NON-MAINSTREAM HISTORICAL LINGUISTIC CLAIMS:
A SKEPTICAL CRITIQUE

Mark Newbrook

Readers who have followed my column *Language On The Fringe* (62 instalments) and my other contributions to the ASKE newsletter over the years will find some of this material familiar. So too will readers of my 2013 book *Strange Linguistics* (Lincom-Europa, Munich). But I hope that such readers will still find it useful to read this summary of my ideas about non-mainstream historical linguistic thought. And of course I also hope that in this new open-access domain I will find new readers. I will always be happy to engage any interested reader in discussion of these matters.

This present paper was originally written as the sole linguistically-oriented chapter of an online volume on non-mainstream historical ideas. I am grateful to the first-point-of-contact editors and the anonymous reviewers who worked with me on the material. After the paper was approved, however, the final-stage editor sought to make many specific changes to the text; I rejected most of these, and the paper was excluded from the volume. I am now grateful to Michael Heap for accepting it as my first contribution to the new-format ASKE website.

Skeptical linguists are part of the world of organised skepticism, which seeks to understand and to critique (fairly) non-mainstream ideas (theories or specific claims) developed in recent decades (or centuries), which are proposed by amateurs (non-linguists) by way of replacements for mainstream positions (or for current mainstream views to the effect that the cases in question cannot at present be resolved) but typically lack supporting evidence that might be judged adequate and which from a scientific, philosophical or historical perspective appear implausible or even irrational (but which may command a degree of support in the relevant communities).

Mainstream linguistics can be regarded as dating from around 1800 CE, when the relevant scholars began to seek scientific status for the discipline and to eschew impressionistic, irrational and non-scientific notions and methods of this kind. The discipline was for many decades predominantly historical in focus, and historical linguistics is still a major branch of linguistics *(note 1)*.

Skeptical linguistics, specifically, involves dubious ideas about human language, which have been proposed and promoted by writers who are/were not themselves professional linguists. There are very few identifying skeptical linguists; they constitute the very small intersection of two specialised and not otherwise associated sets of scholars: qualified linguists on the one hand and active skeptics on the other. Skeptical linguistics arose piecemeal in the context of the growth of organised skepticism from the 1970s onward *(note 2)*.

I deal here (in a summary manner) with aspects of the findings of 'skeptical linguistics' as they apply to amateur historical linguistic claims.

The amateur authors in question have struggled to attract the attention of mainstream linguists (other than committed skeptics) or of journal-editors, and they have typically reacted negatively to such skeptical criticisms as have occasionally been successfully brought to their attention by mainstream commentators, regarding the mainstream as hidebound and biased and declining to engage in serious discussion. For instance, R.A. Strong and Bernard Macklin

(*note 3*), who argue, partly on inadequate linguistic grounds involving alleged 'puns', for stronger links between Greece and Egypt than are generally accepted, dismissively reject the mainstream objections of which they have been made aware (they also display some confusion as to the scope of mainstream historical linguistics, interpreting it very narrowly.)

Like their equivalents in other disciplines, all these writers are 'revisionists'; they proclaim 'alternative' ideas about linguistic history – either revived older ideas or newly conceived ideas. Only a few of these writers have any useful knowledge of linguistics (although some of those who do have some such understanding proceed as if as their knowledge is at a more advanced level than it actually is). Most such writers are simply unaware of established mainstream scholarly linguistic thought. These shortcomings are not usually apparent to the non-specialist reader. Indeed, readers who share some of these writers' background beliefs or attitudes may be inclined to accept their linguistic claims as authoritative, although they themselves often know no linguistics at all.

Some non-mainstream ideas about the origins of languages or influences between languages are primarily motivated not by an interest in the languages themselves but rather by ideas about the histories and cultures of the peoples who have used them; they frequently involve nationalistic, racial or religious notions relating to contentious issues in history (and modern politics). The linguistic notions can thus become inputs to – or support for – racist or other prejudiced positions. This can have practical consequences in fields as important as the content of curricula for the education of major minority groups, such as African-Americans or Australian Aborigines, or even for relations between entire communities espousing different religions and/or having different ethnic or national origins.

Some such writers, indeed, uphold blatantly pseudo-historical theories and the idea that there has been conspiratorial misrepresentation of historical developments; they may cite associated fringe linguistic claims as support for these extreme non-linguistic notions. For example, some writers hold that Indian civilisation and Hindu religion (Vedanta) once dominated the world and that knowledge of this situation and of how it was later illegitimately superseded has been suppressed by non-Hindu interest-groups. These writers often attempt to support this view with extreme claims about the antiquity and ancestral status of the Sanskrit language; for Hindu believers, Sanskrit, as the language of their scriptures, is of overwhelming significance. Sanskrit etymologies are proposed (often dogmatically) for place-names (e.g., English county- and city-names) and for other words from languages all around the world. Ideas of this kind have created tension; for example, this Sanskrit/Hinduism-based work has drawn hostility from Muslim thinkers (*note 4*). Constructive dialogue on such issues is virtually impossible. Some ideas of this kind are therefore potentially very damaging.

Some non-standard historical linguistic claims involve the ultimate pre-historic evolutionary origins – or in some cases the alleged non-evolutionary origins – of human language *per se*, including both *Homo sapiens* and pre-*sapiens* species (as are discussed by mainstream palaeolinguists). Most such claims, however, deal with linguistic change in 'deep' historic times: mainly the years BCE, especially second millennium BCE and earlier. These relatively recent dates are to be regarded as 'deep' in a linguistic context because writing is known only from the last 5,000 years and for the early part of this period is confined to a small number of languages. Spoken and signed language is obviously

ephemeral, and very few specific linguistic facts can be determined for the period before 4000 BCE; earlier languages and language-stages can be reconstructed only to a limited extent and with high levels of uncertainty.

