

Swimming against the current

Text of a “keynote address” for video delivery at the first World Congress of Korean Language, October 2020

신사 숙녀 여러분: 만나서 반갑습니다!

I want to talk today (in English, I’m afraid) about what I believe is a consistent tendency in the historical evolution of scripts around the world. I shan’t call it a “linguistic universal”, because I don’t believe in Chomskyan biological universals of language. Languages are purely cultural products, and like other cultural products their development isn’t controlled by scientific laws. But the tendency I have in mind does seem to hold quite widely, and it can be explained, in terms of the demographic changes we associate with modernity. The reason why this topic may make a suitable opening for your conference is that the one unexplained exception to the tendency which I know of is in fact Korean script. But I’ll come to the case of Korean after I’ve spent some time discussing the tendency more generally.

What I believe to be generally true, as I discuss in my book *The Linguistics Delusion*, is that in newly literate societies writing tends to be a matter of phonetic transcription – if the script is alphabetic, a newly-devised orthography will conform pretty closely to the principle “one sound, one symbol”; while as the script matures and the society modernizes, the orthography tends to evolve in the direction of what I’ll call “contrastiveness”, in the sense that the meaningful units of the language have consistent written shapes which are well-differentiated from the written forms of other meaningful units.

Looking at my own native language, English, it seems fair to say that when Old English became a written language in the middle of the first millennium A.D., its orthography was pretty well “phonemic”, as linguists say – whereas for a long time now, it has been far from phonemic.

Take one word as an example, the word *debt* – something that we owe. The word came into English from Old French in the 13c, and initially it was spelled in phonetically-rational ways: DET, DETT, DETE, DETTE. (Where a final E was written, a vowel may actually have been pronounced at the time.) But from the 16c onwards the standard spelling has been DEBT, with a letter B.

Some odd-looking English spellings can be explained in terms of English pronunciation changing over time while spelling has not changed to catch up with pronunciation – but that’s not the case with *debt*: the word never had a B sound in English. And some special spellings are seen as valuable because they distinguish homophones – the words MEET, for people to get together, and MEAT, edible flesh, sound exactly the same in modern English, but their different spellings (which reflect a phonemic contrast lost in the modern language) are useful in keeping the words apart on paper. But again this is irrelevant to *debt*: there is no other word pronounced that way. The way the letter B got into the spelling was that some French writers took to inserting B on occasion to show that they were educated enough to know that the word had had a B back in the days of Latin – though even in Latin it didn’t have B and T adjacent, there was a vowel in between. In French this habit didn’t last: the modern French word *dette* has no B. But in English, as it happens, this quirky occasional spelling habit became fixed when our spelling was standardized.

Odd-looking spellings will always have some historical explanation, comparable to this story: people don't just pluck irrational-looking spellings out of thin air; though in some cases the explanation is even more far-fetched than in the case of *debt*. The word *foreign* is spelled with -EIGN because, somewhere along the line, someone imagined that *foreign* shared an etymology with REIGN, for a king to rule. It never did: *foreign* ultimately comes from Latin *foraneus*, "out-of-doors", which had no G either in spelling or pronunciation. But the interesting question is not so much where did the B in *debt* or the G in *foreign* come from, as why were they kept? Why didn't people say, as it were, "Well, you're jolly clever to know about Latin, or you're jolly imaginative to link *foreign* with *reign*, but now we're fixing on one standard dictionary spelling for each word, we're going to keep things simple and write DET and FORAN"?

The answer, I believe, is that the phonetically-irrational spellings have an advantage of their own: they make the words visually distinctive, contrasting heavily with the spellings of all other words. So far as I can think, there is only one other word in the entire English vocabulary which ends in -BT, which is the word *doubt* – whereas, if *debt* were spelled DET, there are plenty of other words ending in -ET. *Debt* may be an extreme case, but in general I believe it's true that while modern English spelling is much less phonetic than Old English spelling was, it also tends to possess greater contrastiveness among the spellings of different words than existed when our orthography was young. And this development is not restricted to English. It is particularly marked in English, but then English has a particularly long history as a language used for serious written purposes – English was already a language of scholarship at a period when scholars in Continental Europe automatically wrote in Latin rather than in their own vernaculars. But one can see the same tendency in the history of other European orthographies, for instance French.

