

To be in *International J. of Chinese Linguistics*

## **An unaddressed phonological contradiction**

Geoffrey Sampson  
Sussex University

### **Abstract**

There is general agreement on the main features of the process through which the phonology of modern standard Chinese has evolved over three millennia from that of Old Chinese. However, according to general linguistic theory, that phonological history is impossible: the theory claims that no human language can evolve in the manner in which Chinese is believed to have evolved. Furthermore, this particular strand of general linguistic theory has recently been corroborated through stringent statistical testing. Thus there is a glaring contradiction between two areas of scholarship, and to date there has been little recognition by the scholarly community of the need to resolve this contradiction, indeed little willingness to admit its existence. I argue that the contradiction is real and serious, and needs resolution.

### **A paradoxical history**

The aim of this paper is to draw attention to a paradox arising from the accepted reconstruction of Chinese phonological history.<sup>1</sup>

An international workshop on the reconstruction of Old Chinese pronunciation was held at Jena, Germany, in 2018. The leaflet which announced it said, correctly, that (although there remain plenty of differences of opinion about details) there is by now a high level of agreement on the main features of the process by which the Old Chinese of three thousand years ago evolved through the Middle Chinese of circa AD 600 into present-day Standard Chinese. The trouble is that, according to a well-supported strand of general linguistic theory, this reconstructed history is impossible. No language could behave as Chinese is thought to have behaved.

A very striking feature of the agreed historical reconstruction is repeated losses of important phonemic contrasts, creating the huge number of homophonous morphemes which are such a characteristic property of modern Mandarin. Even Axel Schuessler's 'Minimal Old Chinese' (Schuessler 2009) postulates far more possible syllables than found in Mandarin, and his 'Minimal' is intended to make the point that in reality Old Chinese probably had further contrasts.

Just between Middle Chinese and the present, many separate sound-mergers occurred at different times. All final stops -p -t -k dropped (merged with zero). Final -m merged with -n. The voice contrast was lost, leading to mergers other than in level-tone words (where it survives as the contrast between Mandarin tones 2 and 1). Velars and sibilants merged before front vowels, producing the alveolo-palatal *j q x*

sounds. And so on. When I tried estimating how the count of distinct Middle Chinese syllable shapes compared with Mandarin (Sampson 2013: 587), I came up with a ratio of 2·8 : 1. (William Wang, 1969: 10 n.3, suggested 3 : 1, not very different.) There is less agreement, as one would expect, about what happened between Old and Middle Chinese, but most people postulate similar events there, for instance Old Chinese is believed to have contained consonant clusters which were all reduced to single consonants by the Middle Chinese stage.

Where the Old Chinese vocabulary consisted predominantly of single-morpheme words, Mandarin makes heavy use of disyllabic compounds, often synonym-compounds like *péngyǒu* 朋友 “friend-friend = friend”, and this is standardly seen as having been a response to the loss of phonological contrasts: the individual syllables had become too ambiguous to use in isolation, but the compounds were unambiguous.

On the other hand, if we look at what theorists of general linguistics say about possible sound-changes in languages, there is a well-established doctrine that languages tend to avoid phoneme mergers which create numerous homophones.

This doctrine was originally developed by French-speaking linguists, most notably André Martinet (1955). Martinet wrote about the *rendement fonctionnel* of a given phonemic contrast--literally ‘functional yield’, though some linguists writing in English prefer ‘functional load’. For instance, in English the /t ~ d/ contrast has a high functional yield: it keeps many word-pairs apart, e.g. *tip* ~ *dip*, *tug* ~ *dug*, etc.; but /θ ~ ð/ has a low yield--it distinguishes *thigh* from the near-obsolete *thy* and just a handful of other pairs. Consequently the prediction would be that /θ/ and /ð/ might well merge in the future, but /t/ and /d/ are less likely to merge.<sup>2</sup>

To someone who did not know about Chinese, the idea that languages avoid phoneme mergers which create a lot of homophony (I shall call them ‘high-yield mergers’) might look like simple common sense. Matthew Baerman (2011: 2 n. 4) has quoted a citation count showing that numerous linguistics publications have treated it as an axiomatic truth. But it is an empirical, testable hypothesis, and for a while theoreticians who took the hypothesis seriously were unsure whether to embrace it. Martinet himself (1955: 58) expressed caution about its predictive value; Robert King (1967) tested it against data from Germanic languages and concluded that “functional load, if it is a factor in sound change at all, is one of the least important”.

