

We went in search of the world's hardest language – The Economist – Medium

yesterday

A certain genre of books about English extols the language's supposed difficulty and idiosyncrasy. "Crazy English", by an American folk-linguist, Richard Lederer, asks "how is it that your nose can run and your feet can smell?". Bill Bryson's "Mother Tongue: English and How It Got That Way" says that "English is full of booby traps for the unwary foreigner...Imagine being a foreigner and having to learn that in English one tells *a* lie but *the* truth."

Such books are usually harmless, if slightly fact-challenged. You tell "a" lie but "the" truth in many languages, partly because many lies exist but truth is rather more definite.

It may be natural to think that your own tongue is complex and mysterious. But English is pretty simple: verbs hardly conjugate; nouns pluralise easily (just add "s", mostly) and there are no genders to remember.

English-speakers appreciate this when they try to learn other languages. A Spanish verb has six present-tense forms, and six each in the preterite, imperfect, future, conditional, subjunctive and two different past subjunctives, for a total of 48 forms. German has three genders, seemingly so random that Mark Twain wondered why "a young lady has no sex, but a turnip has". (*Mädchen* is neuter, whereas *Steckrübe* is feminine.)

German has three genders, seemingly so random that Mark Twain wondered why "a young lady has no sex, but a turnip has"

English spelling may be the most idiosyncratic, although French gives it a run for the money with 13 ways to spell the sound “o”: o, ot, ots, os, ocs, au, aux, aud, auds, eau, eaux, ho and ö. “Ghoti,” as wordsmiths have noted, could be pronounced “fish”: gh as in “cough”, o as in “women” and ti as in “motion”. But spelling is ancillary to a language’s real complexity; English is a relatively simple language, absurdly spelled.

Case by case

Perhaps the “hardest” language studied by many Anglophones is Latin. In it, all nouns are marked for case, an ending that tells what function the word has in a sentence (subject, direct object, possessive and so on). There are six cases, and five different patterns for declining verbs into them. This system, and its many exceptions, made for years of classroom torture for many children. But it also gives Latin a flexibility of word order. If the subject is marked as a subject with an ending, it need not come at the beginning of a sentence. This ability made many scholars of bygone days admire Latin’s majesty – and admire themselves for mastering it. Knowing Latin (and Greek, which presents similar problems) was long the sign of an educated person.

Yet are Latin and Greek truly hard? These two genetic cousins of English, in the Indo-European language family, are child’s play compared with some. Languages tend to get “harder” the farther one moves from English and its relatives. Assessing how languages are tricky for English-speakers gives a guide to how the world’s languages differ overall.



Illustration by W. Vasconcelos

Even before learning a word, the foreigner is struck by how differently languages can sound. The uvular r's of French and the fricative, glottal ch's of German (and Scots) are essential to one's imagination of these languages and their speakers.

But sound systems get a lot more difficult than that. Vowels, for example, go far beyond a, e, i, o and u, and sometimes y. Those represent more than five or six sounds in English (consider the a's in father, fate and fat.)

And vowels of European languages vary more widely; think of the unlauded ones of German, or the nasal ones of French, Portuguese and Polish.

Yet much more exotic vowels exist, for example that carry tones: pitch that rises, falls, dips, stays low or high, and so on. Mandarin, the biggest

language in the Chinese family, has four tones, so that what sounds just like “ma” in English has four distinct sounds, and meanings.

That is relatively simple compared with other Chinese varieties. Cantonese has six tones, and Min Chinese dialects seven or eight. One tone can also affect neighbouring tones’ pronunciation through a series of complex rules.

Consonants are more complex. Some (p, t, k, m and n are common) appear in most languages, but consonants can come in a blizzard of varieties known as egressive (air coming from the nose or mouth), ingressive (air coming back in the nose and mouth), ejective (air expelled from the mouth while the breath is blocked by the glottis), pharyngealised (the pharynx constricted), palatised (the tongue raised toward the palate) and more. And languages with hard-to-pronounce consonants cluster in families.

Languages in East Asia tend to have tonal vowels, those of the north-eastern Caucasus are known for consonantal complexity: Ubykh has 78 consonant sounds. Austronesian languages, by contrast, may have the simplest sounds of any language family.

The click

Perhaps the most exotic sounds are clicks — technically “non-pulmonic” consonants that do not use the airstream from the lungs for their articulation. The best-known click languages are in southern Africa. Xhosa, widely spoken in South Africa, is known for its clicks. The first sound of the language’s name is similar to the click that English-speakers use to urge on a horse.

The leading expert on the !Xóõ developed a lump on his larynx from learning to make their sounds

For sound complexity, one language stands out. !Xóõ, spoken by just a few thousand, mostly in Botswana, has a blistering array of unusual sounds. Its vowels include plain, pharyngealised, strident and breathy, and they carry

four tones. It has five basic clicks and 17 accompanying ones. The leading expert on the !Xóõ, Tony Traill, developed a lump on his larynx from learning to make their sounds. Further research showed that adult !Xóõ-speakers had the same lump (children had not developed it yet).

Beyond sound comes the problem of grammar. On this score, some European languages are far harder than are, say, Latin or Greek. Latin's six cases cover in comparison with Estonian's 14, which include inessive, elative, adessive, abessive, and the system is riddled with irregularities and exceptions. Estonian's cousins in the Finno-Ugric language group do much the same. Slavic languages force speakers, when talking about the past, to say whether an action was completed or not. Linguists call this "aspect", and English has it too, for example in the distinction between "I go" and "I am going." And to say "go" requires different Slavic verbs for going by foot, car, plane, boat or other conveyance. For Russians or Poles, the journey does matter more than the destination.

