Phonemic Depletion and the Origin of Speech:

How Humans Won the Language Lottery

By Simon Prentis

As part of the talk I gave this summer at the IJET conference in York, there was a brief discussion about the number of phonemes there are in Japanese and English. It's not something most people give much thought to most of the time, but it was interesting to note that even a group of translators and interpreters seemed quite unsure as to just how many sounds they actually use when speaking. When I asked people what they thought, the suggested answers ranged from 12 to 150—which, oddly enough, is quite close to the known spectrum for the numbers of phonemes used in human languages. What's more interesting, though, is the question of why there should be such a difference anyway: why should English have 44 phonemes, for example, and Japanese be able to make do with only 22? It turns out that this is a rather neglected corner of linguistic research, but at the risk of taking the Bulletin on a somewhat unorthodox off-piste excursion into the mechanics of language, I'd like to explore the issues that arise. I invite you to buckle up your seat belts and join me.

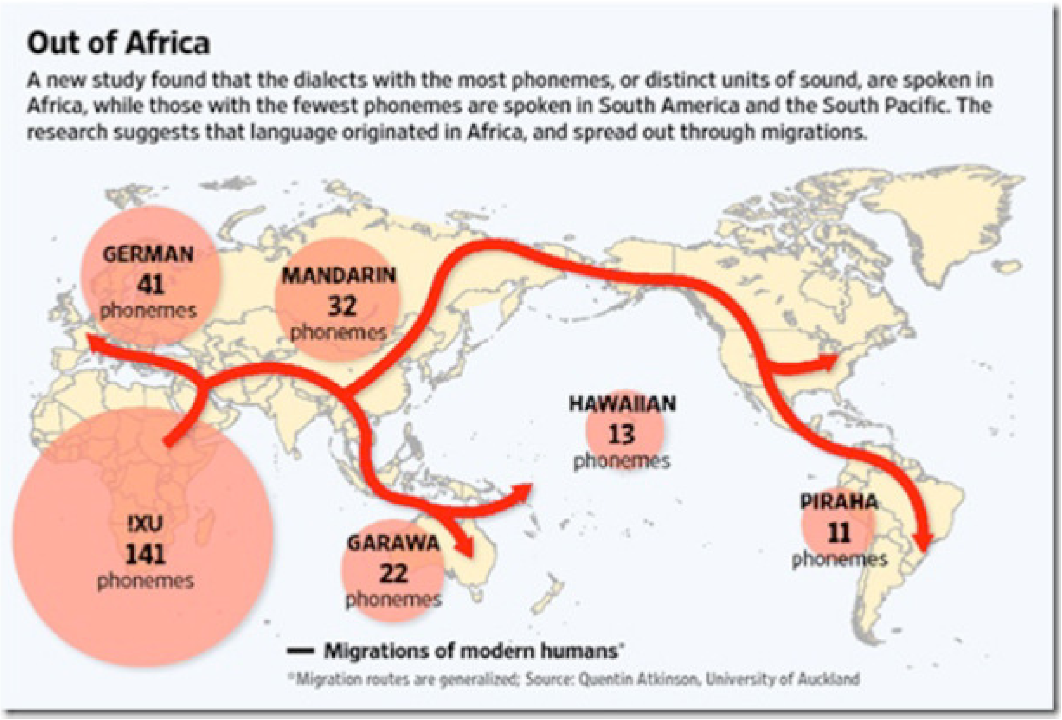
It's perhaps not entirely surprising that the IJET audience should have been so uncertain, despite being professionally engaged with these languages on a daily basis: just as for monoglot civilians, knowledge of deeper linguistic detail such as the precise number of phonemes in a language, or exactly how the rules of grammar work, is not something we necessarily need to be conscious of in order to do our work—any more than you need to have an understanding of cell biology (or even any biology at all) to be able to pass muster as a regular human being. And whether you speak many languages or only one, the fact is that given enough time you can still achieve adult competence with little more than a quite rudimentary understanding of the inner workings of the language you are speaking. That is especially so with our 'native' language, which we all learn without seeming to make any conscious effort at all.