Recently, though, quite a number of general linguistic theorists have revived the hypothesis and have argued that, despite King, we have good grounds for accepting it. I list various publications in Sampson (2015: 681–2). Particularly impressive work has been published by Andrew Wedel and co-authors, as Wedel, Kaplan, and Jackson (2013) and Wedel, Jackson, and Kaplan (2013). (Because these citations are cumbersome, I shall abbreviate them as “WKJ” and “WJK” respectively, and “WKJ OS” will refer to the online supplement to WKJ.) Wedel et al. look at a number of mergers known to have occurred in languages drawn from diverse language families, and compare their functional yields with that of hypothetical mergers which seem equally phonetically natural but which did not occur. They apply a statistical test, and find that the prediction of avoidance of high-yield mergers

is confirmed at a very high level of significance.

It is unusual for linguists to use tests of statistical significance to establish the reliability of their findings. If I did not know about Chinese, I believe I would take Wedel et al. to have settled the issue decisively in favour of the theory that languages tend to avoid high-yield mergers.

However, I know that the arguments for many such mergers in the history of Chinese are also solid. Various lines of evidence--dialect comparison, rhyme tables, graph structure, borrowings into and out of other languages, comparison with other Trans-Himalayan languages--reinforce one another in giving us a picture which in its broad outlines seems scarcely disputable. So we have a real paradox here. As William Labov put it (1981: 269), in connexion with another issue where the evidence of Chinese arguably contradicts a well-entrenched principle of general linguistics, we face the "opposition of two bodies of evidence: both are right, but both cannot be right."

### ***Mergers and splits***

I am not suggesting that the evolution of Chinese phonology has only ever eliminated phonemic contrasts and never introduced any new contrasts. The latter has sometimes happened. For instance, the contrast between *lā* 拉 "pull" and *lá* 刺 "slash" is new: if words had always obeyed the usual sound laws, there would have been no source for a word like *lā* 拉, with a sonorant initial consonant but the first tone. However, cases like this seem to be quite few, whereas the number of word-pairs which at one time contrasted and are now homophones is truly massive.

Perhaps one might think that it is somehow in the nature of historical reconstruction that it can often show us that originally-distinct forms have merged but cannot show us that particular earlier phonological forms have split to give contrasting forms later. In that case the apparent finding that Chinese phonological evolution has greatly reduced the range of phonological contrasts might be an illusion created by our methodology, rather than a real historical fact. But if splits happen, I cannot see why evidence for them would be less available than evidence for mergers. If a language acquires a new phonemic contrast through an internal development (as opposed to through borrowing vocabulary from another language or dialect), that development must presumably be a case where some sound change that creates a new sound or new combination of sounds diffuses through only a proportion of the vocabulary to which it potentially applies, so that the new sound or combination contrasts with its predecessor which lives on in the remainder of the vocabulary. The European *Junggrammatiker* tradition held that sound changes do not do that: they were supposed to apply without exception across the board. As Hermann Osthoff and Karl Brugman famously put it (1878: xiii):

Aller lautwandel, so weit er mechanisch vor sich geht, vollzieht sich nach *ausnahmslosen gesetzen ... alle wörter, in denen der der lautbewegung unterworfenen laut unter gleichen verhältnissen erscheint, werden ohne ausnahme von der änderung ergriffen.*<sup>3</sup>

But it has been claimed, particularly in connexion with Chinese (Wang 1969, 1977), that sound changes have sometimes diffused through the vocabulary gradually and only partially--this is the other possible contradiction between a general linguistic theory and the facts of Chinese, which I alluded to above in connexion with the Labov quotation. That would imply that evidence is there for a mechanism which could have increased the total range of contrasts; yet the cases of incomplete diffusion which actually occurred in Chinese, if they created new contrasts rather than merely giving individual words new shapes in terms of an unchanged range of contrasts, were far outbalanced by the loss of contrasts produced by mergers.

William Wang's theory of 'lexical diffusion' remains highly controversial. Many linguists continue to adhere to Osthoff and Brugman's principle (see e.g. Hill 2014: 211), which directly contradicts the possibility of lexical diffusion. Wang, and Labov (1981: 271), argued that borrowing between dialects is unsatisfactory as an alternative explanation for Wang's star example (about alternative reflexes in a Min dialect of the Middle Chinese departing tone), but their arguments were rebutted, convincingly to my mind, by Edwin Pulleyblank (1982: see esp. p. 405). Nevertheless, lexical diffusion has gained support among linguists studying non-Chinese languages (e.g. Krishnamurti 1978, Bybee 2002).<sup>4</sup> Søren Egerod (1982) argued that Wang's example does not work, but he nevertheless believed that lexical diffusion is a reality.

