and how they may have affected the results of the two Afroasiatic comparative lexica.

3.1 General Calculations

3.1.1 Probability of a Match

To begin with the probability of finding any segment in any given position in a word is equivalent to the frequency of the segment in that position in that language. If half the words in a given language start with “d” then the probability that any given word, say the word for “dog” will start with “d” is 50%. If 5% of the words in a given language start with “d” then the probability that the word for “dog”or any other given word will start with “d” is 5%, or 1/20. Under the formula of PT, therefore, if you take 20 words in English starting with “d” and compare them with exactly one semantic equivalent in this hypothetical language you would expect to find on average that one of these twenty will also start with “d”. The number of words compared is the number of trials. If a hundred English words starting with “d” are compared, then five words starting with “d”in the given hypothetical language can be expected.

Note that the relevant notion of frequency depends on type counts rather than token counts (i.e. dictionary entries rather than running text). But without actually calculating the frequency of each consonant (or vowel), we can easily calculate the average frequency of all consonants (or vowels) in a language by simply dividing 1 by the number of consonants (or vowels respectively) in the consonant (or vowel) inventory of that language. In making the calculations which follow, I will use average frequency. (In order to get some idea of how much distortion this introduces, I calculated the distribution of Arabic consonants in three consonant words in Wehr's dictionary-- i.e. a token count. I found a range of from 90% below to 200% above the average. In other words the most frequent consonant occurs about twice as often as the average, the least frequent one -tenth as often.)

If a cognate is defined as a word containing two or more correspondences, and if both languages contain 20 consonants, each having the same frequency of occurrence in all positions, and correspondences are assumed to be exactly one-to-one between the two languages, then the chance of a match on C1 (on average for any of the 20 C's) is .05 (1/20) and the probability of a match on C1 and C2 is .05x.05=.0025 (1/400), and the probability of a match on C1,C2, and C3 is .000125 (1/8000).

20 C's

C	CC	CCC
.05 (1/20)	.0025 (1/400)	.000125 (1/8000)

If the consonant inventories are larger the probability of a chance match is smaller.

For 25 consonants the probabilities are

C	CC	CCC
.04 (1/25)	.0016 (1/625)	.000064 (1/15,625)

For 30 consonants the probabilities are

C	CC	CCC
.0333 (1/30)	.00111(1/900)	.0000370 (1/27,000)

For 40 consonants the probabilities are

C	CC	CCC
.025 (1/40)	.000625 (1/1600)	.000015625 (1/64,000)

If vowel correspondences are also insisted upon the probability of a chance match decreases further. For a five vowel system the chance of a match on the first vowel is (on average) 1/5 or .2 and for a match on two vowels is 1/25 or .04. For a three vowel system the probabilities are 1/3 (.33) and 1/9 (.11). The possibilities for CVC, CVCV, CVCC, and CVCVC matches are as follows

20C+3V			
CVC	CVCV	CVCC	CVCVC
.000833	.000278	.0000417	.0000139
(1/1,200)	(1/3,600)	(1/24,000)	(1/72,000)
20C+5V			
.0005	.0001	.000025	.000005
(1/2,000)	(1/10,000)	(1/40,000)	(1/200,000)
25C+3V			
.000533	.000178	.0000457	.00000711
(1/1,875)	(1/5,625)	(1/21,875)	(1/140,625)
25C+5V			
.00032	.000064	.0000128	.00000256
(1/3,125)	(1/15,625)	(1/78,125)	(1/390,625)
30C+3C			
.00037	.000123	.0000123	.00000412
(1/2,700)	(1/8,100)	(1/81,000)	(1/243,000)
30C+5V			
.000222	.0000444	.00000741	.00000148
(1/4,500)	(1/22,500)	(1/135,000)	(1/675,000)
40+3			
.000208	.0000694	.00000521	.00000174
(1/4,800)	(1/14,400)	(1/192,000)	(1/576,000)
40+5			
.000125	.000025	.000003125	.000000625
(1/8,000)	(1/40,000)	(1/320,000)	(1/1,600,000)

A number of methodological points emerge from these figures.