In that sense, learning to talk is like learning to walk —which we have all managed to do by the time we are two. But talking is really not like walking at all. We all end up walking in much the same way, but the way we talk depends entirely on where we grew up. By the time we go to school, each one of us will have mastered the basic grammar of one or more of the over 6,000 languages still spoken around the world, but the way we manage to do it still remains an almost complete mystery. Traditionally, this mysterious capability has been explained as a gift from God, or the gods—for Christianity is not the only tradition that speaks of the divine origin of language. Without any other way to explain it, that has clearly seemed a plausible enough account for most of history, and makes almost as much sense as the only other "explanation" on the block, Chomsky's still widely-accepted theory that a chance genetic mutation has left us with what he calls a Language Acquisition Device, a putative accident of cerebral biology that supposedly somehow, suddenly, enabled us to speak.

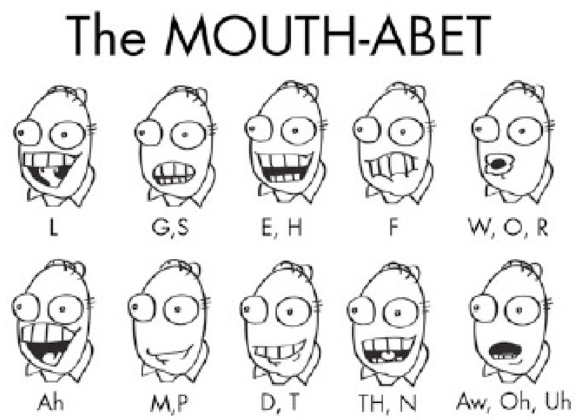
But both of these explanations contradict everything we know about evolutionary biology. Language cannot "suddenly" have appeared out of nowhere, any more than the elephant's trunk did—it must represent an evolution from an earlier, more primitive form of communication. And despite the remarkable ingenuity of generative grammar theory, the more practical experience of linguists in the field is slowly casting doubt on the idea of an inbuilt Universal Grammar—in favour of a more pragmatic understanding of language as a largely random expression of our consciousness, jerry-rigged to fit pre-existing cognitive patterns that we mostly share with our animal cousins. In this view, language is an adaptive means, simply a mechanism that allows us to tag our innate biological perceptions and enable our brains to communicate with each other. And a closer look at the fate of the humble phoneme offers a way to understand how that process might have got under way.

There is still a fierce debate about the origins of language, both about when and how it could have come into being. Until relatively recently, historical linguistics took the view that nothing useful could be said about the early history of language due to the absence of physical evidence. But though we may not be able to say anything very meaningful in morphological terms, there is a marker that provides an interesting clue as to how language is likely to have evolved. In a paper published in the magazine *Science* in 2011, Quentin Atkinson, a researcher from the University of Auckland, showed how a map of phonemic diversity could be used to track the origin of language—which he placed in Sub-Saharan Africa. Though this largely agrees with the archaeological evidence, his notion of a "serial founder effect" to explain the loss of phonemes attracted a storm of controversy from linguists—mainly on the grounds that different methods of classifying and counting phonemes produce different results, and that the method is not sufficiently rigorous.

What does not seem to be in dispute, however, is that as a general tendency—whatever the precise details of a particular language—there is a discernible drop-off in the number of phonemes used by human languages over time: in other words, the younger a language (measured broadly by its distance from Africa) the fewer phonemes it uses. This is clearly the case with Japanese and English, but the line can be followed through with other languages—consider Chinese (32), Korean (27), Hawaiian (13) and Pirahã (11)—whereas at the other end of the scale the Xoisan group of click languages, thought to be some of the oldest languages still in use, can have as many as 141 phonemes. And although it's well established that the defining feature of human language is the use of discrete units of sound which are combined to form words, the question of how this system might have originated, or why the number of such phonemes should deplete, has yet to be adequately addressed.