A high proportion of claimed cases of lexical diffusion in the literature relate to Chinese in various of its dialects. Conversely, the neogrammarian principle went largely unchallenged by Western linguists for about a hundred years after it was formulated.<sup>5</sup> The salience of Chinese in the lexical diffusion literature could merely indicate that Wang's writings have been specially influential among Chinese-language linguists. (See Hill 2016: 277–9 for a recent case where Chinese data seem to have been explained in terms of lexical diffusion although they are amenable to explanation in neogrammarian terms.) But, if incomplete diffusion of sound-changes through vocabulary really were more characteristic of Chinese than of European languages, that is, Chinese had more than European languages have of a mechanism which can potentially increase the number of phonological contrasts in a language, then the fact that Chinese in general and Mandarin in particular have nevertheless been characterized by a heavy *reduction* of contrasts would become all the more remarkable, a fact which any general theory of language change must take seriously.

And even if lexical diffusion were illusory, the contradiction between Wedel et al.'s evidence for homophony avoidance and the history of Chinese mergers would remain paradoxical.

### ***Statistical tendency versus absolute rule***

One can imagine various ways in which the apparent paradox might be resolved, but no way seems satisfactory.

A preliminary issue I need to mention, before discussing alternative resolutions, is that the published writings of recent believers in homophony avoidance are not always as explicit as they might be about their intellectual

assumptions. I have repeatedly found these assumptions made much more explicit in anonymous referees' reports on papers I have submitted. It is not usual to quote anonymous referees' reports in published research, but in the present case these are so revealing that I hope it will be allowable to do this, though whenever a point can be illustrated from published writings I shall do that by preference.

The line that has been most commonly taken, by believers in homophony avoidance, in response to my critiques is to point out that what Wedel et al. have put forward is a statistical tendency, not an absolute rule against high-yield mergers, and a single exception cannot refute a statistical tendency.

I find this response surprising, because it implies an odd interpretation of the homophony-avoidance theory. I had always taken this to be a theory about individual sound changes. But, if so, then what happened in Chinese was not a "single exception"; it was many exceptions, separate sound changes affecting different classes of sound and occurring at different times, and linked only in that each of them erased a contrast with high functional yield. Yet Abby Kaplan (2015) rebuts my 2013 article by writing "There is a statistical tendency for women to be shorter than men; one cannot argue against that claim by producing a single tall woman." And different anonymous referees have objected to my refuting Wedel et al. by presenting "a single counterexample" or "an alleged counterexample from Chinese", when I presented a list of these Chinese sound changes.

These sound changes would constitute a "single counterexample" only if we interpret the homophony-avoidance theory as a theory about languages as wholes, rather than about individual sound changes. That is not how the theory was understood by those originally responsible for it. The detailed illustrations of his thesis given by Martinet (1955), for instance, are all drawn from Indo-European data, yet Martinet does not argue that concepts like 'functional yield', 'push chain', 'drag chain' are relevant specifically to languages of the Indo-European family: he argues that they are relevant to language in general, and uses Indo-European languages to supply examples. Likewise Robert King (1967) does not use Germanic data to question whether functional yield is relevant to the history of the Germanic subfamily, but whether it is relevant to human language in general.

More important: homophony avoidance has appealed to linguists as widely as it has because, interpreted as a constraint on changes in any language, we can easily fit it in to what we know about human behaviour in general. People speak in order to communicate, so we can readily understand why people everywhere might tend to avoid developments that introduce ambiguity. It is far harder to imagine a reason why speakers of one language might regularly avoid this but speakers of another language not avoid it.

One anonymous referee writes: "It is ... possible that--for cultural or linguistic reasons that are not under statistical control in Wedel et al.'s models--a given language variety may be completely exempt from the pressure towards homophony avoidance." (Baerman 2011: 25 makes a similar suggestion, not quite so explicitly.) In the absence of any hint about what these "cultural or linguistic reasons" might be, this looks like empty handwaving. (I have considered the possibility that the unusual

nature of Chinese script could have made this language more open than others to homophony-increasing sound changes, Sampson 2015: 683, and argued that it does not stand up.)

### ***Chinese-specific paradox resolutions***

If one aims to resolve the paradox by reference to facts about Chinese rather than to general methodological principles, perhaps the most obvious suggestion would be that we are mistaken about the timing of the shift to disyllabic vocabulary. This is commonly taken to have occurred as a response to increasing ambiguity of syllables, but if the vocabulary became predominantly disyllabic before most of the mergers had happened then words would not have become ambiguous when the mergers did happen. And I am not sure that we have enough empirical evidence about the colloquial spoken language at early periods to be certain about the timing of the vocabulary change. But, if things happened that way, we would have solved one puzzle at the cost of creating another. Why would Chinese speakers have taken to replacing single words with two-root compounds while the single words remained unambiguous? In particular, why would they have taken to using so many synonym-compounds? So far as I know, on a world scale synonym compounding is a quite unusual word-formation mechanism, and if the separate roots are unambiguous then this mechanism seems to conflict with the usual assumption that people tend to economize effort. Surely we would be startled if we encountered a community of English-speakers who had taken to saying things like “It is timehour for me to gowalk homehouse”!