First, consonants are more important than vowels. That is, insofar as consonant inventories are larger than vowel inventories in a language, which seems to be universally the case, the probability of a chance a match on consonants is lower than that on vowels. It is thus safer to ignore vowels than consonants in the initial analysis. The cost (in terms of increased probability of a chance match) of ignoring one consonant out of three in three consonant words is higher than the cost of ignoring the vowels. Ignoring vowels increases the probability of a chance match usually by an order of magnitude or less, but ignoring the third consonants when three are present increases the probability of a chance match by two orders of magnitude. This is important to note in this context, because both E and to a lesser extend OS reconstruct to a CVC target, rather than to a C-C-C target.

Second, insofar as distribution of consonants is more or less even through the languages compared, chance matches are less likely between languages having large consonant inventories than for those with small ones. The threshold of statistical significance is reached with fewer examples in these cases.

Third, where distribution is uneven, matches between statistically infrequent consonants is less likely than for frequent ones.

3.1.2 Increasing the probability of a (chance) match

There are a number of steps a compaartist might take, which while not inherent–ly illegitimate, have the effect of increasing the probability of a chance match, and allowing more error into the results. The two most obvious of these are allowing many-to-one correspondences and systematically ignoring some feature of phonological structure.

3.1.2.1 Many to one and many to many correspondences

Since mergers and splits do occur, it is not inherently illegitimate to allow many-to-one or many-to-many correspondences, but this increases the probability of a chance match. For example, under the scenario above of a language with twenty consonants the chance of finding a match on one of them is 1/20 (05). But if this consonant inventory is distributed through four places and five manners of articulation, and the researcher allows a consonant in one language to match with any consonant at the same place of articulation in the other, then the chance of a match is 1/4 (.25). Here the possibility of a match on both one and two is 1/16 (.0125) and of 1, 2 and 3 is 1/64 (.0156). If we were to take 20 words in English starting with “d” and compare them with exactly one semantic equivalent in this hypothetical language we would expect to find on average that five (20 x .25) of these twenty would start with a dental of some kind.

3.1.2.2 Ignoring features of structure

Another way to increase the probability of a chance match is to ignore certain features of phonological structure. This too can be perfectly legitimate. For example, because developments affecting stress or tone can be complicated a researcher may choose to ignore them in an initial stage of comparison. In fact ignoring such features has a low cost in terms of increased probability of a chance match because such features generally have few values, for stress either stressed or unstressed, for register tone, usually either high or low. We have already noted, that for the same reason, ignoring vowels is less costly than ignoring consonants.

Another way to increase the probability of a chance match is to treat as cognate word pairs in which only two out of three consonants corresponds. (We have already noted that both O/S and E do this to some extent and in different ways). Since loss of segments is not at all unusual, and fusion of an affix with a stem is possible, there is nothing inherently wrong with allowing for this possibility, but it increases the probability of a match, and hence the margin of error, considerably. The cases to be considered here are the costs of allowing 2 out of 3 (ABC = ABD) and one type of 2 to 3 (AB = ABC), as in E, and the cost of allowing several types of 2 to 3 matches (AB=ABC, AC= ABC, BC=ABC) as in O/S.

The chance of finding a two out of three match for the first two consonants (ABC=ABD) is no different from the chance of finding a two to three match (AB=ABD). In either case C3 is ignored so the chance of finding a match on C3 is 100%. For 25 consonants, the average probability is .04x.04x1.0 = .0016. This is of course also the same as the probability of finding a two-to-two match (AB=AB). The real problem with 2 out of 3 matching is that it increases the number of trials-- the number of potential comparisons. I will elaborate on this in the following section.

If multiple types of two to three matching is allowed, as in OS, then C1 is allowed to match either C1 or C2, and C3 is allowed to match C2 or C3. The chance that C1 will match either C1 or C2 is the frequency of C1 plus the frequency of C2. In the above case with 25 consonants evenly distributed this would be

.04 + 04 = .08 (or 1/12.5)

Multiplying this by the probability that the final C will also match gives the probability of finding either an AB=ABC or BC=ABC match. Again in the case of 25 consonants this would be

08 x .04 = .0032 (1/312.5)

If C2 is also allowed to match either C2 or C3 the probability of a match on one consonant is again, in the hypothetical case, .08.and the probability of a chance match of either AB=ABC, BC=ABC or AC=ABC is

.08 x .08 = .0064 (1/156.25)

The point here is that all types of two to three or two out of three comparison allows a probability of a chance match two orders of magnitude higher than that for a three out of three match. Thus under the formula of PT, if the same number of comparisons are made, at least a hundred times more spurious cognates are expected under the assumption of 2 to 3 or 2 out of 3 match than under the assumption of strict 3 to 3 matches.

