

# **An Analysis of the Difficulty of Learning Various Foreign Languages for English speakers – Non-Indo-European Languages**

**By Robert Lindsay**

**Method.** 73 non-IE languages were examined. A literature survey, combined with interviews of various L2 language learners was conducted. In addition, 100 years of surveys on the question by language instructors was reviewed. The US military's School of Languages in Monterey's ratings system for difficulty of learning various languages was analyzed.

Results were collated in an impressionistic manner along a majority rules line in order to form final opinions. For example, a minority said that Portuguese or Spanish were very hard to learn, but the consensus view was that they were quite easy. In this case, the minority opinion was rejected, and the consensus view was adopted.

Clearly, such a project will necessarily be more impressionistic than scientific. Scientific tests of the relative difficulty of learning different languages will have to await the development of algorithms specifically designed to measure such things. In the meantime we will have to do with the best we have at the moment, which are impressionistic analyses.

**Results.** A ratings system was designed in terms of how difficult it would be for an English-language speaker to learn the language. In the case of English, English was judged according to how hard it would be for a non-English speaker to learn the language. Speaking, reading and writing were all considered.

**Ratings:** Languages are rated 1-6, easiest to hardest. 1 = easiest, 2 = moderately easy to average, 3 = average to moderately difficult, 4 = very difficult, 5 = extremely difficult, 6 = most difficult of all.

Ratings are impressionistic.

**Time needed.** Time needed for an English language speaker to learn the language "reasonably well": Level 1 languages = 3 months-1 year. Level 2 languages = 6 months-1 year. Level 3 languages = 1-2 years. Level 4 languages = 2 years. Level 5 languages = 3-4 years, but some may take longer. Level 6 languages = more than 4 years.

**Conclusion.** The results of this study indicate that languages do indeed differ dramatically in how difficult they are for L2 English language learners.

**Time needed:** Time needed to learn the language "reasonably well": Level 1 languages = 3 months-1 year. Level 2 languages = 6 months-1 year. Level 3 languages = 1-2 years. Level 4 languages = 2 years. Level 5 languages = 3-4 years, but some may take longer.

## **Northeast Caucasian, Northwest Caucasian and Kartvelian**

Of course the Caucasian languages like Tsez, Tabasaran, Georgian, **Chechen**, Ingush, Abkhaz and **Circassian** are some of the hardest languages on Earth to learn.

Chechen and Circassian are **rated 6**, hardest of all.

## **Northeast Caucasian**

NE Caucasian languages have the uvulars and ejectives of Georgian in addition to pharyngeals, lateral fricatives, and other strangeness. They have noun classes like the Bantu languages (but usually fewer). Nevertheless, they have noun class agreement markers on verbs on adjectives. One thing NE Caucasian has is lots of case. Some

languages have 40+ cases. They are built from the ground up via two forms - one a spatial form such as *in*, *on* or *around* and the other a directional motion form such as *to*, *from*, *through* or *at*.

## Tsezic

**Tsez** has 64-126 different cases, making it by far the most complex case system on Earth! It is one of the few languages on Earth that has two genitive cases - Genitive 1 (-s) and Genitive 2 (-z). Genitive 1 is used when the genitive's head noun is in absolutive case and Genitive 2 is used when the genitive's head noun is in any other case. It also has four noun classes. It is said that even native speakers have a hard time picking up the correct inflection to use sometimes.

In Tsez, you need to know a lot of Tsez grammar to communicate at a basic level. The sentence:

English: "I like your mother."

Tsez: *Д̄ар̄ деби энийу йетих.* (*Dār debi eniyu yetix.*)

In order to speak that sentence in Tsez, you need to know:

- the words themselves (word order is not as important)
- that the verb *-eti-* requires the subject to be in the dative/lative case and the object to be in the absolutive
- the noun class for *eniyu* (class II)
- the dative/lative form of *di* ("I", which is *dār*)
- the genitive 1 form of *mi* ("you", which is *debi*)
- the congruence prefix *y* corresponds to the noun class of the absolutive argument of the phrase, in this case "mother"
- the present tense ending for vowel-final verbs *-x*

**Tsez** is rated **6**, hardest of all.

## Lezgič

### Archi

**Archi** has an extremely complex phonology and one of the most complicated grammars on Earth. The extreme fusional aspects and the verbal morphology are what make the grammar so difficult. Every verb root has 1,502,839 possible forms!