If one looks for a case of a European language whose orthography is pretty well perfectly phonemic, one of the best examples I know is Finnish, the language of Finland (or one of its languages – Swedish has equal official status in that country). It isn't a matter of chance, I believe, that Finnish spelling is thoroughly phonemic: it's also one of the youngest European orthographies. The Finnish language was not written at all before the 16c, when written English already had almost a millennium of history behind it, and Finnish has ranked as an official language only since the 1890s, after centuries when Finland was ruled by Swedes and later by Russians.

Finnish script stands at one extreme of the spectrum of possibilities between straight phonetic transcription, and maximally contrastive representation of meaning units. What's the other extreme? For that we have to look outside Europe to Chinese script, which is not phonetic at all but what linguists call "logographic": each word has its own written form, which reflects pronunciation only tenuously if at all, but which is often extremely visually distinct from the graphs for all other words, with huge diversity of visual component elements, dwarfing the level of contrastiveness that's possible with an alphabet of 26 letters. What I'm saying is that, on a world scale, the historical evolution of scripts tends to make them less like Finnish and more like Chinese. Writing comes to have less to do with reflecting sounds, and more to do with identifying meanings.

This isn't just true of spelling: we see it also in the history of European punctuation. When writers began to use punctuation marks, the hierarchy of comma, semicolon, colon, and full stop was thought of as marking silent pauses of different lengths – the logic of the word "semicolon", half-colon, was that its pause was supposed to be half the length of a colon pause. Of course, at the time people had no way of actually measuring pauses in the flow of speech, and I don't imagine this idea was ever objectively true; but nowadays no writer thinks of punctuation in those superficial physical

terms. Punctuation reflects aspects of the logical structure of language: a colon says that what follows will redeem some kind of promise implied in the wording before the colon, whereas a semicolon merely separates clauses which may have no relationship beyond sequence in time.

And the evolution away from Finnish-type towards Chinese-type script isn't just about making writings for different meanings look distinct from one another, but also about making the writing of any given meaning-unit consistent even when its pronunciation varies. The spelling of English words affected by the Great Vowel Shift is an obvious example. If we take the words *metre* and *metric*, both deriving from the root of Greek μέτρον, “measure”, we find that the vowel of the root comes out different ways in modern English: [i] in *metre*, [ɛ] in *metric*; and indeed the same root is pronounced a third way, as shwa, in a word like *telemetry*. In a thoroughly phonetic script, the spellings would be different and the shared meaning would be obscured. But our actual spelling system ignores these phonetic differences in favour of giving a common element of meaning an invariant visual shape, METR-; and the same thing is true in general for the hundreds, or thousands, of cases of phonetic alternation produced by the Great Vowel Shift.

People commonly suppose that these cases of phonetically-irrational spelling reflect nothing deeper than brute conservatism. Before the Great Vowel Shift, the vowels didn't alternate so the spellings were rational, after that sound-change applied the spellings were “wrong” and ideally ought to have been changed, but we just never got round to changing them and now we're stuck with the crazy spelling system. Many individuals and groups have advocated reforming English spelling – the best-known example undoubtedly being the playwright George Bernard Shaw, who actually left a large part of his fortune after his death to fund a competition to devise a whole new alphabet for English; and spelling reform provided much of the motivation for the academic study of phonetics, a subject which flourished particularly actively in Britain in the late 19c and early 20c. But society at large never saw the movement for spelling reform as much more than a harmless eccentricity, it's scarcely ever been taken seriously. (So far as I know, the most recent occasion when it received any governmental support was not in Britain nor in the USA but in Australia, perhaps surprisingly.) And although the shift along the Finnish-to-Chinese spectrum does often in the case of English orthography relate to spellings reflecting earlier states of pronunciation, I believe the fact that society accepts that shift, and is willing to live with phonetically-irrational orthography, has a deeper explanation than simple conservatism.

Indeed, in Korea you have a case where a very comparable shift demonstrably has nothing to do with conservatism, in the orthographic developments associated with the linguist Ju Si-gyeong whose life spanned the 19c and 20c – developments which I discuss in my book *Writing Systems*. As I understand it (and please forgive me if I blunder badly in speaking about a language that, alas, I know only a very little about), when Korean sound-laws cause roots to have varying pronunciations in different environments, older *hangul* usage reflected that by using varying shapes for the roots; but more recently, roots have been given a constant written shape, ignoring morphophonemic variation. For instance, the root for “price” or “value” is /gabs/, and with the case suffix /i/ you get /gabsi/, but before the consonant in the suffix /do/ the /s/ drops and you say /gabdo/. Traditionally, the words would have been spelled <gab si, gab do>; now, they are spelled <gabs qi, gabs do>. This move away from spelling phonetically towards invariant visual shapes for meaningful roots is quite comparable to the way that *metre* and *metric* are spelled with the same vowel, but in your Korean case the explanation cannot relate to conservatism. I am pretty sure, though please correct me if I'm mistaken, that the sound-law which leads to the -s of /gabs/ dropping before /-do/ was already in operation well before the 19c.