3.1.2 Increasing the number of trials.

The number of trials in a comparative study is the number of comparisons made. If we accept only one-to-one comparison, each word in language A compared only with its exact equivalent (assuming it exists) in language B, then for two 1000-word word lists or dictionaries containing the same entries there will be 1000 trials. Assuming all the words in the list have three consonants and we only accept as cognate those words in which all three consonants correspond, the probability of a chance match, in the example case of 25 consonants equally distributed, is .000064, and the average expected number of chance "cognates" is .000064 x 1000= .064, i.e. less than one. Even a single cognate in this case is statistically significant and unlikely to be due to chance. If however, we allow the third consonant to be ignored, as in E, then the probability of a chance match is .0016, the expected number of chance "cognates" is 1.6, i.e. around 2. Even here the threshold of statistical significance begins to be crossed with as few as four or five cognates.

The expected number of chance matches increases dramatically, however, if semantic leeway is allowed into the comparison. Again since semantic shifts do occur, it is not inherently illegitimate to allow for them. But if, for example, one compares each word in the hypothetical 1000 word list above not with its exact semantic equivalent only but with a range of 10 semantically close terms in the other word list, the number of comparisons then increases to 10,000, with the expected number of chance "cognates" also increasing ten-fold, in the case above .64 3-to-3 matches, 16 2-to-3 matches. If the semantic range is 100 the number of expected chance "cognates" again increases 10-fold, with 6.4 2-to-3 and 160 3-to-3 matches expected in the example case. (Note that under the binomial distribution as the average number of expected chance cognates increases, the number of cognates needed to cross the threshold of statistical significance also increases, but at a higher rate.)

How much semantic range do compartists normally allow for, and how much have E and OS allowed for? Researchers rarely spell out how they have gone about searching for cognates, how many comparisons they have made before they have found anything. But often this can be deduced from the results. One interesting cause and effect relation, is that broad semantic leeway in making the comparison leads to the reconstruction of forms with vague semantics and also not unusually to the reconstruction of a large numbers of synonyms.

This can be illustrated by example. Suppose one compares the word for "bird" in one language not only with the equivalent for "bird" in the other language, but with the words, for sparrow, pigeon, vulture, etc. If there are 50 "bird-words" in each language, and one compares each of these in one language with the fifty in the other, there is a total of 2500 comparisons made. Of course if one finds apparent cognates on the basis of assuming semantic equivalence between pairs like swallow-hawk, parrot-quail, kite-ostrich, butterfly-pelican, etc. one can only reconstruct a superordinate term like "bird" or "kind of bird." If several such matches are found, the researcher ends up reconstructing several synonyms for "bird."

This example may sound exaggerated, but in fact it seems to be exactly what O/S have done in this particular case. The index under "bird" indicates 52 sets reconstructed with the meaning "bird". This includes pairs like the following.

10 Sem., Eg. "kind of bird," ECh "duck"
320 Eg. "duck," ECh "hen"
356 Eg. "falcon," CCh "vulture," "hen," ECh "great bustard," Agaw "kind of bird"
397 Sem. "swallow," Rift "hawk"
432 Sem. "sparrow," Ch. "guinea fowl"
443 Eg. "kite" Ch. "parakeet"
714 Ch. "guinea fowl," Rift "stork"
736 Sem. "thrush" Ch. "kite"
748 Sem. "parrot," WCh. "quail"

1261 Sem. “kite,” ECh. “ostrich”
1505 Sem. “crane,” CCh. “dove, francolin”
1539 Eg. “cuckoo,” WCh “rooster,” CCh “hen”
1598 CCh. “hawk” ECh. “dove”
2072 Eg. “goose” HEC “crow”
2090 Sem. “crane,” ECh “vulture”
2190 Berber “butterfly, small bird” Ch. “guinea fowl,” Beja “pelican”

It is quite likely that there are several valid cognate sets among these 52. Certainly the material provides an interesting starting point for a future monograph on Afroasiatic birds (along the lines of Indo European trees). However, it is not plausible that Proto-Afroasiatic had 52 synonyms for the generic term “bird” and that these later differentiated in the individual languages. Natural languages do not allow such a high degree of synonymy. The reconstruction of synonyms is an artifact of method-- specifically of allowing broad semantic leeway in making comparisons.