In considering these matters we must first understand how the maths stacks up. If we think of spoken words as being mainly composed of syllables, the number of syllables any set of phonemes can theoretically generate is given very approximately by [the number of consonants] x [the number of vowels]. In the case of English that's around 480, and for Japanese it's 85 (although not all possible syllables are actually used in a language: the standard Japanese syllabary uses only 50). In the case of the language with 141 phonemes, the number of potential candidates is over 4,500. Yes, that's 4,500, not 450. Why? Simply because that's what happens with as many as 92 distinctive consonants and 49 vowel sounds available. So even with just one syllable, that's already enough potential words to have a meaningful conversation—and unsurprisingly, much of the core vocabulary of click languages is monosyllabic. But language also allows for the combination of syllables to make words. The magic of multiplication then means that with just two syllables, the potential for English is around 230,000 words. For Japanese it is still only 7,225, but with one more syllable that number then rises to over 600,000—nearly 30 times as many words as all but the most educated of speakers usually know, in any language. Even Pirahã, with its tiny inventory of 11 phonemes, has the potential to generate over 300,000 words with just four syllables, and it's no surprise that the language is polysyllabic.



What does this tell us? First of all, it shows that a large inventory of phonemes is not required for a language to function properly. That means most languages are over-engineered in phonetic terms, making it easier for phonemes to drop away as new language groups form—in much the same way that grammar simplifies over time—and that is the phenomenon we actually observe. But if we then imaginatively rewind that process back through time, simple logic points us to a rather more interesting conclusion: language must have first evolved from a much larger inventory of sounds. And with that in mind, the suspicion must also be that the over-abundance of phonemes in human languages is really just a relic of a time when such otherwise meaningless noises probably did have a meaning, and were used for communication—just as some other species, notably monkeys, are able to differentiate discrete vocal calls today, using different cries to signal different threats, for example.

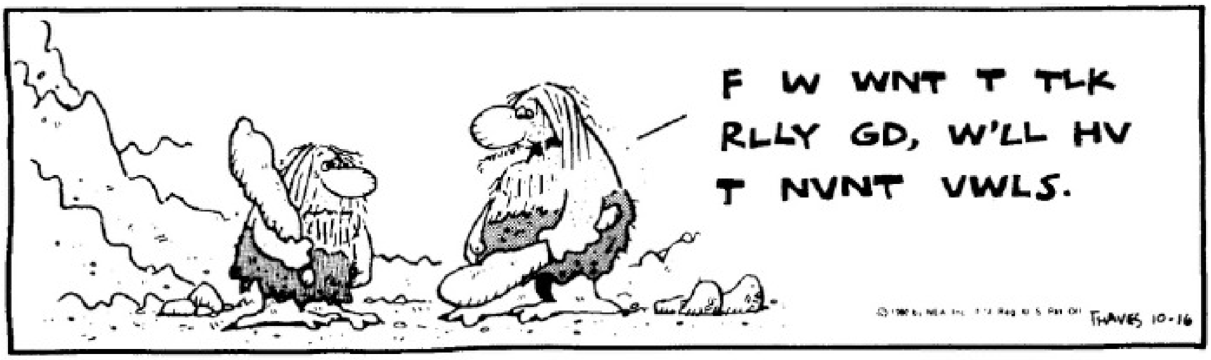
The crucial difference between such animal communication and human language is simply that we have stumbled onto a way to organize these animal calls, which are essentially random noises that have become associated with specific meanings. Their phonology may be fixed, but apart from that there is no discernible system, and they must be uttered and learnt as one-off analogue noises. But there's a problem with that—the limit of memory for random, unrelated things is quite small, even with humans. Imagine trying to remember all the numbers up to 100 if there were no system in place, and each number had a discrete, unrelated noise attached to it instead of recursively linking digits in units of ten—which is how we are able to remember a number like 178,654 for example. Without a system, you couldn't do it. But in learning to combine these calls, and using a restricted number of them to form new words in combination—in other words, by digitizing existing analogue call noises, isolating them as a limited set of discrete sounds which can then be combined and easily remembered... we have unlocked the door to language.

Is there any evidence that other species are doing this? Remarkably, there is. The clearest evidence to date is in the calls used by the aptly named babbler bird, but several species of monkeys, notably putty-nosed monkeys and Campbell's monkeys, have been found to concatenate, or combine individual sounds used to mean one thing on their own into combinations that then mean something else. Researchers studying these species have used edited playback calls to confirm that these combined sounds are indeed being used digitally to convey abstract meaning in the same way as humans do with language. This means that, far from needing to posit the miraculous emergence of a Language Acquisition Device to explain human language, the jump from the existing communicative behaviour of animals to the languages that humans use today can be understood to have begun as a relatively small evolutionary step: all that is required is that you restrict the number of existing analogue calls to a limited digital set of discrete sounds—phonemes—and you have a potential repertoire of thousands of easily memorable words on your hands.