The claims in question here deal with the origins of specific languages (in supposed ancestor languages) and with historical relationships (involving common origin or later contact; see below on this contrast) between languages normally thought of as unrelated and/or unconnected. Such claims typically rely mainly upon unsystematic, superficial similarities between words and the like.

Many of the claims discussed here involve 'diffusion': the differentiation or (here) the alleged differentiation of single earlier languages into groups of related later languages used in different locations, and the spread of specific words (etc.) from one language to another, often with modification of spoken and/or written form and/or with change of meaning. The notion of diffusion is itself very well-established in mainstream historical linguistics (see below); the issue here involves egregiously ill-supported diffusionist interpretations of particular sets of linguistic data.

Indeed, non-mainstream claims of this kind often go further, being 'hyper-diffusionist' in nature, involving differentiation from much older common ancestor languages than is accepted as demonstrated by mainstream historical linguists and historians. Their proponents may even trace all human languages to a single identified ancestor language, the *Ursprache* ('primeval language') or, in more recent terminology, 'Proto-World'.

The mainstream view is that if such a single ancestor language ever existed (see below) it cannot now be identified, still less described in any detail, because of the lack of hard evidence about very early (pre-literate) languages and the large time-depths involved.

As noted above, many claims of this kind (like some other claims with a non-historical focus) relate to nationalistic, racial or religious stances with their own pseudo-academic backgrounds. These claims often involve nationalistic or religion-derived biases in favour of languages which their proponents believe were especially ancient, often indeed to be identified with the *Ursprache*, and thus widely dispersed and/or influential at early dates. These are often earlier stages of writers' own native languages, or languages of religious or ethnic significance which are important in their own cultures.

Examples of this latter type include Classical Sanskrit, Biblical Hebrew and Ancient Greek. In the last two centuries, Sanskrit has been especially popular in this context, especially among Hindu believers (Indians or converts). The long superseded early nineteenth-century mainstream view that it was the ancestor of all the 'Indo-European' languages (the large language 'family' which includes most of the languages of Europe and Western Asia) has often been exaggerated into the idea that it represents a pan-human *Ursprache*. Many amateur writers trace many words from many languages to Sanskrit roots, providing no persuasive evidence.

Languages of unknown, disputed or non-apparent affiliation, especially genuinely mysterious, 'isolated' languages with no known related languages, such as Basque (extant) and Sumerian (ancient), are also commonly identified as ancestor languages (or as otherwise especially important). In still other cases, the language presented as the *Ursprache* (or as a later but still very early ancestor language for a specific group of languages) is one which has

been reconstructed or invented by the author in question (and sometimes is presented in detail).

The various specific non-mainstream diffusionist claims made may be divided into three main types:

Claims about the origins of specific languages and about the historical relationships in historic times between languages normally thought of as unrelated and/or unconnected

Claims about the identity and nature of the ancestor languages from which later languages or groups of languages are said to be descended

Claims about the etymologies of specific words; this includes onomastics (the etymological origins of personal names and place-names)

As noted above, linguistic diffusion is not a concept invoked **only** by non-mainstream thinkers. Indeed, it was observation of diffusion that led to the 'birth' of modern historical linguistics around 1800. A language may differentiate into various later languages by way of diffusion as it comes to be used in various areas remote from where it was first used. Such languages typically display equivalent forms which retain some similarities but which have diverged in some respects. The affiliations/ relationships between languages which manifest patterns of this kind are described as 'genetic'; as they diversify over time, 'families' of languages develop cladistic 'tree-structures' resembling those found in the biological differentiation of species or in families made up of individual animals *(note 5)*. Indeed, the discipline of historical linguistics developed in parallel with the work on biological differentiation which originated with Linnaeus in the eighteenth century. It must be emphasised, however, that the term *genetic* involves only metaphor/analogy in a linguistic context. The linguistic characteristics in question are acquired, often with modification, by each new generation of speakers, not genetically inherited.

The established method of comparing known, possibly related languages with a view to arriving at accounts of the language families in question and at reconstructions of the earlier languages from which they are judged to derive is known as the 'comparative method'. For example, it can readily be deduced by comparison that French *pain*, Italian *pane* and Spanish *pan* are all derived from Latin *panis*; all four words mean 'bread'. These words resemble each other but also display obvious differences. Sets of words such as *pain*, *pane*, *pan* and *panis* are known as 'cognates': they descend from a common ancestor word or word-root in the common ancestor ('mother') language of the family in question, or one of them (in this case, *panis*) is itself that common ancestor word.

This specific case is a favourable case for analysis because we have (now mainly in writing) a form of the mother language of these 'Romance' languages: Latin. In most cases, the mother language of a family is not itself found (it is 'unattested'), because a) it is no longer spoken or known and b) it was seldom or never written (or the written records are lost). However, some such lost mother languages can be reconstructed in part by careful comparison of the 'daughter' languages. For instance, the ancestor of the Germanic languages (English, Dutch, German, Scandinavian, Gothic, etc.) does not itself survive in either spoken or written form, but it can be reconstructed in part by comparison of the existing languages (especially what we know of their older forms) and from older Germanic languages which do survive in written form. Mother languages of this kind are called 'proto-languages'; and

unattested and reconstructed mother languages, which naturally have no existing names, are given names commencing with 'Proto': for example, the reconstructed mother language of Germanic is 'Proto-Germanic'.

In context, the crucial point regarding cases such as that of the Latin/Romance words for 'bread' is that diffusion/differentiation of this kind is not haphazard; it is subject to principled constraints. Any claim to the effect that there is a common source behind a range of similar linguistic word-forms, having similar meanings and a) found in languages which are here said to be related or b) said to be related as individual word-forms, must be justified against these patterns of differentiation which have been discovered through over two hundred years of historical linguistic scholarship. Most importantly, it has been found that linguistic differentiation of this kind is largely systematic, that is to say regular and indeed largely predictable once the patterns are known. (This is one of the most important respects in which historical linguistics can claim to be 'scientific'.) For instance, not only *panis* but most Latin words which ended in *-anis* came to end in *-ane* in Italian, *-an* in Spanish, etc.: Latin *canis* ('dog') became *cane* in Italian, etc., etc.