This kind of semantic leeway is a consistent feature of the O/S study. I found that on average O/S reconstruct 10 items for each word in the 100 word Swadesh list, the highest number being the 52 for “bird.” In trying to determine how many comparisons have been made, the highest number found rather than the average should be taken as a guide. If 50 matches were found, at least 50 items were looked at (although admittedly not necessarily in any one language). If 10 items were found, more than 10 must have been looked at. There are roughly 1600 entries in O/S’s appendix, and 2672 items reconstructed, many of them synonymous. We can assume, conservatively I believe, that these figures reflect word lists of about 2000 words. If 2000 words are compared with each word in one language tested against 50 semantically similar ones in the other, then the total number of trials or comparisons is 100,000. If all the comparisons were 2 out of three, and the number of consonants in each language compared is 25, then based on the probabilities figured out above, 640 chance 2 of 3 matches are expected, and if all matches are 3 of 3 then 6.4 chance matches are expected.

These calculations are based on the assumption that comparison involves two languages, or two branches. Another way to increase the number of comparisons, and hence the average number of chance matches is to directly compare languages of different branches without attempting a reconstruction to the branch level. This is equivalent to reconstructing a variety of synonyms for the proto language of the branch. Take the example of words for “hen” in Chadic. The OS entries under “bird” include 25 words for “hen” in 21 Chadic languages (Kwang, Kera-2, Gisiga , Sumray, Mofu, Gude, Munjuk, Musgum, Sibine, Bolewa-2, Dera, Tangale-2, Pero, Ngamo, Gulfey, Sura, Angas, Montol-

2, Nanchere, Kabalay, Mokilko) with four languages apparently having two synonyms (although the two Montol words analyzed as going back to different prot-Afroasiatic etyma look suspiciously like mere transcriptional variants by different field workers). These are grouped into 13 cognate sets, reflecting 5 words for “hen” reconstructed for proto-East-Chadic, 5 reconstructed for proto-Central-Chadic, and 4 reconstructed for proto-West-Chadic. In two cases, the Central Chadic and East Chadic sets are grouped together as cognates, but without any suggested proto-Chadic reconstruction. If this example is typical it means the number of comparisons, and hence the number of expected chance correspondences, between, any non-Chadic language or proto-branch language should be multiplied by 13.

Chadic “hen” (figures after entry indicate: No. of branches outside of Chadic.-Type of match- No.of Chadic words)

301 ECh *bwagur Kwang bogor-to, Kera d\b\rg\ w/ Eg. bd™3 “duck” 1-3/3-2

356 CCh *bwak Gisiga bokoy w/ ECh ʔ abuka great bustard, Eg. byk “falcon”, Agaw bik ”kind of bird” 2-2/2&2/3-1-

748 ECh *dur Sumray dure: w/ WCh durwa “quail” Sem. durr “parrot” 1-2/3-1-

943 CCh *gwa-gwar Mofu gwagwar,
ECh *gu-gur Kera gu-gur w/LEC Omotic gogorrii “guinea fowl” 1-3/3-2-

965 CCh *gVya Gude gyagya w/ Eg.(sarc.) d-wy.t “kind of bird” 1-2/3-1-

988 CCh *yVgur Munjuk yugur Musgum yugur, igur 1-2/3-3-
ECh *gurVy Sibine g\ray w/ Eg. gry poultry

1088 WCh *Hyabi Bolewa yawi, Dera ya:we, Tangale yabe, Pero yabe Ngamo yabi w/ Eg. ʕ bw ”kind of bird” 1-1/3?-5-

1478 WCh *kwam Tangale kom, Bolewa kom w/ Sem. kumVy ”waterfowl” 1-2/3-2-

1539 CCh *kwak Gulfey kwaku w/ ECh Bid keeke “bird,” WCh Fyer “rooster,” Eg. (NK) kʕ kʕ “cuckoo,” Sem. kʕ akʕ Vy “bird” 2-2/2&2/3-1-