The true explanation for the tendency of the world's scripts to shift away from the Finnish end towards the Chinese end of the spectrum, I believe, has to do with different kinds of script having a different balance of advantages and disadvantages from different user perspectives.

For learners – people first acquiring the ability to read and write – it's unquestionable that Finnish-style phonetic transcription is easiest. Because English spelling is so unphonetic, during my lifetime there has been a controversy among the primary schoolteaching profession about whether literacy is best taught via the so-called “phonics” or “look-and-say” methods – look-and-say treats each word as a single visual *Gestalt*, as if it were a Chinese graph to be learned as a unit, whereas phonics teaches learners to break spoken words down into phonemes and link the phonemes to the letters of the spelling. Even though these linkages don't always work in English, by the beginning of the 21c this controversy was decisively settled in favour of phonics: there's a wealth of hard empirical evidence showing that phonics is the best way of teaching literacy even in English.

But after individuals have become literate, once they are skilled readers of a script, for them the pros and cons change. They no longer have a reason to break words down into their phonemes: they know how words are spelled, and they may well recognize words as visual *Gestalts*. They want to “read for meaning”: to absorb the sense of a document as efficiently as possible, irrespective of precisely how it would sound if read aloud. For skilled readers, an orthography which makes meaning-units as distinct and contrastive as possible is best.

This is more than just a theoretical idea – there's plenty of hard evidence backing it up, as I discuss in the 2015 edition of my *Writing Systems* book. For instance, an article from the year 2000 by Perea and Rosa showed that for a word to be an “orthographic neighbour” of a different, high-frequency word – that is, to have a near-identical spelling – has an inhibiting effect on reading, at least among skilled readers: people read more efficiently if words are spelled distinctively, irrespective of their pronunciations. Conversely, though, a 2008 article by Duñabeita and Vidal-Abarca showed that the opposite is true for early-years schooling: reading at that stage is helped if the reader can make analogies between the spelling he's looking at and spellings of other words he's familiar with. And, remarkably, a 2012 article by Zhao et al. shows a very comparable effect for reading Chinese. Young Chinese children identify graphs more efficiently if they share their phonetic elements with many other graphs, but for older readers that's actually a negative factor – with the crossover coming at about age eleven.

Incidentally, in the West discussion of different types of orthography has been hindered by the fact that no European language uses a logographic script, and in consequence Westerners have taken Chinese writing to be a strange and almost impossibly difficult system. In the 20c there were serious scholars who ought to have known better, who wrote as if the Chinese people were struggling with a horribly cumbersome script, which was limiting the spread of literacy and holding back their economic development. Of course they didn't guess what was going to happen with the Chinese economy in the 21c, none of us did. In the 20c there were no objective, internationally-comparable measures of educational success. But nowadays we have the three-yearly Programme for International Student Assessment, PISA, and it's been giving us a pretty consistent story. In the latest PISA results, for instance, from 2018, the four leading countries for 15-year-olds' reading ability were the People's Republic of China, Singapore, Macao, and Hong Kong – the only Chinese-speaking country not in the top five was Taiwan. And the previous round of PISA tests had given very similar results. I know there are large problems of comparability in carrying out PISA-style

international assessments, but these results are so striking that they make nonsense of Westerners' traditional Eurocentrism about alphabetic script.

And a conflict of interests like the one between learners and skilled users of a script probably applies also to literate people in their role as writers versus their role as readers. Even literate English-speaking adults are often momentarily unsure of details of spelling, and have to think about how to write a word; if the spelling is unusual in some way, that might make it harder to retrieve every detail of the correct spelling. But the reader only needs to recognize which word is intended – if the spelling is unusual, that helps to identify the word, and the fact that the reader might not remember every detail of the spelling of a long word doesn't matter – nobody's asking him to.

As the psycholinguist Uta Frith summed the situation up in 1980, “the ideal orthography for spelling is incompatible with the ideal orthography for reading”. One might gloss this by saying that, for learners, Finnish-type scripts are most favourable; but for skilled readers, something closer to Chinese-type script works better.