A variant of this solution would be to say (as Daniel Silverman has suggested to me--in personal communication, but see also Silverman 2015: 698) that of the two processes, phoneme mergers and vocabulary change, neither preceded the other but they went hand in hand. The suggestion seems to be that this removes the paradox, because there would never have been a stage when either ambiguity was created or people used unnecessarily verbose vocabulary--as fast as one thing threatened to occur, the other mechanism stepped in just sufficiently to avert it. As a hypothesis about the historical timing of Chinese language changes this might be very reasonable. The problem with it is that it makes the theory about avoidance of high-yield mergers unrefutable and empty. No hypothetical merger is predicted to be impossible or unlikely, because if it occurred it could always be claimed not to have been a high-yield merger at that point. And if the theory is empty, we are left without an explanation of how Wedel et al. achieved that high level of statistical significance.

A quite different kind of solution might be to suggest that at some point the colloquial spoken Chinese vocabulary became very small, so that mergers which reduced the range of possible syllables did not lead to ambiguity because the spoken lexicon was too small to overfill even that shrunken space of possibilities. Educated Chinese could communicate in writing, where phoneme mergers were irrelevant. Then, when new spoken vocabulary was coined, it would necessarily have had to be by means of compounding.

But I cannot take this idea seriously. In the first place, even the most basic,

everyday vocabulary which any language would need to contain often consists in Mandarin of synonym-compounds. Furthermore, while I am not a very ‘politically correct’ person I am sceptical about the idea that peasant societies manage with tiny vocabularies. Agricultural workers may not use terminology relating to abstruse intellectual disciplines, but they use plenty of agricultural terms which abstruse intellectuals are ignorant of.

I have discussed some other possible resolutions of the paradox in Sampson (2015), but I cannot find any which seems satisfactory. That paper was a ‘target article’ in an issue of the *Journal of Chinese Linguistics* which invited a number of scholars to comment on my claim that a paradox exists. Unfortunately, while the commentators made various interesting remarks around the topic, few of them can be said to have addressed the issue directly, by saying whether it is the homophony avoidance theory or the standard account of the history of Chinese which they believe to be mistaken, or by giving a reason to see the contradiction as apparent rather than real. (Abby Kaplan did do this, by raising the point about statistical tendency versus absolute rule which I dealt with above.) Some of the commentators in 2015 seemed to misunderstand the problem. When I discuss the growth of Chinese homophony with academics who are not knowledgeable about this language, they often say things like “But surely they solve that problem with their tones?”, as if tone contrasts had been introduced into Chinese as a device which prevented homophony from increasing. Wolfgang Behr’s comments on my target article seemed to associate himself with this idea (Behr 2015: sec. 2). But the fact that a number of segmental contrasts in Old Chinese have been replaced by tone-contour contrasts in the modern language does nothing to contradict the point that, overall, homophony has increased hugely. Syllables differing in tone are not homophones, but nevertheless there are very many homophones.

The journal issue contained many suggestions that the contradiction I had identified is illusory, but no serious explanation of why it is not real. I believe it is real.

### **Questionable data**

I take it to be common ground that a contradiction within the body of human knowledge is unacceptable. We cannot rest easy while one group of scholars believe X and another group believe not-X; the fact that one group may be working within Departments of Oriental Studies while the other group are working within General Linguistics departments may explain the persistence of the contradiction, but does nothing to make it legitimate. The question is what we do about this particular contradiction.

One tactic would be to ‘carry the battle into the enemy camp’, as it were, by querying the reliability of the data used by Wedel et al. to argue for homophony avoidance. Statistical findings can be only as reliable as the data from which they are extracted. Wedel et al.’s papers are less explicit about their language data than about statistical operations, but from what they do say it appears that in many respects the data are questionable or wrong.

WKJ and WJK both use data on phoneme mergers in nine languages--the same set of nine, except that where WJK uses Slovak, WKJ uses Turkish. The first language listed in each set of nine is “English (RP)”, that is English Received Pronunciation, the English accent regarded as standard in England and Wales (and to a lesser extent elsewhere in the British Isles); I shall focus my critique on this case. As their authority for statements about RP-related data, Wedel et al. cite the three-volume work Wells (1982). In my citations of passages in Wells I shall use roman numerals to identify volumes.

Wedel et al. proceed (for each language) by comparing the functional yield of phoneme pairs which contrast in the standard language in question, but which have recently merged (unconditionally, or in specified phonological environments) in some “otherwise phonemically similar dialect” (WKJ, p. 180), with the yield of phoneme pairs which have not merged.