1593 WCh *keyar << kewari Montol kier w/CCh Mofu kwerekwere “duck,” Sem. kʕ ariʕ “kind of bird” 1-2/3-1-

1598 WCh *keway Sura kw´´ ky´, Angas ki, Montol kiy´ w/ ECh *kway “bird, ”
CCh *kuy “hawk,” Eg. *key “bird” 1-2/2-3-

2443 ECh *tur Nanchere turoba, Kabalay turo w/ CCh Gisiga turo “partridge,” Sem. tʕ aaʔ ir 1-2/3-2-

2494 ECh *ʔ was Mokilko ʔ osso w/ Eg. (MK) wS3 .t “poultry” 1-2/3-1-

Of course this example may not be typical. Many of the words taken back to different etyma look similar, suggesting that there may be considerably fewer than 13 distinct etyma here. (Even if they are not cognate, similar looking words offer essentially only one term of comparison.) Nonetheless there are a number of quite distinct items here, and this degree of lexical diversity within Chadic is by no means atypical. Jungraithmayr and Ibriszimow (1993:xiii) observe that even between Hausa and Tangale (both West Chadic) only 20% of the Swadesh list vocabulary is cognate and and between Sura and Migama (West and East, respectively) 26%. Given the great divergence among Chadic languages, and that OS maintain a separate comparison on the basis of Ech, Cch, and Wch, we should assume that at least three different comparisons have been made and that in most cases (roughly 80%) etymologically distinct items are involved. This would mean 240,000 to 300,000 trials, resulting in 1536 to 1800 chance two-out-of-three matches and 15 to 18 three-to-three matches. Note that in the actual case of "hen" above only two of the thirteen comparisons involve a three-to-three match (301 with Eg. and 943 with Oromo), but one of these (943) involves a reduplication of the first consonanat and looks suspiciously like an onomatopeia. Note too that only two of the "hen" comparisons (and neither of the 3-to-3) involve comparison with more than one non-Chadic branch (356 involving Eg. and Cu. and 1539 involving Eg. and Sem). Six comparisons are based on some Chadic language and Eg. only; four comparisons are based on some Chadic language and Semitic only; and one comparison is based on some Chadic language and Cushitic (Oromo) only. Given that O/S also do not reconstruct Proto-Cushitic, but treat it as seven independent sub-branches, the same methodological problem arises here. (There are, however, as OS point out, good reasons for thinking that Cushitic is not a single branch of Afroasiatic, but should be reclassified as several distinct primary branches.) Judicious reconstruction of the individual branches of Afroasitiac is the best way to eliminate this source of error.

Another way to increase the number of comparisons, in effect another way to broaden semantic leeway, is to ignore derivational and etymological history. If one considers not just the most basic or etymologically oldest sense of a word but secondary semantic shifts, dialect variant meanings, derived words no longer closely semantically connected with the source, etc., one can increase the number of comparisons, and hence the margin of error. Conversely a careful use of etymological and derivational history to restrict the number of comparisons can reduce the margin of error.

This too can be best illustrated by example. Suppose someone looks up a word in an Arabic-English bilingual dictionary, takes all the English equivalents, and looks them up in an English-Somali dictionary (a not unrealistic description of what a compartist might actually do), then the number of trials is a factor not of the number of Arabic words being compared, but of the number of English translation equivalents. Since Arabic dictionaries are traditionally panchronic, incorporating material from a viariety of sources, time periods, and places, without any indication of provenance or date of attestation, each word frequently has a variety of meanings and a variety of English translations. Further since Arabic is a morphologically rich language, in which exploitation of native derivational resources, rather than loanwords has been the preferred means of expanding the lexical stock, each root entry typically contains a variety of deverbal nouns, denominal verbs, secondary derived verb forms, lexicalized participles, etc., which have acquired a distinct semantics from the root. This observation applies most emphatically to Steingass's Arabic dictionary which E has used as his primary source for Semitic. (Note Kaye's observation "It was not a reliable dictionary in 1884, and to use it today does not make much sense.")

Based on analyzing every 200th page (p.200, 400, 600, 800, 1000, and 1200) in Steingass's dictionary, I deduced the following averages:

avg. no. of Arabic words/ page: 23
avg. no. of three consonant root entries/ page: 8
avg. no. of Eng. translation equivalents per/word: 5.2
avg. no. of Eng. translation equivalents for all words under a root entry: 13.8

Since the dictionary has more than 1200 pages, if these averages hold good, there should be 9,600 root entries in the dictionary. If only the first or most basic meaning of each of these is taken and compared against its translation equivalent in Somali or any other language then 9,600 trials have been made. In the standard hypothetical case of 25 consonants evenly distributed, this should yield an average of .61 three-to-three chance "cognates". (In other words a chance cognate should be found for every third language compared.) It should also yield for any language about 15 cases of chance matches on just the first two consonants. If a range of 5 meanings is taken into account for each Arabic entry the number of trials increases five times, and hence the number of chance matches increases to 3 and 77 respectively. If a range of 14 meanings is considered, the expected number of chance matches increases to 9 and 215 respectively.