How long ago might this have happened? Of course, we still can't say for sure. The disputed study by Quentin Atkinson previously cited quotes a figure of around 100,000 years ago, and some researchers believe that we may have had some form of proto-language for as far back as 500,000 years. Given how slowly evolution moves, that does not seem unreasonable. The habit of combining sounds into words may have taken a long time to bed in. But although purists may complain that there is still no actual evidence for language prior to 10,000 years ago, there are two well-documented events in the archaeological record that seem to strongly imply its existence: first, the emergence of burial practice (dating from around 100,000 years ago), and then the so-called great leap forward—the advent, around 50-60,000 years ago, of what archaeologists term behavioural modernity. Other animal species do not bury their dead, and it's hard to imagine how the kind of funerary routines practiced by early humans could be explained and propagated without language, in however rudimentary a form—and harder still to explain the sudden jump in the diversity and complexity of our behaviour some fifty millennia later.

The dramatic increase in the sophistication of tools and other technology around the time that humans first started to leave Africa in large numbers is well known, a change that was apparently unaccompanied by any obvious anatomical changes. In other words, our mental software appears to have been suddenly upgraded without any change to our cranial wetware. The obvious suspect here is language, which in allowing individuals to pass on useful information, knowledge and experience to others is more than sufficient to account for a sudden jump in collective awareness. But there is likely to have been a two-stage process involved. Even assuming that the digitization of existing communicative sounds was what enabled the creation of an open-ended (because systematized and memorable) vocabulary, knowledge of lexical items alone does not allow much communication to take place, as anyone who has struggled to learn a foreign language will confirm; without the ability to use grammar effectively there are still serious limitations on what can be conveyed.

Nor is grammar likely to have developed simultaneously with the growth of words—it would have taken time to figure out and agree a system of function words that allow the relationships between vocabulary items to be expressed. We can see young children struggling with this process, even though they have fully-developed models to follow in the adults around them. And function words are the key to effective grammar: though they only make up about 2% of our active vocabulary, they account for over half of what we actually say, because without them an unstructured succession of nouns and verbs is hardly more useful than a shopping list. We know it took long enough—millions of years, even—for us to progress from the first use of simple tools by homo habilis to the more sophisticated flint-flaking skills of homo erectus, so if we allow that words were beginning to come on stream some 100,000 years ago, it might easily have taken another 50,000 years to



flesh out a functional grammar. But once we had that, the world would have been our oyster.

Until we find a way to travel back in time, this must remain speculation—but just as our new-found ability to track back the evolution of genes has added greatly to our knowledge of the history of our species, revealing clear evidence of population bottlenecks and clues to the speed and timing of the spread of populations across the globe, so language has more to tell us than we can ever hope to glean from mere physical remains. Philologists have had great success in prising the history of language from written records; but just like our genomes, living spoken language also holds the history of its past, hidden in the very sounds that we use to produce it. And if we worry that the chain of evidence is too indirect, as back-up we have only to look at the evolution of writing to see a remarkably similar story in the emergence of a digital means of communication from primitive analogue antecedents, this time with the incontrovertible evidence that all known early forms of writing have independently evolved from analogue pictograms into digital morphemes, be they letters of the alphabet or the radicals of characters.

For an even more radical analogy, we can perhaps step further back and realise that digitization is nature's signature means of storing and conveying information, right down to the very structure of matter ('quantum' theory is at bottom describing the discrete, indivisible—and thus digital—units that underlie our reality)—not forgetting the DNA helix itself, a base-four digital system which encodes the language of life using four nucleotide 'digits' to create three-letter words that spell out the templates for proteins—which then fold into the grammar and extended ongoing narrative of our entire biochemical life. Given that trajectory, we should not be surprised to understand that our brains eventually came up with a digital means of organising and transmitting consciousness, even though it took us long enough to get there.