Women, fire, and dangerous things

Beyond Europe things grow more complicated. Take gender. Twain's joke about German gender shows that in most languages it often has little to do with physical sex. "Gender" is related to "genre", and means merely a group of nouns lumped together for grammatical purposes. Linguists talk instead of "noun classes", which may have to do with shape or size, or whether the noun is animate, but often rules are hard to see. George Lakoff, a linguist, memorably described a noun class of Dyirbal (spoken in north-eastern Australia) as including "women, fire and dangerous things". To the extent that genders are idiosyncratic, they are hard to learn. Bora, spoken in Peru, has more than 350 of them.

Agglutinating languages — that pack many bits of meaning into single words — are a source of fascination for those who do not speak them. Linguists call a single unit of meaning, whether "tree" or "un-", a morpheme, and some languages bind them together obligatorily. The English curiosity "antidisestablishmentarianism" has seven morphemes ("anti", "dis", "establish", "-ment", "-ari", "-an" and "-ism"). This is unusual

in English, whereas it is common in languages such as Turkish. Turks coin fanciful phrases such as “Çekoslovakyalılastiramadıklarımızdanmissiniz?”, meaning “Were you one of those people whom we could not make into a Czechoslovakian?” But Ilker Aytürk, a linguist, offers a real-life example: “*Evlerindemisçesine rahattılar*”. Assuming you have just had guests who made a mess, these two words mean “They were as carefree as if they were in their own house.”



Yes we (but not you) can

This proliferation of cases, genders and agglutination, however, represents a multiplication of phenomena that are known in European languages. A truly boggling language is one that requires English speakers to think about things they otherwise ignore entirely. Take “we”. In Kwaio, spoken in the Solomon Islands, “we” has two forms: “me and you” and “me and someone else (but not you)”. And Kwaio has not just singular and plural, but dual and paucal too. While English gets by with just “we”, Kwaio has “we two”, “we few” and “we many”. Each of these has two forms, one inclusive (“we including you”) and one exclusive. It is not hard to imagine social situations that would be more awkward if you were forced to make this distinction explicit.

Berik, a language of New Guinea, also requires words to encode

information that no English speaker considers. Verbs have endings, often obligatory, that tell what time of day something happened; *telbener* means “[he] drinks in the evening”. Where verbs take objects, an ending will tell their size: *kitobana* means “gives three large objects to a man in the sunlight.” Some verb-endings even say where the action of the verb takes place relative to the speaker: *gwerantena* means “to place a large object in a low place nearby”. Chindali, a Bantu language, has a similar feature. One cannot say simply that something happened; the verb ending shows whether it happened just now, earlier today, yesterday or before yesterday. The future tense works in the same way.

A fierce debate exists in linguistics between those, such as Noam Chomsky, who think that all languages function roughly the same way in the brain and those who do not. The latter view was propounded by Benjamin Lee Whorf, an American linguist of the early 20th century, who argued that different languages condition or constrain the mind’s habits of thought.

Whorfianism has been criticised for years, but it has been making a comeback. Lera Boroditsky of Stanford University, for example, points to the Kuuk Thaayorre, aboriginals of northern Australia who have no words for “left” or “right”, using instead absolute directions such as “north” and “south-east” (as in “You have an ant on your south-west leg”). Ms Boroditsky says that any Kuuk Thaayorre child knows which way is south-east at any given time, whereas a roomful of Stanford professors, if asked to point south-east quickly, do little better than chance. The standard Kuuk Thaayorre greeting is “where are you going?”, with an answer being something like “north-north-east, in the middle distance.” Not knowing which direction is which, Ms Boroditsky notes, a Westerner could not get past “hello”. Universalists retort that such neo-Whorfians are finding trivial surface features of language: the claim that language truly constricts thinking is still not proven.

Language diversity

Greenberg's diversity index, 1= total diversity (where no two people meeting at random would have the same mother tongue), latest available year



Source: Ethnologue.com

Language diversity measured in two very different ways: the number of languages spoken in the country and Greenberg's diversity index, which scores countries on the probability that two citizens will share a mother tongue

And the winner is

With all that in mind, which is the hardest language? On balance *The Economist* would go for Tuyuca, of the eastern Amazon. It has a sound system with simple consonants and a few nasal vowels, so is not as hard to

speak as Ubykh or !Xóõ. Like Turkish, it is heavily agglutinating, so that one word, *hóabāsiriga* means “I do not know how to write.” Like Kwaio, it has two words for “we”, inclusive and exclusive. The noun classes (genders) in Tuyuca’s language family (including close relatives) have been estimated at between 50 and 140. Some are rare, such as “bark that does not cling closely to a tree”, which can be extended to things such as baggy trousers, or wet plywood that has begun to peel apart.

Most fascinating is a feature that would make any journalist tremble. Tuyuca requires verb-endings on statements to show how the speaker knows something. *Diga ape-wi* means that “the boy played soccer (I know because I saw him)”, while *diga ape-hiyi* means “the boy played soccer (I assume)”. English can provide such information, but for Tuyuca that is an obligatory ending on the verb. Evidential languages force speakers to think hard about how they learned what they say they know.

Linguists ask precisely how language works in the brain, and examples such as Tuyuca’s evidentiality are their raw material. More may be found, as only a few hundred of the world’s 6,000 languages have been extensively mapped, and new ways will appear for them to be difficult. Yet many are spoken by mere hundreds of people. Fewer than 1,000 people speak Tuyuca. Ubykh died in 1992. Half of today’s languages may be gone in a century. Linguists are racing to learn what they can before the forces of modernisation and globalisation quieten the strangest tongues.

This article originally appeared in the [2009 Christmas double issue](#) of The Economist.