There are often individual exceptions to a given pattern of systematic correspondences, but these are relatively few, and where detailed information is available they can generally be explained. In addition, of course, some words may be altogether replaced; for example, the main modern Spanish word for 'dog' is *perro*, a form unrelated to Latin *canis (notes 6 & 7)*.

However, most non-mainstream, often hyper-diffusionist proposals involving hitherto unrecognised common origins for word-forms (etc.), such as those under discussion here, are grounded in **unsystematic** (and often individually superficial) similarities which can readily be attributed to chance; they do not meet these criteria for acceptance as genuine.

Very damagingly for non-mainstream claims of the type under discussion here, there are superficially similar forms – often with similar meanings and sometimes even found in related languages – which are nevertheless themselves demonstrably unconnected and only accidentally similar. Even with some knowledge of linguistics, one might imagine, for example, that Latin *habere* and German *haben* are cognates; after all, they are very similar (the stems, as opposed to the grammatical 'infinitive' suffixes, differ only with respect to phonetic detail), they both mean 'have', and in this case we know independently that the languages themselves are 'genetically' related (they are both members of the large Indo-European family, albeit from different more specific families). But in fact these words are not cognate; they are unrelated, and their similarity is unsystematic and accidental. German *haben* does have a Latin cognate, but this is *capere* ('take', 'capture'). German and other Germanic words beginning with *h-* normally have Latin/Romance cognates which systematically display *c-* – as in *Hund* and *canis* ('dog'), *hundert* and *centum* ('hundred') – and not *h-* (because of different changes within the two language 'sub-families' as they diverged from their Indo-European origins).

We know that cases of this kind involve 'false cognates' because the forms and the entire languages are well documented. Where we do not have this information (for instance, where very ancient and/or unwritten languages are involved – and where few specific words are known, so that degrees of systematicity cannot be determined), we are simply unable to establish whether or not such pairs and sets of forms are cognate.

It is, therefore, agreed among linguists that sets of unsystematically similar forms with similar meanings are not at all likely to be cognates *(note 8)*; and there is certainly no reason to regard them as demonstrably cognate (or as otherwise historically connected, for instance through 'borrowing', on which see below). However, high levels of systematicity are very rare indeed in sets of cases given as evidence in amateur work. In fact, amateurs typically appear altogether unaware of this theoretical and methodological requirement.

One example of non-mainstream thought in this area involves the self-published amateur Michael Tsarion's proposal *(note 9)* that the unsystematically similar forms *arya* (a key term in Indic studies which has had a range of uses), English *area*, the *-erra* in *terra* (Latin: 'earth'), the *-aria* in *Bulgaria*, *Hera* (Greek goddess-name), etc. (drawn or derived from a range of Indo-European languages and themselves usually deemed unconnected), were in fact cognates, from an ultimately Irish Gaelic root (and that their meanings were thus originally the same). Tsarion is himself Irish: his focus upon Gaelic illustrates the tendency of authors of this kind to identify a language of personal significance as an ancestral language or as otherwise especially significant.

Like some other such writers, Tsarion attributes the unacknowledged diffusion of linguistic forms in historic times to the aftermath of a catastrophic event, in this case the destruction of Atlantis, which of course is itself not acknowledged as factual by mainstream scholars *(note 10)*.

Some of the word-forms presented by Tsarion have other, known etymologies and origins, and even where this is not the case the only reason for considering the possibility of common origin is the superficial, unsystematic similarity between them. Tsarion shows no sign of awareness of the question of systematicity or of the obligation that lies upon him to defend his proposed equations (as opposed to any supposed obligation lying upon critics to provide evidence that they are invalid).

Using the loose methods adopted by most non-mainstream thinkers, one can in fact 'prove' (spuriously) that almost any two languages share considerable amounts of vocabulary, or that a given word – or even a longer text *(note 11)* – can be identified as 'really' being in a language other than what it appears to be in. In fact, non-mainstream etymological work of this kind is itself so varied, in its 'findings' at any rate, that any major loosening of the standards of evidence for cognatehood, which non-mainstream claims such as those discussed here require, would have the consequence that many alternative proposals (involving, for example, a whole range of different languages of origin for the same words) would be roughly equally plausible. However, such proposals would all obviously contradict each other. Only one of them, if any, could be correct in each case – but there would be no rational way of identifying which of them was correct. In that event, the reasonable conclusion would probably be that we could not say much at all about deep-time philology or older etymologies with any confidence. Mainstream linguists would regard such a conclusion as a last resort and as not in any way warranted by the actual evidence.

The onus therefore has to lie upon those who present either novel etymologies or novel equations of purported cognates, based on unsystematic similarities, to argue that the scholarly tradition of historical linguistics is mistaken in these respects. But this has seldom even been attempted (even by the few such writers who appear aware of these issues), and so far has never actually been accomplished. In fact, most non-mainstream authors focus only

on language data (mainly individual words rather than systems) which they perceive (or, because of extraneous religious or political motivations, want to perceive) as supporting their own ideas; they pay no attention to data which suggests other sets of cognates or links involving other languages and cultures. Indeed, they typically ignore both mainstream notions about such links and rival non-mainstream claims. Furthermore, they usually make no serious attempt to support their own ideas; typically, as exemplified by Tsarion, they merely invite their readers to accept their own preferred equations and links as obviously valid once they have been noticed.