An initial query (not the most important) would be how Wedel et al. have decided to treat particular dialects as related to “English (RP)” rather than to “English (American)”, their second language. For instance, one merger they list for RP is 3: ~ εə, that is the NURSE and SQUARE VOWELS (WKJ, pp. 181 and OS 3, misprinted in the former place as εɪ ~ εə). This merger is characteristic of Liverpool and nearby parts of Lancashire (Wells, 1982 ii, p. 372), and it is not obvious that the phonological structure of those dialects is closer to RP than to General American English. Wedel et al. may simply have chosen to group regional accents with “RP” rather than with “American English” when the regions in question are within Britain, but if so that would be misleading. RP is a notably un-conservative lect, with innovating features that are young relative to the settlement of North America; in terms of phonological structure regional British accents often seem closer to General American English than to RP. (For instance, many of them are rhotic, containing postvocalic /r/ in words like *car*, *part*, while RP is non-rhotic.)

More worrying questions relate to Wedel et al.’s decisions that particular mergers are or are not found in RP-like dialects, however these are defined.

For instance, Wedel et al. list a merger /θ ~ f/. Realizing the /θ/ phoneme as [f] is characteristic of London speech, but Wells (1982 ii, pp. 328–9) points out that this is not a phoneme merger: all mature speakers have /θ/ in their phoneme inventory and know which words it occurs in, though in some speech styles they may pronounce it as [f]. WJK (pp. 400, 410) argue that it is contrasts between lemmas rather than between surface forms which matter in connexion with functional yields, implying that they ought not to include /θ ~ f/ in their list of mergers. The Chinese mergers discussed in my opening section were absolute mergers: present-day Mandarin speakers have no awareness at any level of the contrasts which existed previously (unless they happen to be knowledgeable about other dialect(s)).

Another merger listed by Wedel et al. is /ð ~ z/. I know of no English lect which merges these phonemes, other than native speakers putting on a mock German or French accent. I have searched Wells’s three volumes without finding anything to justify this merger claim.



Conversely, /s ~ z/ is listed (WKJ OS, p. 3) as a phoneme pair which are never merged. Yet according to Wells (i, p. 180) these phonemes are merged, as [s], in some Celtic-influenced areas. (And it is a standard cliché that they have merged as [z] in much of the West Country, as in the well-known song which begins “Oh, we’ m come up from Zummerzet, where the zider apples grow”--though Wells, ii, p. 343, states some caveats about this latter merger.)

Thus it is not clear how Wedel et al. derived their lists of occurring and non-occurring mergers from Wells (1982), or how they obtained their data if they did so independently of that work, and one could question further individual entries. But a separate issue concerns Wedel et al.’s assumption that mergers are to be treated as events occurring between single phonemes. For instance, their list of mergers (WKJ OS, p. 3) includes, as two entries, /θ ~ t/ and /ð ~ d/. This appears to relate to Irish varieties of English, which are heard by English people as substituting /t d/ for /θ ð/. Wells (ii, pp. 428–9) says that in most environments the phonemes are not in fact merged: the interdental fricatives become dental stops, which continue to contrast with the alveolar stops /t d/. However, Wells adds that before /r/ even /t d/ in Irish speech become dentals, which is possibly what justifies Wedel et al.’s claim of merger (though Wedel et al. do not identify /θ ~ t/ and /ð ~ d/ as conditioned mergers, whereas in other cases they specify conditioning factors explicitly). But in any case, that is *one* sound change, not two: alveolar stops become dental irrespective of voice. Sound changes often do affect classes of sounds rather than single sounds, and there are many other examples in Wedel et al.’s lists. For instance, the five mergers they list for Korean are all between lax obstruents and their tense counterparts (obscured in their listing by accidental omission of the apostrophe marking tenseness in two of the five cases, /tʃ’ ~ tʃ/ and /s’ ~ s/). As a language development this is surely a single change: lax and tense obstruents merge irrespective of place of articulation.

This is not a pedantic nicety; it is crucial for computing functional yields. Treating the Korean tense ~ lax merger as five separate mergers, rather than one merger with a much higher total yield, gives artificial support to Wedel et al.’s hypothesis.

A problem which may be merely an unclarity in Wedel et al.’s exposition (but which is severe if real) relates to how they counted functional yields in the case of conditioned mergers. For instance, if /θ ð/ merge with /t d/ only before /r/, then presumably it is the functional yield of the contrasts just in that environment which might predict propensity to merge, not their yield over the whole vocabulary. Wedel et al. nowhere state that they counted functional yields for conditioned mergers separately.