If only the first two consonants are considered as constituting an entry, as under E's root theory, then the range of translation equivalents increases dramatically. But since we are in effect defining the number of trials as the number of Arabic entries times the number of translation equivalents, the number of entries will decrease if only the first two consonants are taken as constituting an entry (even as the number of translation equivalents rises), this step will not contribute significantly to the number of trials. The only difference with the maximal case so far considered is that if one pays no attention to semantics and normal derivational processes and simply groups all words that start with the same two consonants together, nonce words, four consonant roots, etc. will be grouped together with the productive three-consonant roots. In short the number of trials will simply be the total number of translation equivalents for all the words in the dictionary. At an average of 120 per page for 1200 pages this will be 144,000. (Of course this doesn't necessarily mean 144,000 distinct English words. Some semantic overlap is expected.) For this number we expect 230 two-out-of-three matches and 9 three-out-of-three. The total possible number of two consonant stems is only 600 and not all actually occur. (Total possible number of 2C combinations in Arabic is 28^2=784 minus excluded homorganic combinations 184 = 600.) Thus under the assumption of the biconsonantal root and a liberal use of a source like Steingass, one can expect that 40% to 50% of putative Arabic "biconsonantal" roots will have purely chance "cognates" in any language in the world for which a suitably large English-x dictionary exists.

Under this way of increasing semantic leeway, too, one might expect semantically vague reconstructions and many synonyms. This is exactly what is found in E. There are 84 instances of synomous entries reflecting 286 reconstructions. There are also many near synonyms, e.g. be moist (4 entries), be wet (another 4), be very wet (1), become wet (4), become thoroughly wet (1); or raise (7), rise (20), rise up (2), rise above (1). Interestingly, in complete contradistinction to O/S the vast majority of these (65 of 84) are verbs. There are only 15 synomous noun entries, and no reconstruction at all for "hen" or "bird" or most of the nouns in the Swadesh list. One strongly suspects this bias reflects the traditional organization of Arabic dictionaries, where a verb stem is always given as the head entry, even where a noun of the same root is more etymologically primitive. If so it indicates a use of the Arabic sources much in the way outlined above.

The question raised at the outset can now be answered in general terms at least. Both E and O/S have adopted assumptions which have insured a high probability of a chance match and have used sources in such a way as to insure a high number of comparisons. The two things together insure a high number of accidental "cognates"-- a high margin of error. Yet because the assumptions adopted and the use of sources is slightly different, the margin of error extends, as it were, in a different direction in each case.

4. Where do we go from here

The essential paradox of long and medium range comparison can be stated as follows. Over time words are lost from languages, obscuring evidence of cognateship.

Genuine cognates that remain are subject to semantic shifts which may make them difficult to recognize. At the same time loss of segments, fusion of segments, and odd changes like metathesis make regular sound correspondences increasingly difficult to identify. In order to compensate for these factors, a comparatist may relax the criteria for a correspondence, and broaden the search for cognates by allowing more and more semantic leeway. The result of this however is necessarily to allow in a high degree of "noise" or purely chance matches, until eventually anything found becomes statistically meaningless.

Fortunately, in the case of Afroasiatic we have not reached that point yet. Although I have given both authors credit for applying the comparative method correctly, insofar as they have both respected the principle of regular sound correspondences and the uniformitarian principle in reconstructing the sound system, a closer look reveals certain problems that might under a stricter analysis be termed errors of method. These can be grouped under four headings: failure to reconstruct intermediate nodes; failure to restrict possible meanings to the earliest etymological attestation; failure to respect the uniformitarian principle in lexico-semantic reconstruction; and failure to respect the uniformitarian principle in morphological reconstruction. Interestingly while each of these may be judged problematic on general principles, each has a methodological implication as well, because each increases the margin of error and allows a measurable degree of randomness into the results. Correcting them should lead to a more plausible reconstruction for Afroasiatic.