In addition to 'genetic' cognatehood, linguistic forms may come to be shared across languages in a second way: contact. Unrelated or distantly related languages can influence each other by way of diffusion, through the movement of their users around the world. For instance, members of one cultural group may undertake journeys, for various reasons, and may thus arrive in the territory of another group; or members of two groups may come to meet frequently in a third location, for instance if both groups are trading with a third group. In such cases, knowledge of other groups' languages (sometimes to the point of bilingualism) is common, and the cultures and languages may influence each other. Grammatical constructions and features of sound-systems may be adopted from one language into another; but the most common aspect of this phenomenon involves the 'borrowing' or 'transfer' of individual words. A familiar example is the English word *restaurant*, which was 'borrowed' from French some 200 years ago. (Of course, the English word now has an Anglicised pronunciation, and most English-speakers do not even know that it is a 'borrowing'.)

Some non-standard historical linguistic claims involve not cognatehood as described above but rather the supposed 'borrowing' of individual word-forms between languages which are said not to be 'genetically' related as wholes but to have been in significant and influential contact in earlier times – involving historical scenarios which are not recognised by mainstream linguists and historians. As with claims involving cognatehood, the linguistic (and other) evidence for such claims is typically inadequate. Systematicity is not as important in such cases as it is where cognatehood is involved, because individual words are usually in question; but it is still a factor.

Alleged 'borrowing' is often contentious even where there is some supporting non-linguistic evidence. It has been claimed that the Polynesian word *kumara*/*umara* ('sweet potato') is a 'borrowing' from Ecuadorian languages where there are phonetically similar words referring to this vegetable. The sweet potato does not appear to be indigenous to Polynesia and clearly did spread westwards across the Pacific in pre-modern times in some manner or other. The suggestion is that Polynesian voyagers reached South America and returned bearing the hitherto unfamiliar vegetable and its extraneous name, or that third-party voyagers brought the thing and the word across the Pacific (*[note 12](#)*). This is no means historically impossible or even implausible. But on the specifically linguistic front no assessment of degree of systematicity is possible where only one word-form is in question.

Especially (but not only) in cases where a given language is identified by a non-standard thinker as the ultimate ancestor (mother) language of humanity, or at least of all known later languages (the *Ursprache*), it is often alleged that mainstream scholars, with their very different views on early history, on language origins and on the dating and recoverability of

any *Ursprache*, have ignored or suppressed the 'truth' of the theory in question, because of cultural, political and/or religious biases.

In the context of mainstream linguistic thought on such matters, two points should be noted. Firstly: it is not regarded by mainstream linguists as certain that there ever was one single *Ursprache*; humanity may have developed language more than once. It **is** possible that all known human languages (whether still used or not) descend from one common ancestor ('monogenesis'), either because humanity did in fact develop language only once and the phenomenon diffused from that one starting-point, or because only one ultimate ancestor language, out of a number which once existed, has left any known descendant languages. On the other hand, it is possible that the known languages descend from multiple ancestors. There are on present reckoning over one hundred language families which by definition cannot currently be shown to have ever had common ancestors *(note 13)*.

We remain ignorant on this first point precisely because of the second point needing to be made here: it is generally agreed by linguists, on the evidence available, that any *Ursprache* or multiple ancestor languages must have been spoken so long ago (at least 70,000 years BP, probably more like 150,000) that given the recent origin of writing and the observed range of rates of linguistic change it/they cannot possibly be reconstructed in any detail *(note 14)*.

It will be seen from the above that the main method employed by non-mainstream diffusionist thinkers involves non-standard amateur etymologising: claims about the origins and relationships of individual words. The author starts with sets of superficially and unsystematically similar words and word-parts with similar meanings, taken from a range of (semi-)relevant languages. These languages are generally identified by mainstream linguists as unrelated in historic or late pre-historic times and as having had no influential pre-modern contact; the words themselves are therefore regarded as unrelated, and any phonetic or semantic similarities they possess are thus deemed accidental. In contrast, it is claimed in the relevant non-mainstream works that the similarities of form and meaning between the members of many such sets of words show that they are in fact cognate. This claim is then used to argue that the cultures and groups of speakers associated with the languages are connected, and often forms part of a case for a pseudo-historical account involving important events not recognised by mainstream historians *(note 15)*.

For instance, it is observed that the male name *Madoc* is found in Welsh and that the male name *Modoc* is found in Mandan (USA). It is held that the two forms are so similar that they are very probably etymologically related and that cognatehood or 'borrowing' is involved. On the basis of a limited number of individual cases of this kind, involving individual pairs of words displaying unsystematic and often superficial similarities, it is deduced that the Welsh and the Mandans had a common ancestor culture or else experienced influential contact (not recognised by contemporary mainstream scholarship) in remote times. This particular case involves contact/'borrowing' arising from the supposed voyages of the medieval Welsh prince Madoc to North America *(note 16)*.

In addition to the lack of focus on systematicity, non-mainstream etymologising of this type may be countered by two further general observations. Firstly, there are millions of words and word-parts in the several thousand known languages; and there are only so many common sounds and sound combinations. The calculable probability of pairs of superficially and unsystematically similar words in apparently unrelated languages having very similar or

the same senses by chance is in fact much higher than most non-linguists – including non-mainstream writers of this kind – generally imagine. See above on the similar but unrelated forms *haben* and *habere*, involving two languages which are themselves related.

Therefore, on these grounds too, it can again be stated that superficial, unsystematic phonetic similarity between isolated words and/or meaningful word-parts taken from different languages is in itself no evidence of cognatehood or of any genuine, non-accidental connection, even if there are many such words or if their meanings too are similar. As mainstream linguists argue, there is vast scope for accidental similarity between the words of unconnected languages *(note 17)*.

For remote periods or scantily-represented language 'families' where the comparative method cannot be applied, statistical considerations of this kind, under names such as 'mass comparison', can be applied (with caution) in attempts to determine the approximate likelihood that genuine relationships obtain between words, as opposed to accidental similarities. Linguists themselves differ as to the reliability of such alternative methods. Some linguists have urged that the comparative method and the reconstruction of proto-languages can legitimately proceed (and indeed often have in fact proceeded, often without this being acknowledged) only **after** 'genetic' relationships have been provisionally established. Indeed, the comparative method is **not** to be seen as having the pre-eminent status which is ascribed to it by most historical linguists. The best known linguist of this kind is Merritt Ruhlen *(note 18)*. A clear majority of linguists regard these linguists as 'mavericks' and treat their ideas as suspect. For example, R.L. Trask presents a fair-minded but ultimately (provisionally) negative assessment of Ruhlen's claims *(note 19)*.