A further problem which is certainly real relates to Wedel et al.’s choices of “unmerged pairs” to compare with the mergers they list. They say (WKJ, p. 180):

Because phonemes that merge tend to be phonetically similar ... we limited the comparison set of non-merged phoneme pairs to pairs which differ in only one phonological feature such as *voice* or *place of articulation*.

That might be a sensible decision, but the lists of “unmerged pairs” in their online supplement show that they did not follow it in practice. Thus, for RP, alongside pairs like /t ~ p/ (which differ only in place of articulation) and /t ~ d/ (differing only in voice) there are pairs like /v ~ tʃ/ (differing in place, manner, and voice).<sup>6</sup>

The problems I have exemplified are enough to cast severe doubt on the conclusions Wedel et al. draw with respect to the RP English case. Even if space permitted, I am not qualified to criticize their treatment of other languages in similar detail. But then, English is one of the world’s best documented languages. I question whether phonological descriptions on a par with Wells (1982) in terms of comprehensiveness and precision exist for all of the other languages used by Wedel et al. So there is little reason to expect their statistical data on those other languages to be more reliable than their figures for RP English.

### **Noisy data no problem?**

However, believers in homophony avoidance see objections like these as irrelevant. One anonymous referee, after acknowledging that my objections to Wedel et al.’s English-language data may be justified, wrote:

There is, however, an important point to note here. If these are honest mistakes by Wedel et al., they simply introduce further noise into their data, which would actually *weaken* any existing statistical patterns, not strengthen them. Such mistakes could only create the illusion of a pattern where there is none if the authors systematically underrepresented mergers that did not fit their prediction, and overrepresented mergers that did. Suggesting that they did so would amount to an accusation of academic malpractice.

In case it needs saying, I do not suspect Wedel et al. of intellectual dishonesty. It is true that finding a statistical significance effect in a data-set despite random errors in that data-set might be seen as making the finding more rather than less credible. But, in the first place, it is not true that systematic rather than random errors would imply dishonesty: confirmation bias is a powerful unconscious psychological force (see e.g. Kahneman 2011), from which *bona fide* academic research enjoys no special exemption. Furthermore, the high level of errors I have encountered in the area of Wedel et al.’s work which I am best qualified to assess, the phonology of English, makes it reasonable to suspect that with respect to other areas, including data on languages I know less well or not at all, and details of statistical techniques used, the research could be equally rocky. It would be quixotic to argue that yes, Sampson, you have correctly identified lots of mistakes in this research, so you really have to believe in its findings.

### **Potential falsifiers**

A scientific hypothesis is worth putting forward if it has ‘potential falsifiers’:

logically-possible observations which, should they actually be observed, would refute the hypothesis. A theory lacking potential falsifiers is empty. The central question I would put to believers in homophony avoidance is: accepting that Wedel et al.'s hypothesis is statistical rather than absolute, nevertheless if you do not regard the hypothesis as refuted by the occurrence of a series of separate sound changes which create so much homophony in a language that its vocabulary has to be almost entirely replaced (as happened in Chinese), then what would be enough to refute it? What makes the hypothesis non-empty?

I have found little attempt in the published literature to answer this question. One anonymous referee does suggest an answer, though: "A true counterexample to Wedel et al.'s findings should be a language where mergers with a high functional yield are *more* likely to happen than mergers with a lower functional yield."

That is a clear statement about a logically possible observation. But it makes Wedel et al.'s theory too weak to be of interest. I had supposed that the alternative to Wedel et al.'s theory is that homophony is simply irrelevant to the issue of whether particular sound changes occur. But, for the referee, that situation is compatible with Wedel et al.'s theory: the *only* thing they are ruling out is the possibility that a language might systematically prefer to adopt high-yield mergers. If that were truly all Wedel et al. were trying to say, then I am sure they are correct, but whoever would doubt it? Wedel et al. would be in the position of some hypothetical linguist who announces a universal finding that, say, no speakers of any language systematically prefer to use vocabulary sharing an initial with the current month-name, so that they utter more /m-/ words than usual in March, more /s-/ words in September, and so on. Most people would respond to that by asking "Who on earth imagined that speakers might do such a bizarre thing? Your 'universal' does technically have content, but it is too close to being empty to be publishable." Most of Wedel et al.'s readers have surely believed, rightly or wrongly, that they were arguing for a more interesting thesis than this.

This is not the only voice, though, which defends Wedel et al. against my critique by arguing in effect that their thesis is weaker than I suppose. Another anonymous referee objects to my describing Wedel et al.'s work as requiring to be taken seriously because of the very high levels of statistical significance they report: "Given that Wedel et al.'s arguments are couched in a frequentist statistical framework, 'level of significance' is not actually a meaningful concept. Any result that has a  $p$ -value under 0.05 is considered significant ... Obtaining a particularly low  $p$ -value does not strengthen the evidence for a given hypothesis".