4.1 Use of multiple language data from single sub-branches with failure to reconstruct at the subbranch level

This is a criticism that relates to O/S principally. E is apparently not guilty of it. Although since he does not publish the evidence upon which his branch level reconstructions are based, it is difficult to tell. As noted above this allows for a choice of potential cognates among many languages of a given subbranch, increasing the number of comparisons and hence the margin of error. The solution is of course judicious reconstruction at the lower nodes of the tree.

4.2 Panchronic use of the sources, choosing meanings from various time depths

This relates especially to E's use of Arabic data. But possibly O/S are also guilty of it in their use of Egyptian data. (See the entries under "water", where items from several different stages of Egyptian are correlated with words from several different branches of Chadic.) This too leads to an increase in the number of comparisons.

To take an analogy from archeology, the historical linguist who works with a dictionary like Steingass (or most Arabic dictionaries except Quranic concordances based on clearly defined corpus) is in the position of an archeologist presented with a truckload of looted artifacts stripped from a variety of sites and historical strata and thrown together at random. Without the crucial information about provenance and time depth, it is difficult or impossible to interpret the items in a historically meaningful way. A major desideratum of the field is an etymological dictionary of Arabic. In the meantime the solution might be to work with existing etymological dictionaries of languages like Hebrew and Ge'ez and to avoid using Arabic meanings which can not be confirmed elsewhere in Semitic. As it is E ends up reconstructing back to Proto-Afrasiatic level polysemies which have arisen comparatively recently in the history of Arabic.

4.3 The uniformitarian principle and semantic reconstruction

The uniformitarian principle applied to historical linguistics says in effect that reconstructed languages should look like "real"-- attested-- langauges in their overall structure. I have observed that both E and O/S respect this principle in phonological reconstruction, but the same cannot be said about lexical and grammatical reconstruction. With regard to lexicon both sources reconstruct a large number of homonyms and synonyms which is excessive for a natural language. I have noted in passing above that about 40% of the entries in O/S are part of a group of synonyms, as are about 28% of those in E, not counting near synonyms. In other words only 60% and 72%, respectively of the reconstructed items represent the unique item reconstructed for that meaning. There are also a large number of homonyms. I calculated that in O/S 858 reconstructed items or 32% of the total are part of a group of homonyms. The most prolific of these is *mar which is given as the reconstruction for 14 distinct cognate sets. Ehret too, reconstructs a high degree of homonymy for Proto-Semitic, as he acknowledges: "The analytical groupings of roots distinguished by this approach reveal more sharply than ever before the pervasive polysemy and homonymy of a large proportion of Semitic verb roots... it becomes clear that in pre-proto Semitic, because of the loss of stem vowel distinctions in verbs, a great many previously distinct roots had fallen together." But he tries to avoid this at the AA level by reconstructing distinctive vowels and tones for otherwise identical two-consonant stems.

Excessive synonyms and homonyms violate the uniformitarian principle and are therefore problematic on substantive grounds. But they also reveal something about methodology. Excessive synonymy results from allowing a broad range of semantic equivalence. This also increases the number of comparisons made, and hence the margin of error. Excessive homonymy results from ignoring features of phonological stucture (usually third consonants). For example in O/S we find Arabic qarn “horn”, qaar -at “hill”, qarra “be cold”, and qaraʔ a “read”, reconstructed as qar “horn”, qar “hill”, qar “be cold”, and qar “shout”, with the extra segmental material which distinquishes the words reconstructed out. Ignoring this material increases the probability of finding a match, hence the margin of error. Thus the following equations can be set up:

broad semantic leeway = increased number of trials = reconstruction of synonyms
ignoring segments = increased probability of a match = reconstruction of homonyms