If all this is right, then we shouldn't be surprised that in modern times scripts have tended in practice to shift to some extent further from the Finnish-style extreme and closer to the Chinese-style end of the spectrum. When a society is newly literate, at first only a minority learn to read and write (in the Middle Ages, I believe even our kings were not always literate), and those who do acquire these skills often do so as adults. In modern societies, almost everyone learns to read and write in primary school. Furthermore, life expectancy has become much longer under modernity. For both these reasons, the period when individuals are skilled readers is typically very much longer, relative to the period while they are acquiring literacy, than it was in pre-modern societies. So extra effort needed to acquire literacy becomes a worthwhile investment, because it's rewarded by increased reading efficiency over a relatively long period. Commonly, people who think about these issues are over-influenced by witnessing the very visible problems children have in learning to read. We see our sons and daughters, or in my case our grandchildren, struggling with unphonetic English spellings, and we naturally think “Wouldn't it be better if our spelling was simpler, and didn't put these stumbling-blocks in children's paths?” On the other hand the relative efficiency of the adult reading process is not immediately open to observation in the same way. It can only be examined through obscure psycholinguistic experimentation; so quite naturally people fail to place it in the balance against children's struggles. But modern life is so heavily dependent on documentation that efficiency of skilled reading deserves to be assigned a heavy weight in deciding what kind of script best suits a society overall.

And another difference between pre-modern and modern societies is that acts of reading have become far more frequent relative to acts of writing. In the days before printing, when all documents were individually hand-written, it wasn't possible for any single physical document to receive a wide circulation; if something needed to be read by many different people, perhaps a prayer-book or the Bible, it had to be laboriously copied out by scribes many times. But with printing, and even more so now with the Internet, it's quite possible and in fact commonplace for a single document, written just once, to be read by many thousands of readers. So properties of a script which make it challenging for a writer may be worth accepting, if they make reading more efficient: the extra effort by a writer on one occasion will buy extra efficiency for readers not just once or a handful of times, but very many times.

Both these differences between pre-modern and modern life push scripts in the same direction: they push scripts further away from the Finnish end of the spectrum, closer to the Chinese end. And in the history of English as a written language, that's indeed what we find. I'm not suggesting, of course, that this development resulted from any conscious decision by English-speakers to embrace the advantages of phonetically-irrational but contrastive orthography: the pressures which led to it were unconscious, but they were real nevertheless. It was the kind of case which the economist Friedrich Hayek saw as characteristic of successful cultural evolution: "the result of human action, but not of human design".

Now there's an obvious objection here, which has probably already occurred to many people in my audience. The kind of orthographic shift I'm discussing has happened in English, and in some other alphabetically-written languages such as French; but it hasn't happened universally. Finnish is a rather newly-written language, but there are languages with longer written histories whose orthographies remain pretty well phonemic. Spanish is a major world language with a very rational phonemic orthography. It may not have been in heavy use as a written language quite as long as English, but it has certainly been so for a very long time: so why hasn't its orthography developed in a similar way to that of English?

The answer there, I believe, has to do with the point I made earlier: people don't just pluck orthographic complications out of thin air. Even if there would be advantages in adopting a range of different spellings for the same sound, people won't do it unless they happen to be exposed to a range of alternative spelling conventions. English had that exposure from an early date. It's a Germanic language, but much of its learned vocabulary was borrowed from Latin. It was spoken in a country where the official language for several centuries after the Norman Conquest was French. When printing was invented, the earliest printers in Britain were largely Dutchmen. All these factors exposed English-speakers to diverse spelling conventions: they provided material that was available for creating the phonetically-inconsistent but relatively highly contrastive orthography that became standard. Spanish, on the other hand, is a language which evolved out of Latin, and (although I'm no Hispanicist) I'm not aware that Spanish-speakers were ever exposed in the same way to a variety of alternative conventions.

So I'm not saying that all orthographies are sure to become less phonetic and more contrastive as their society modernizes: whether an orthography has the chance to become more contrastive depends on the accidents of history. But what one wouldn't expect to see is an orthography moving in the reverse direction, becoming less contrastive and more phonetically rational than it was at an earlier period.