I find no explicit statement in Wedel et al.'s writings that  $p < 0.05$  is the only probability threshold that interests them, and they do not explicitly discuss 'frequentist' versus alternative statistical frameworks. I am not qualified to discuss the latter issue, but it is clear to me that the rhetorical persuasiveness of Wedel et al.'s writings owes much to their repeated references to far lower  $p$ -values, e.g.  $p < 0.001$ --not just in WKJ and WJK but in other publications by members of the group, e.g. at several places on p. 663 of Kaplan (2011). Observing something that would be expected to happen by chance one in twenty times on average is not very remarkable,

but even to relatively statistically-naive readers like myself it seems obvious that a hypothesis which correctly predicts something that would happen by chance less than once in a thousand times is much more impressive. If Wedel et al. really only mean to claim that their findings achieve the  $p < 0.05$  threshold, then why did they not write “ $p < 0.05$ ” at each point where they actually wrote “ $p < 0.001$ ” and similar? If the referee I quoted is correct, this would have been less misleading. But it would also have given their work much less public impact.

### ***A call to arms***

To my mind, the defences being mounted to save the homophony-avoidance theory are often intellectually far-fetched. But at the same time there seems very little willingness to entertain the idea that the theory might be wrong. Its defenders are many, and so far as I know I am at present the only one seeking to reveal its flaws publicly.<sup>7</sup> I do not flatter myself that I can win this debate single-handedly. Unless others care to enter the lists, it will be an established truth of linguistics that languages tend to avoid high-yield sound changes, and Chinese historical linguists will risk being seen as resembling inventors designing perpetual motion machines.

Conversely, specialists in Chinese historical linguistics seem unmoved by the apparent incompatibility between their views and the theory of homophony avoidance (as witness the failure of the Sinologists among the respondents to my ‘target article’ to engage with the contradiction). It could be that many specialists in this language think along the lines “We know (to a reasonable approximation) what the facts of Chinese are; if linguistic theorists are making rash generalizations in ignorance of those facts, it is they who have dug themselves into a hole and it is up to them to clamber out of it.” Bernhard Karlgren, for many years the doyen of Western studies of Chinese historical phonology, was openly scornful of the idea that general linguistic principles could help shed light on the history of Chinese (see e.g. Karlgren 1954: 366–7).

One can sympathize with a feeling that the onus is on general linguistic theorists rather than Sinologists to resolve the contradiction. If it had been the English language which posed comparable difficulties for the homophony-avoidance theory, that theory would surely have been laughed out of court the moment anyone put it forward. If an abstract general theory turns out to be incompatible with particular concrete facts, I certainly feel instinctively that the first priority should be to reconsider the abstract theory, rather than to try to explain away the awkward facts.

Nevertheless, it would seem somewhat intellectually irresponsible for either group of scholars to pursue their researches while ignoring the fact that, according to beliefs which are well established among the folk in the next-door university department, the accounts which they are developing cannot be true. And happily Karlgren’s present-day successors, or some of them, do not share his hostility towards general linguistic theory. William Baxter (1992: 4) wrote: “We are on firmest ground ... when we reconstruct systems and changes which are well within the range of variation actually observed in human languages.” (Baxter has not to my knowledge

discussed the specific issue of homophony avoidance.)

Thus I hope that suitably-qualified scholars can be persuaded to join me in taking seriously the incompatibility of the theory of homophony avoidance and the established account of Chinese phonological history.

### References

- Baerman, M. (2011). Defectiveness and homophony avoidance. *Journal of Linguistics*, 47, 1–29.
- Baxter, W.H. (1992). *A handbook of Old Chinese phonology*. Berlin: Mouton de Gruyter.
- Behr, W. (2015). Comments on Sampson 2015. *Journal of Chinese Linguistics*, 43, 719–32.
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261–90.
- Chen, M. (1977). The time dimension: contributions toward a theory of sound change. In Wang 1977.
- Egerod, S. (1982). How not to split tones--the Chaozhou case. *方言* 1982 vol., issue 3, pp. 169–73.
- Hill, N.W. (2014). Grammatically conditioned sound change. *Language and Linguistics Compass*, 8, 211–29.
- Hill, N.W. (2016). A refutation of Song's (2014) explanation of the 'stop coda problem' in Old Chinese. *International Journal of Chinese Linguistics*, 3, 270–81.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin.
- Kaplan, A. (2011). How much homophony is normal? *Journal of Linguistics*, 47, 631–71.
- Kaplan, A. (2015). Reply to Sampson 2013. *Diachronica*, 32, 268–76.
- Karlgren, B. (1954). Compendium of phonetics in Ancient and Archaic Chinese. *Bulletin of the Museum of Far Eastern Antiquities*, 26, 211–367.
- King, R.D. (1967). Functional load and sound change. *Language*, 43, 831–52.
- Krishnamurti, Bh. (1978). Areal and lexical diffusion of sound change. *Language*, 54, 1–20.
- Labov, W. (1981). Resolving the Neogrammarian controversy. *Language*, 57, 267–308.
- Martinet, A. (1955). *Economie des changements phonétiques*. Bern: Francke.
- Osthoff, H. & Brugman, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, vol. 1. Leipzig: S. Hirzel.
- Pulleyblank, E.G. (1982). Review of Wang 1977. *Journal of Chinese Linguistics*, 10, 392–416.
- Sampson, G.R. (2013). A counterexample to homophony avoidance. *Diachronica*, 30, 579–91.
- Sampson, G.R. (2015). A Chinese phonological enigma. *Journal of Chinese Linguistics*, 43, 679–91 & 740–53. A version is reprinted in Sampson, *The linguistics delusion*, Sheffield and Connecticut: Equinox, 2017.
- Schuessler, A. (2009). *Minimal Old Chinese and Later Han Chinese: a companion to Grammata Serica Recensa*. Honolulu: University of Hawai'i Press.