Because amateurs proposing non-mainstream theories are typically unaware of the comparative method and indeed are often concerned especially with remote periods where it is not applicable in any case, these statistical methods, to the extent that they are deemed valid, are frequently relevant to the consideration of their claims.

The **precise** statistical likelihood of accidental (unsystematic) similarity between pairs of words, as opposed to genuine connectedness of one kind or another, depends upon a number of factors. (I am not here taking into account non-linguistic factors such as the likelihood of contact between the communities in question, or of common origin, in geographical or chronological terms.). The main factors are these:

(a) The degrees of phonological and semantic similarity between words which are required if they are to be regarded, *prima facie*, as probably shared (although such decisions are often arbitrary to a degree; for example, scholars might disagree as to whether a word *san* meaning 'scarlet' in one language and a word *zen* meaning 'orange' in another language were similar enough in form and/or meaning to be regarded, *prima facie*, as probably shared). (Of course, where sufficient material is available and the comparative method can be applied, the degree of phonological systematicity will resolve such cases, altogether superseding the matter of degrees of phonological similarity.)

(b) The phonological systems of the relevant languages. These considerations are often rather technical, and even closely related languages may have diverged in these respects. But the main issue here is that of how a phonological sequence in one language is likely to be or indeed **can** be represented in a language with a very different phonological system,

for instance in a contact situation. For instance, the English philosopher's name *Russell* is necessarily represented in Cantonese and other Chinese 'dialects' which lack both the phoneme /r/ and word-final /l/ as *Lo So* or the like. In such cases a 'borrowed' form may not be recognisably similar to the source form. The forms *san* and *zen* mentioned under a) above are more likely to be connected (by 'borrowing') if the language featuring *san* lacks the phoneme /z/ and has replaced it by /s/ in 'borrowing' the word. Forms such as these may be genuinely connected, or probably so, despite their limited phonetic similarity. (But in fact amateur theorisers are naturally unlikely to identify such pairs or sets of forms as cognate or as 'borrowings', whether they are in fact connected or not; they will typically not realise that such forms are connected. They will thus **miss** evidence of this kind, unless it is noticed and could possibly be invoked in support of their own claims. See also *note 7*.)

(c) The lengths of the words; for example, if two languages not known to be connected share a very short word-form such as [sa] with the same meaning, this could very well be accidental (in fact, some entire non-mainstream theories are vitiated through dependence on known or proposed very short forms of this kind), whereas if they share a polysyllabic form such as [tolpesveblig], again with the same meaning, or with transparently related meanings, this is much less likely, and the sharing of the form requires a historical explanation.

(d) The cross-linguistic frequency of the sounds and sound-sequences in question; very widely-shared sounds such as [e], [s], etc. or common sound-sequences such as [til] or [po] are more likely to be shared by chance than sounds and sequences found in relatively few languages.

The second key observation here is that etymologising should (for obvious reasons) be based not upon well-known contemporary forms (as often occurs in amateur work) but instead upon the forms of the relevant words in the oldest available versions of the languages in question – and upon the ranges of known cognate forms in the relevant language 'families'. This observation naturally applies mainly to 'genetic' links between languages rather than to those involving contact – although even in the latter case the etymologies proposed must obviously involve forms which were current at the relevant dates, not modern forms. Very many writers of this kind (often ill-informed about earlier forms) fall into this 'trap'.

Another issue here involves known or very well-grounded established etymologies for words, involving the histories of the relevant language families. Many of the novel etymological claims discussed in the non-mainstream literature fly in the faces of known or very probable etymologies, which are often very well supported with historical and linguistic evidence. Proposers of alternative etymologies need to argue that theirs are more plausible than the established ones. But this is very seldom even attempted; readers are simply invited to accept the alternative etymologies, and the established ones are hardly ever even mentioned. (See again Tsarion's proposals, instantiated above.)

Other etymological claims deal mainly with the very remote past where the actual etymologies for words and word-parts are obscure and uncertain, or simply cannot be established, at least by current methods. The point here is not that the novel etymologies offered by non-mainstream writers are known to be incorrect but rather that there is no

particular reason to believe that they are correct (especially if they are historically implausible).

In fact, as indicated earlier, it is widespread systematicity, as revealed by the comparative method, which is normally decisive, not superficially similar words or the contested etymologies of individual words *per se*. In addition, 'genetic' relationships between languages are often shown accurately by specific grammatical similarities. But few of the writers in question here know enough linguistics to deal adequately with phonology or grammar. Indeed, the vast bulk of the argumentation associated with non-mainstream amateur claims involves vocabulary, which is replete with superficial (mostly accidental) similarities and which requires much less understanding of linguistic theory or the techniques needed for describing and explaining linguistic systems *(note 20)*.

There are two special, overlapping groups of claims which are in other respects similar to those discussed above. These claims involve (a) the conscious, deliberate or semi-deliberate (often conspiratorial) concoction of known languages or language data out of other known or reconstructed (or invented) languages or data; and/or (b) ancestor languages of a specific type involving very short words.

These ideas form part of the wider non-mainstream tradition of 'pseudo-history'. The authors in question here hold that churches or supra-governmental and other powerful covert forces, which have long controlled the world or at least some of its most important institutions, have furthered and are still furthering their own agendas by jealously guarding the true knowledge of the remote past which they alone possess and promulgating the (in fact false) alternative ideas which most people accept as true. The deliberate manipulation of linguistic forms and entire languages (and the suppression of the truth regarding these matters) is one aspect of this conspiratorial activity. Orthodox scholars, it is claimed, are either part of this conspiracy or set of conspiracies, or simply dupes. The authors discussed here believe that, in contrast, they themselves have unearthed the essentials of the real truth despite the existence of this vast conspiracy to conceal it.