4.4 The uniformitarian principle and morphological reconstruction

Both sources reconstruct lexical relationships in the attested languages as going back to derivational relationships in the proto-language. (In at least one case OS also reconstruct a derivational relationship-- an Arabic singular-plural pair qarya(tun), qura(n)-- as going back to lexical ones in Proto-Afroasiatic, reconstructions 1568, 1589.) E does this in a thorough-going way and the result is proto-language in which the basic vocabulary consists of a set of polysemous verbal roots with abstract and general meanings, while verbs with more specific meanings, and almost all nouns are derived by suffixation. Further all consonants in this language can serve as suffixes. I would argue that both points are violations of the uniformitarian principle. In general the underived, basic vocabulary of a language and specific and concrete, while abstract words are formed by derivation. Further it is rare for the full consonant inventory of a language to be used in its productive derivational morphology. Finally, given the well-known homorganic cooccurence restrictions on Afroasiatic roots (Greenberg 1950, Bender 1974), each suffix would have to have at least one allomorph at a different point of articulation and a hideously complex system of dissimilation rules would be needed to account for their distribution. E’s justification for this is revealing “With respect to triconsonantal roots in Semitic, a[n] ... explanation of the third consonant as lexicalized pre-proto-Semitic suffixal morphemes has now been put forward (Ehret 1989).... It has been applied here without apology because, quite simply it works.” This is the worst possible argument in favor of the hypothesis. As the above calculations have shown, such a procedure should indeed work quite well as a way of generating random noise.

When considering in general the problem of biradicalism vs. triradicalism, the following calculations should be taken into account. The proportion of potential two consonant combinations to three consonant combinations is x^2::x^3, where x is the total number of consonants in the inventory. Unless the language has a rich vowel and/or supra-segmental inventory, two-consonant stems will not provide an adequate vocabulary. The percentage of potential two-consonant words as a total of all two- and three- consonant words (assuming no distributional restrictions) is given by the formula x^2/x^2+x^3, where x is the number of consonants in the inventory. If distribution of actual words is proportional to potential, then you would expect to find that biconsonantal words are only about 4% of the total:

For 20 consonant inventory 400/8400=.0476
for 25 consonant inventory 625/625+15625=625/16250=.038
For 28 C 784/784+21952=784/22736=.0345
For 30 C 900/27900=.0323

If two genetically related languages each have 4% biconsonantals and 96% triconsonantals, it is logical to assume that the common proto-language had the same proportion. However, pure chance matching would yield a completely different proportion. Assume that lexical resources and the comparatist's imagination is such as to allow 200,000 comparisons, 4% of these or 8000 should involve a biconsonantal and 96% or 192,000 should involve a triconsonsonantal. For a 25 consonant inventory the chance of finding a 3 out of 3 match is .000064, yielding 12.3 expected matches out of 192,000. For multiple two to three matches as in O/S the chance of match is .0064 yielding 51 expected matches out of only 8000. In other words, if one starts by comparing two languages in which the proportion of 2-consonant words to 3-consonant words is 4::96, the proportion of chance matches involving 2-consonant words is in a proportion of 81::19 to those involving 3 to 3 matches in 3-consonant words, and this remains constant regardless of the number of comparisons. If two-out-of-three matches are allowed as in E, the probability of finding a match is lower (.0016) but 100% of the items are available for comparison, yielding 320 expected matches, a proportion of 96::4 in favor of two-consonant matches, the exact opposite of what was started with.

References

Bender, M. Lionel. 1978. "Consonant Co-occurence Restrictions in Afroasiatic Verb Roots" in Atti del Secondo Congresso Internazionale di Linguistica Camito-Semitica, ed. Pelio Fronzaroli. Istituto di linguistica e di lingue orientali universita di firenze.

Ehret, Christopher. 1995. Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, tone, consonants, and vocabulary. University of California Press. Berkeley.

Ehret, Christopher. 1989. “The Origin of the Third Consonants in Semitic Roots: an Internal reconstruction (Applied toArabic). Journal of Afroasiatic Linguistics, 2,2, pp. 109-202.

Greenberg, Joseph. 1950. The Patterning of Root Morphemes in Semitic” Word 6, 162-181.

Jungraithmayr, Herrmann & Dymitr Ibriszimow. 1993. Chadic Lexical Roots (2.vols.). Dietrich Reimer Verlag. Berlin.

Kaye, Alan. 1996. Review of Ehret (1995) in Canadian Journal of Linguistics.

Orel, Vladimir and Olga V. Stolbova. 1995. Hamito-Semitic Etymological Dictionary: Materials for a reconstruction. E.J. Brill. Leiden.

Ringe, Donald. 1992. On Calculating the Factor of Chance in Language Comparison. The American Philosophical Society. Philadelphia.

Steingass [1884] 1978. Arabic-English Dictionary. Curzon. New Delhi.