I'm aware of just two clear counterexamples to that prediction, but one of them doesn't really count. The case that doesn't count is Vietnamese. Vietnamese used to be written with a script which was 100% logographic: the large majority of its vocabulary which derives from Chinese was written with Chinese graphs, and the minority of native Vietnamese words were written with graphs created for the purpose on the Chinese model. In modern times, by contrast, Vietnamese is written with an alphabetic script that's not perfectly phonemic, but fairly close. It represents a large shift away from the Chinese end towards the Finnish end of the spectrum. But there is a special reason for that: Vietnam was a French colony from the second half of the 19c till the Second World War. I believe the French administrators imposed alphabetic writing because they themselves couldn't cope with logographic script. I'm not sure the Vietnamese had any control over the change.

That leaves the one case I know of where a society which controls its own orthography has shifted it to a significant extent away from the Chinese and towards the Finnish end of the spectrum: and that's my audience's society, Korea. Again correct me if I am mistaken about details, but as I understand it Korean was traditionally written in a mixed script, with the large number of words derived from Chinese roots written with their Chinese graphs, *hanja* as you call them, while the inflexional endings and little grammar words, and I suppose the native vocabulary not borrowed from Chinese, were written in *hangul*. Still when I visited Korea in the 1980s I saw a fair amount of *hanja* in use, though I believe not nearly as much as there would have been a generation or so earlier. But since then, as I understand it, the movement has been in the direction of eliminating *hanja*, so that many or most categories of document now use pure *hangul* with no *hanja* at all. Korean script has apparently been swimming against the current that I see affecting the other languages of the world.

To me as an outsider, this exception is specially surprising, for two reasons. One is the issue of homophones. I said earlier that the advantages of orthographic contrastiveness for readers are not just about distinguishing homophones. It helps rapid word identification that *debt* ends in BT, even though there's no other word pronounced the same way and written without the B. But if there are homophones, then non-phonetic, contrastive writing is particularly helpful to the reader – if spelling is perfectly phonemic, the only way a reader can distinguish one homophone from another is by drawing inferences from the context, and that takes mental effort which isn't needed if each word has a unique written shape. For European languages, homophones are a marginal issue. But you know better than I do that for a language like Korean whose vocabulary is largely derived from Chinese, homophones are a huge issue. The very words *hangul* and *hanja* illustrate the point: the same syllable *han*, pronounced the same way, means “Korean” in *hangul* but in *hanja* it means “Chinese”. I really wonder how law, for instance, works in a script which depends on context to disambiguate the meanings of words in statutes.

And the other special factor which makes the move away from *hanja* surprising to me is that, compared to roman script, *hangul* seems to be a script with relatively low visual contrastiveness. The 26 roman letters clearly can't include the high level of visual diversity that we find in Chinese script, but there is a reasonable degree of variety. Some lower-case letters have ascenders, some have descenders, some have neither. Some are open, some like p or d have one closed loop, the letter g has two. And so forth. Not all alphabetic scripts have this degree of visual diversity: the Hebrew and Arabic alphabets are notable for having much less. And to my eyes *hangul* script also has relatively few unique visual elements for the eye to fasten on: it consists largely of horizontal and vertical lines, right-angles, and one circle. One might have expected that mixing *hangul* with *hanja* would be specially advantageous for Korean readers.

So the recent evolution of Korean script constitutes a particularly emphatic counterexample to what otherwise is the general direction of script evolution under the conditions of modernity. Yet whatever the explanation may be, it surely cannot be that there is anything un-modern about South Korea. Seen from Britain, your country appears to be a highly modern society, with notably greater penetration of new technology, higher broadband speeds, and so forth, than we in Britain are accustomed to.

As I understand it, the vocabulary of science and technology in modern Korean is largely borrowed from English rather than from Chinese, and since English hasn't got a high incidence of homophones, the homophony problem wouldn't apply much there. But science and technology are

only one specialized area within the life of a society. The soul of any nation resides in discourse about current affairs, creative literature, discussion of politics, historical references, religious and philosophical ideas; and (again as I understand it) Korean vocabulary in these areas remains largely Chinese-derived. So it isn't plausible to see the use of English-derived vocabulary for science and technology as a factor which exempts Korea from the pressure on modern societies to evolve reader-friendly scripts.

As an outsider to Korean society, I am left puzzled. I have no explanation for why Korea has chosen to "swim against the current", but then, not being familiar with your society, it is not for me to explain it and I shouldn't expect to be able to do so. But it's surely a question worth thinking about. I hope some of my audience may see answers, where I can only wonder.

And on that note, I shall end this virtual visit to Korea, and wish you all:

신사 숙녀 여러분, 안녕히 계세요 !