There are important features which separate these ideas off from those discussed above. Most importantly, proposals of type (a) as introduced above involve **planned** changes rather than normal unplanned linguistic change. Theories of this special type are more difficult to refute than claims of a more 'normal' nature such as those discussed above, and hence less scientific in character. It is much more difficult to apply either the comparative method or statistical considerations to these theories, since these methods involve unconscious psycholinguistic processes and assume 'normal', unplanned linguistic change. These particular theories, although they are typically both implausible and indemonstrable, are thus almost immune to effective disproof, either along statistical lines or on the ground that the similarities and differences between the relevant forms are too unsystematic. Even if a set of alleged changes were highly unsystematic (involving only superficial, inexact similarities) and/or otherwise implausible, it nevertheless **could** occur if it was deliberately planned as part of a project of language concoction. With deliberate manipulation, especially if applied repeatedly, forms can alter in any way whatsoever, and almost any word or language can be used as a source for a form if this is desired. Where 'anagrammatisation' or uncontrolled ordering of morphemes (meaningful word-parts) is proposed, the scope for multiple sources

becomes truly enormous and the choice of a source becomes arbitrary. Such theories are immune to empirical testing and hence vacuous.

If one accepted the reality of significant deliberate manipulation of word-forms on a large scale, the reasonable conclusion (even more surely than in the cases discussed earlier) would probably be that we could not say much at all about philology or older etymologies with any confidence, at least with respect to the specific languages in question. In fact, however, there is no good reason to believe that such massive linguistic manipulation has occurred on the scale proposed by most of these writers. Even if adequate motivation existed, the task would surely be infeasible, especially before the development of linguistic analysis in the nineteenth century.

Furthermore: where very short words are reconstructed as ancestor forms (or indeed are heavily invoked as evidence), the scope for chance similarity is typically enormous (see the example given above involving [sa]). Theories of this kind allow their proponents far too much freedom to be regarded as plausible, or (again) even as capable of being empirically tested.

If these two approaches are employed together (as often occurs), even higher levels of freedom are generated for the writers in question, rendering their claims arbitrary and altogether untestable.

One author of this kind is the linguistically-untrained Polat Kaya *(note 21)*, who claims that most non-Turkish languages were deliberately concocted and that almost all words of all languages are really combinations of short Turkish words/morphemes, deliberately corrupted – and in this case often 'anagrammatised'; that is, the order of their letters/phonemes has been 'shuffled' – so as to conceal their origin. For instance, English *accelerate* is derived by haphazard phonetic modification and anagrammatisation from Turkish *acele-eder* ('increases speed'). Furthermore, Kaya ignores the word's very clear Latin etymology. In fact, he ignores or rejects vast amounts of established information about the history of Turkish, the Turkic 'family' and other languages. He also fails to demonstrate understanding of the need to attempt justification of his novel Turkish etymologies for words in various languages against the established etymologies for these words – which could not be accomplished, because derivations of this kind are irreducibly arbitrary.

Another such claim is that of another non-linguist, John J. White, III *(note 22)*, whose publications appear in non-mainstream 'epigraphicist' journals which concern themselves mainly with supposed inscriptional evidence supporting hyper-diffusionist theories about the early diffusion of Old World languages to the Americas. White traces the words of all languages back to a morphologically simple *Ursprache* called 'Earth Mother Sacred Language' (EMSL; the name indicates supposed connections with early religion). Grammar is largely ignored; EMSL was supposedly used in remote times when a complex language was apparently not required. EMSL allegedly had only very short morphemes. White proposes fifteen basic morpheme-shapes; two of these are monophonemic, consisting of one phoneme (structurally-defined speech-sound) each (these are the vowels /a/ and /u/), nine are monosyllabic (eight of these have the form Consonant-Vowel, the ninth is /en/) and the remaining four are disyllabic (Consonant-Vowel-Consonant-Vowel). This very limited inventory yields huge amounts of homophony (many morphemes pronounced the same). A vocabulary of a usable size could exist in EMSL only by means of polymorphemic

compounds (long, each consisting of several stems in sequence) which, after centuries of modification, now appear in modern languages as synchronically monomorphemic, polyphonemic words (still long, but now perceived as consisting of only one stem). In addition, each of EMSL's very short morphemes has variant, often quite distinct phonological forms – 'allomorphs', as in *a* vs *an* (indefinite article), *knife* vs *knive*-[*s*], etc. in English) – to the point where (even without alleging deliberate manipulation) almost any of them could be interpreted as sources for the syllables of any given word in any known language. In combination, they are also subject to uncontrolled ordering (thus a three-morpheme EMSL sequence XYZ might equally well appear as YZX with the same meaning), rendering White's etymological claims even more obviously immune to testing.

Another especially interesting sub-set of cases involves the surprising alleged mutual intelligibility of languages generally believed to have no common ancestor in historic or late pre-historic times and to have had no significant pre-modern contact – or about 'out-of-place' spoken languages, said to be (or to have been) used, or at least understood, in unexpected areas. Older non-mainstream works reported many such incidents – Irish Gaelic used/understood by Mexican Amerindians, Latvian by Tatars, Welsh by speakers of Mandan (see above on this case), Navajo by speakers of Uighur (Chinese Turkestan) and *vice versa*, etc., etc. However, actual evidence was never forthcoming.