- Silverman, D. (2015). Comments on Sampson 2015. *Journal of Chinese Linguistics*, 43, 697–702.
- Wang, W. S.-Y. (1969). Competing changes as a cause of residue. *Language*, 45, 9–25.
- Wang, W. S.-Y. (Ed.). (1977). *The lexicon in phonological change*. The Hague: Mouton.
- Wang, W. S.-Y. & Lien Chinfa (1993). Bidirectional diffusion in sound change. In C. Jones (Ed.), *Historical linguistics: problems and perspectives*. London: Longman.
- Wedel, A., Jackson, S., & Kaplan, A. (2013). Functional load and the lexicon: evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech*, 56, 395–417.
- Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: a corpus study. *Cognition*, 128, 179–86.
- Wells, J.C. (1982). *Accents of English* (3 vols). Cambridge: Cambridge University Press.

### **About the author**

Sampson studied Chinese and the history of the Chinese language at Cambridge and Yale Universities, before researching and teaching Linguistics at Oxford, LSE, Lancaster, and Leeds Universities and later Informatics at Sussex University. After becoming emeritus in 2009 he spent several years as a Linguistics research fellow at the University of South Africa. His books include *Love Songs of Early China* (2006), *Writing Systems* (2nd edn 2015), and *The Linguistics Delusion* (2017).

- 1 I am grateful to various scholars who have commented on work from which this paper emerges, including Nathan Hill, Abby Kaplan, Daniel Silverman, and Andrew Wedel. I apologize to anyone whose name I have overlooked, and of course I take full responsibility for any shortcomings in the paper.
- 2 One might object that sound-changes tend to apply to classes of phonemes rather than to single phonemes: a development affecting, say, all voiced fricatives would be more plausible than a change to just the single phoneme /ð/. But the functional-yield concept applies to any hypothetical merger, whether between single phonemes or classes--I gave a single-phoneme example just to keep the exposition simple.
- 3 Orthography and emphasis as in the original. The qualification *so weit er mechanisch vor sich geht* could be read as making Osthoff and Brugman's principle an empty one: a sound-change which has exceptions would simply be one which did not 'proceed mechanically'. But remarks on their next page, particularly their footnote 1, make clear that this would be a misreading.
- 4 For allusions to lexical diffusion earlier than Wang's 1969 paper see citations in Chen (1977: 214–15). Wang attempts to deal with criticisms of his theory in Wang and Lien (1993).
- 5 Beginning fifty years ago, a number of linguists have argued against the neogrammarian principle by claiming that sound laws are sometimes conditioned by grammatical rather than exclusively by phonetic factors. Nathan Hill (2014) has exhaustively defended the neogrammarian principle against these claims, but in any case this is a separate issue from that of lexical diffusion.
- 6 As received at my computer, the list even includes a pair comprising /f/ and a phoneme symbolized as /ŋ/ with a subscript syllabicity mark. Apart from the fact that these sounds would differ in five features (place, manner, nasality, voice, and syllabicity), English is not standardly analysed as containing a syllabic velar nasal. However, I suspect this symbol could be a mistake in electronic transmission for an intended /dʒ/, which differs from /f/ in only three features.
- 7 Since drafting this paper I have learned about relevant work under way at the University of Pennsylvania by the doctoral candidate Andrea Ceolin; see <[www.ling.upenn.edu/~ceolin/Ceolin2019.pdf](http://www.ling.upenn.edu/~ceolin/Ceolin2019.pdf)>, accessed 23 May 2019.