Darwin and language diversity
Conservation and diversification

Rissa:
Pagel, Atkinson and their colleagues have been applying their models to studying the evolution of languages, particularly from the Indo-European family. What are the key questions that interest them? Quentin Atkinson.

Quentin Atkinson:
The first is how the languages in the world today are related, how different language families are related, where did they come from and when did they arise? A second question is if we can characterise how languages are related, we might be able to also start studying the process by which they change.

Mark Pagel:
One of the things we can do when we've got a phylogenetic tree of languages is that we can begin to ask how do various aspects of language evolve along that tree.

If we look at individual elements of language on that tree, we see that some of the elements evolve very, very rapidly and other elements evolve very, very slowly,

So for example the word that Indo-European language speakers use to describe two objects is the same sound across all of the Indo-European languages and so the Spanish say 'dos'; the French say 'deux', the Italian say 'due' the German say 'zwei': all of these sounds are what linguists call homologous and what Darwin would have recognised that as they are sounds that all evolve from a common ancestor.

Cecile:
Un, deux, trois, quatre, cinq …

Joseph:
Un, dos, tres, quatre, cinc ...

Becky:
Uno, due, tre, quattro, cinque ...

Martha:
Uno, dos, tres, cuatro, cinco ...

Jim Donohue:
Een, twee, drie, vier ...

Ursula:
Eins, zwei, drei, vier, fünf ...

Malihe:
Yek, doe, se, char, panj.

Mark Pagel:
But the puzzle is that that other words in the same Indo-European languages evolve very, very rapidly, so, for example, what English speakers call a bird the Italians call 'ucello', the Spanish call 'pajaro' the Germans call 'vogel' and Caesar back in Roman times would have said 'avis' and so here we have a concept, 'bird', that has acquired a whole lot of different words.

Quentin Atkinson:
So the first question we were interested in was how quickly are these words changing and we were able to model that process on these language family trees, and quantify the rate of change. And we calculated what we call a word half-life, which is the time after which there's a fifty percent chance that the word has changed to something else. The shortest half-lives for the most rapidly evolving words were about 4 or 5 hundred years; the longest half-lives were of the order of 70 thousand years, so these words really don't change very much at all.

Rissa:
This degree of conservation for some words is remarkable when you consider that languages are evolving simultaneously and independently of each other.

Mark Pagel:
When we study languages we have to bear in mind that each of these languages is existing independently and evolving through time, and so if we have ten different languages each of which has been around for one thousand years what we really have is ten thousand years of language evolution. These are language years, and if a single word has been used by all of those language speakers in all of those different languages for those years that represents ten thousand language years that that word has been conserved.

Rissa:
But why are some words so highly conserved while others seem free to evolve very rapidly? What's the basis for this variation?

Mark Pagel:
Drawing an analogy from biological evolution we hit on the idea of a word's 'expression level'. How often is that word used in everyday speech? In biological evolution sometimes how often a gene is expressed in your body, how often it's used in your body is related to how rapidly it evolves. So we tried to investigate a word's expression level, and we were able to find information on the frequency with which words are used in everyday speech.

Quentin Atkinson:
We chose four languages: English, Greek, Spanish and Russian as examples from across the Indo-European language family, and these were high quality data sets with millions of conversations, and what we found was the more frequently used words changed much more slowly than the less frequently used words.

Mark Pagel:
Right across the board, in the Indo-European languages, words that are used frequently in everyday speech are highly conserved; they tend not to change. So words like two objects that we describe or three objects that we describe or a pronoun like 'I' or 'who' or 'what', these words all evolve rather slowly and they come up in our speech over and over and over .But words like 'bird' or 'belly' or 'dirty', these words that we don't use so frequently in our everyday speech, they seem to evolve rather rapidly.

Rissa:
In some ways the result was a surprise: one might argue that words that we use more often might have more opportunities to change. But to explain why frequently-used words evolved more slowly, the group came up with two possible explanations.

Mark Pagel:
One possible explanation is that the more we say a word, the better we get at saying it. So when we use the word we're less likely to make a mistake. We introduce fewer mutations into that word, if you want to use a genetic analogy. So it could just be that there are fewer production errors in words that we use frequently because we're better at using those words. Another possibility was that if we use a frequently used word incorrectly, we will hear somebody else using it correctly, rather soon and we'll correct our mistake. And so in some sense our mistake will not be allowed to propagate because others will be using the word correctly, whereas with an infrequently used word, we can use it incorrectly and that mistake can propagate for a long time before somebody corrects it.

Rissa:
But could it be that how often a word is used is simply a feature of what part of speech it is, that is, whether it's a noun such as 'dog' or 'cat' or a verb such as 'run' or 'walk.' So Pagel and his colleagues took this into account when they performed their analysis. They found that for any part of speech like a noun or a verb, the more often it's used, the more slowly it evolved. But overall, what part of speech a word corresponds to, does affect how quickly it evolves.

Quentin Atkinson:
What we found was that numbers evolve most slowly, given how often they're used, then pronouns, things like 'he', 'she', 'I', then nouns, naming words like 'cat', 'dog' and so on, then adjectives, the describing words like 'yellow', 'tall, 'thin', then the verbs were slightly faster again, and then the most rapidly evolving word types were conjunctions and prepositions. So conjunctions are words like 'and', 'but', 'because' and prepositions, words like 'in' and 'on' and 'at', and they evolved relatively quickly given how frequently we used them.

Rissa:
What could the explanation for these findings be? Pagel offers some suggestions which could provide exciting avenues for future research.

Mark Pagel:
Nouns and verbs carry a lot of information. It's extremely important to get them right in a sentence, whereas conjunctions and prepositions are sort of place-fillers in sentences, and so maybe it's the case that parts of speech influence the rate at which a word evolves by how much information they carry. And so it may be there's been a lot of pressure on us to get nouns and verbs right, and so when we look across a selection of words, it's words like the nouns and the verbs that evolve slowly and it's interesting that numerals, the number words, evolve the most slowly, and so here again if I'm trying to describe a band of people coming over the hill who might be trying to rob me and I say there's two when in fact and twenty or two hundred there could be big consequences, and so it may be the case that some words are very, very informative and so there's been pressure to get those right and as a result they evolve rather slowly.

Rissa:
Given how complex the structure of language can be, it's remarkable how well the proposed models were able to explain linguistic evolution.

Quentin Atkinson:
These two simple predictors of rate of change: the frequency with which we use words and the part of speech, can explain 50% of the variation in the rates at which different meanings change. By any standards that's quite a lot of the variation in rate of change. Contrary to what a lot of people think, languages aren't changing in a completely unpredictable way. We're able to use these two simple predictors to really make a pretty good guess about how quickly different meanings are changing. So that's really promising for any research attempting to model the process of language change, to really get an understanding of what is going on when we're using languages day-to-day and how that shapes the form of languages over centuries and millennia.