There is a further very large set of non-mainstream (quasi-)historical claims and theories involving ancient written languages and in particular the purported decipherment of the scripts (many of them unfamiliar) used to write them. This set of claims includes several well-known bodies of epigraphic material (Cretan 'Linear A', Easter Island 'Rongorongo', the 'Indus Valley Script', etc.) and individual documents (books or manuscripts, or inscriptions/alleged inscriptions on rocks or tablets such as the Phaistos Disk, the Cascajal Block or the Yarmouth Runestone). None of these bodies of data has an authoritative decipherment. Some of the scripts are otherwise unfamiliar to scholars; others are supposed instances of familiar scripts (such as Irish Ogam/Ogham) used in unexpected linguistic contexts (said here to have been adopted or adapted to write languages which are generally deemed to have been unwritten until modern times, or are usually found written in other scripts). In a few cases the language itself is otherwise unknown. Some other claims involve familiar but geographically 'out-of-place' written languages in essentially familiar scripts (for instance the Kensington Stone found in Minnesota and written in Runic Norse, or some inscriptions discussed by Menzies) *(note 12)*. But this epigraphic material involves a different range of issues from the above and would require a further paper.

Although by no means all of the claims which fall under the remit of this chapter have had seriously damaging non-linguistic upshots or have even generated much discussion outside the very specialised world of skeptical historical linguistics, they are typically associated with non-mainstream theories in domains such as history and archaeology; and if they really are as weakly supported as I have suggested here they contribute to mistaken – often tendentiously mistaken – pseudo-historical perceptions of the human past. To that extent, they require exposure by suitably qualified skeptical commentators. Many such proposals have indeed been exposed over the decades, but much of this skeptical work is unfamiliar to most non-linguists with interests in the associated non-linguistic historical claims and to lay people who have been persuaded to accept non-mainstream accounts of their favoured languages. More

work is clearly needed, especially for important languages which have hitherto been neglected in this context. It has been my intention here to outline the skeptical case against such proposals and to promote interest in such further work and in these matters more generally.

Notes & References

Click on the note number to return to the text. (A second note number refers to a second mention in the text)

1. There are many introductory books and other sources dealing with general linguistics. One such book, aimed at those who are not enrolled in a course, is David Hornsby, *Linguistics: A Complete Introduction: Teach Yourself* (Paris & London, John Murray Press, 2014). On historical linguistics specifically, introductory books include Lyle Campbell, *Historical Linguistics*, 3rd edn (Edinburgh, Edinburgh UP, 2013). With the marginal exception of the 'Ruhlen School' (see below), the views of mainstream linguists are fairly uniform on the points at issue here.

2. This paper is based on the relevant sections of my book *Strange Linguistics* (Munich, Lincom-Europa, 2013) and on my various shorter pieces dealing with specific claims of this kind. My book was the first of a very small group of book-length survey works in this area and discusses very many non-mainstream claims. The main other relevant book is Karen Stollznow, *Language Myths, Mysteries and Magic* (London: Palgrave Macmillan, 2014). There are a very few book-length and more numerous chapter-/paper-length skeptical treatments of individual claims/theories in this area, but these deal very largely with non-historical issues.

3. R.A. Strong and Bernard Macklin, *The Real Birth of Aphrodite: Multiple Proofs of the Secret Amarna Presence in Greek Mythology* (Melbourne, B.M. Minton, 1997 and 2000; two editions).

4. One such Hindu author is P.N. Oak; see for instance *Some Missing Chapters of World History* (Pune, Itihas Shodh Sansthan, 1973); *Some Blunders of Indian Historical Research* (New Delhi, Hindi Sahitya Sadan, 2008; original publication 1994); *World Vedic Heritage: A History of Histories*, 2nd edn (Pune, Institute for Rewriting Indian History, 1994).

5. Standard historical linguistics texts routinely work with this notion of 'genetic' relationship. See for instance Campbell 2013 *(note 1,* above*)*.

6. The diversification of Latin forms (as with other such cases) arose partly through contact with other languages (non-Romance); with different languages in each case, because French, Spanish etc. were being used in different regions. Some of these other languages are distantly related to the Romance languages (belonging to other branches of the larger 'Indo-European' family) and some are not related to them at all (except perhaps through a now lost pre-historic ancestor in very early times). The changes also arose partly for internal reasons; there are pressures within the structures of any language which can eventually lead to major changes (often different changes in different related languages) even if there is **no** contact with other languages.

7. (7.) In fact, demonstrable cognatehood is one thing and recognisable similarity is quite another. Some languages which ultimately share a mother language may become so far diversified that after a long period they no longer resemble each other at all (and naturally are

in no way mutually intelligible). The (remote) family relationship between Spanish, Polish and Welsh (all Indo-European) is beyond dispute, but it is in no way obvious from the written or spoken forms of these languages as they are today.

8. Again, see standard texts such as Campbell (2013) *(note 1,* above*)*.

9. Michal Tsarion, *The Irish Origins of Civilization: An Alternative History of Ireland and the World* (two volumes) (no location given, Taroscopes, 2007); http://www.irishoriginsofcivilization.com.

10. Tsarion's work is linked with a tradition of non-mainstream archaeological and historical proposals relating Ireland to Atlantis. One such work is Bob Quinn, *Atlantean Irish: Ireland's Oriental and Maritime Heritage* (Dublin, The Lilliput Press, 2005), which itself includes non-mainstream linguistic claims, for instance denying the Indo-European identity of Irish Gaelic. (Of course, very many locations have been proposed for Atlantis. I list some of these by way of background to my comments on non-mainstream linguistic proposals involving Atlantis in my 2013 book *Strange Linguistics* (pp. 42-46).)

11. For example, it has been claimed that the sentence reportedly uttered by Jesus during his crucifixion was not in Aramaic or Hebrew but in Mayan. See for instance Augustus le Plongeon, *Queen Moo and the Egyptian Sphinx* (Alpha Editions, 2019; original publication 1896), p. 38.

12. (12.) See for example Gavin Menzies, *1421: The Year China Discovered the World* (London, New York, Toronto, Sydney & Auckland, Transworld Publishers Ltd, 2003). For detailed comments on Menzies, see the relevant sections of my 2013 book *Strange Linguistics* (especially pp. 54-55), also Mark Newbrook, 'Zheng He in the Americas and other unlikely tales of exploration and discovery', *Skeptical Briefs* (CSICOP, now CSI), XIV:3 (Amherst, NY, September 2004), pp. 1-2; 'Language On The Fringe' #9, *Skeptical Adversaria*, 2010 (1) pp. 4-6.

13. While this question appears immune to empirical investigation and would not in fact explain as much as might be imagined even if it were to be settled, it is frequently discussed by linguists; see for instance Bernard Comrie's *Language Universals And Linguistic Typology* (2nd edn) (Chicago & Oxford, Blackwell, 1989) (pp. 23-24).

14. A few professional linguists have argued that more recent dates for an *Ursprache* should be accepted, and thus that the *Ursprache* (along with other very ancient ancestor languages closely descended from it) **can** (perhaps) be reconstructed in part. One work of this kind is Marge E. Landsberg, ed, *The Genesis of Language: A Different Judgement of Evidence* (Berlin, New York and Amsterdam, Walter de Gruyter, 1988). However, much of this work is now dated, and all of it is marginal to the mainstream at best; much of it involves methodology which, at least nowadays, is regarded by most linguists as too loose and approximate to be reliable. The theory of a recent *Ursprache* is in fact difficult to reconcile with various pieces of evidence, notably the modern archaeological evidence that modern humans spread as far as remote Australia by 60,000 years BP, presumably already using languages which were related to other human languages – as are, in all probability, the contemporary Aboriginal languages, given the relatively recent date of their speakers' arrival in Australia).

15. As noted, not all of the non-mainstream writers under discussion here are mainly interested in language *per se*. One reason why linguistic forms (spoken and/or written) are very commonly invoked in cases of this kind (whether or not involving an *Ursprache*) – including those where language is not itself of especial interest to the writers in question – is that they (with their meanings) appear *prima facie* much more specific and much more easily identifiable than most other cultural traits. The probability of chance similarity thus appears much lower; and the non-specialist author therefore believes that a case for a significant connection between cultures can be supported especially well through language data.

16. See partisan treatments of the Madoc legend, often with a local focus, such as Dana Olson, *Prince Madoc: Founder of Clark County, Indiana* (Jeffersonville, IN, self-published, 1987), and more critical/skeptical treatments such as Gwyn A. Williams, *Madoc: The Making of a Myth* (Oxford, Oxford University Press, 1979), Kenneth L. Feder, *Frauds, Myths and Mysteries: Science and Pseudoscience in Archaeology* (Oxford, Oxford University Press, 2017 & earlier editions (pp. 104-105 or check index), Ronald H. Fritze, *Invented Knowledge, False History, Fake Science and Pseudo-Religions* (London, Reaktion Books, 2009) (pp. 74-77), etc.

17. See for instance Donald A. Ringe, *On Calculating the Factor of Chance in Language Comparison* (*Transactions of the American Philosophical Society*, LXXXII, Part 1), Philadelphia, 1992); also, again, standard texts such as Campbell (2013).

18. See for instance Merritt Ruhlen, *The Origin of Language: Tracing the Evolution of the Mother Tongue* (Stanford, CA, Wiley (now Hoboken, NJ), 1994).

19. R.L. Trask, *Historical Linguistics* (London, Taylor & Francis, 1996) (pp. 391-396).

20. Those few non-mainstream authors who do seek to defend their claims (albeit without knowledge of the above observations) generally regard the objections of linguists with incomprehension; as far as they can see, the linguists are denying plain fact. For instance, Laura Knight-Jadczyk, *The Secret History of the World* (Grande Prairie, AB, Red Pill Press, 2008 & other editions) accepts Wilkens' linguistically naïve equations of Greek and English river-names (Iman Wilkens, *Where Troy Once Stood: The Mystery of Homer's Iliad and Odyssey Revealed*, 2nd edn (St. Martin's Press, New York, 1991) (pp. 68-71) as clearly valid, and wonders how linguists can justify ignoring his 'obviously' impressive success in correlating the two sets of names.

21. See for instance http://www.storm.ca/~cm-tntr/tur1.html. Polat Kaya has revived some ideas dating back to the 1920s, when the new republican regime in Turkey began 'modernising' the Turkish language and attempted to persuade Turks that their language was the ancestor of all human languages – partly as a pragmatic political move, with a view to persuading conservative Turks to accept borrowed words for innovations (if all words were originally Turkish, it was surely legitimate for Turkish to 'reclaim' them), but also for nationalistic reasons. The Turkish people are said to have existed since very early times and to have established the earliest civilisation; this later diffused, with the Turkish language, all over the world. This information has allegedly been suppressed by conspiratorial activity on the part of various powerful non-Turkish elements who wish to deny the above 'facts'. The words of contemporary Turkish and of all other languages were said to have diversified from ancestral Turkish roots by way of sequences each consisting of five Turkish syllables of the

form Vowel-Consonant; these syllables had basic, primeval meanings (often vague) which summed to yield the meanings of each derived word. This system had allegedly developed in very ancient times, eventually arriving at a system featuring eleven consonants and eight vowels. The system was known as 'Sun Language' (see for instance Erik Jan Zürcher, 'La théorie du <<langage-soleil>> et sa place dans la réforme de la langue turque', in *La Linguistique Fantastique*, eds. Sylvain Auroux, Jean-Claude Chevalier, Nicole Jacques-Chaquin and Christiane Marchello-Nizia (Paris, Joseph Clims / Denoel, 1985), pp. 83-91). Belief in this theory was never universal even in Turkey and waned rapidly after the death of Atatürk in 1938; but many Turks were persuaded to entertain it. For more on 'Sun Language' and Polat Kaya, see my 2013 book *Strange Linguistics* (pp. 88-89).

22. See for instance, John J. White, III, 'Earth Mother Sacred Language: A Key to Ancient Names Worldwide', *Midwestern Epigraphic Journal*, X/1 (Westerville, OH, 1996), pp. 22-33.