It is also an ergative language, but there is irregularity in its ergative system. Some verbs take the typical ergative/absolutive case (absolutive for the subject of an intransitive verb and ergative for the subject of a transitive verb - where the direct object would be in absolutive). In others the subject is in dative rather than the expected ergative/absolutive case. These are usually verbs of perception like "love/want", "hear", "see", "feel", and "be bored".

For instance, the verb:

-эти- "to love/want" must have its subject in dative case instead of the expected absolutive or ergative case.

Among non-click languages, Archi has one of the largest consonant inventories, with only the extinct Ubykh having more. There are 26 vowels and between 76 and 82 consonants, depending on the analysis. Five of the six vowels can occur in five varieties: short, pharyngealized, high tone, long (with high tone), and pharyngealized with high tone.

It has many unusual phonemes, including contrasts between several voiceless velar lateral fricatives, voiceless and ejective velar lateral affricates and a voiced velar lateral fricative. The voiceless velar lateral fricative  $\text{ɬ}$ , the voiced velar lateral fricative  $\text{ɮ}$ , and the corresponding voiceless and ejective affricates  $\text{k}^{\text{h}}\text{ɬ}$  and  $\text{k}^{\text{h}}\text{ɮ}$  are extremely unusual sounds, as velar fricatives are not typically laterals.

There are 15 cases, 10 regular cases, five spatial cases and five

directional cases. The Spatial cases are:

Inessive "in"  
Intrative "between"  
Superessive "above"  
Subessive "below"  
Pertingent "against"

The directional cases are:

Essive "as"  
Elative "out of"  
Lative "to/into"  
Allative "onto"  
Terminative "specifies a limit"  
Translative "indicates change"

There are four noun classes:

- I Male human
  - II Female human
  - III All insects, some animates, and some inanimates
  - IV Abstracts, some animates, and some inanimates that can only be seen via verbal agreement
- Archi is **rated 6**, hardest of all.

## Samur

### Eastern Samur

#### Lezgi–Aghul–Tabasaran

**Tabasaran** is rated the 3rd most complex grammar in the world, with [48 different noun cases](#).

Tabasaran is **rated 6**, hardest of all.

## Nakh

### Vainakh

**Ingush** has a very difficult phonology, an extremely complex grammar, and furthermore, is extremely irregular. Ingush also has a proximate/obviate distinction and is the only language in the region that has this feature.

Ingush along with Chechen both have a [closed class](#) of verbs, an unusual feature in the world's languages.

New verbs are formed by adding a noun to the verb "do":

"shoot" = "do gun"

Ingush is **rated 6**, hardest of all.

## Kartvelian

### Karto-Zan

**Georgian** has [also](#) has lots of glottal stops that are hard for many

foreigners to speak; consonant clusters can be huge - up to eight consonants stuck together (CCCCCCCCVC)- and many consonant sounds are strange. In addition, there are uvulars and ejectives. Georgian is one of the hardest languages on Earth to [pronounce](#). It regularly makes it onto [craziest phonologies](#) lists.

Its grammar is exceedingly complex. Georgian is both highly agglutinative and highly irregular, which is the worst of two worlds. Other agglutinative languages such as Turkish and Finnish at least have the benefit of being highly regular. The verbs in particular seem nearly random with no pattern to them at all. The system of argument and tense marking on the verb is exceedingly complex, with tense, aspect, mood on the verb, person and number marking for the subject, and direct and indirect objects.

Although it is an ergative language, the ergative (or active-stative case marking as it is called) oddly enough is only used in the aorist and perfect tenses where the agent in the sentence receives a different case, while the aorist also masquerades as imperative. In the present, there is standard nominative-accusative marking. A single verb can have up to 12 different parts, similar to Polish, and there are six cases and six tenses.

Georgian also features something called polypersonal agreement, a highly complex type of morphological feature that is often associated with polysynthetic languages and to a lesser extent with ergativity.

In a polypersonal language, the verb has agreement morphemes attached to it dealing with one or more of the verbs arguments (usually up to four arguments). In a non-polypersonal language like English, the verb either shows no agreement or agrees with only one of its arguments, usually the subject. Whereas in a polypersonal language, the verb agrees with one or more of the subject, the direct object, the indirect object, the beneficiary of the verb, etc. The polypersonal marking may be obligatory or optional.

In Georgian, the polypersonal morphemes appear as either suffixes or prefixes, depending on the verb class and the person, number, aspect and tense of the verb. The affixes also modify each other phonologically when they are next to each other. In the Georgian system, the polypersonal affixes convey subject, direct object, indirect object, genitive, locative, and causative meanings.

*g-mal-av-en* "they hide you"  
*g-i-mal-av-en* "they hide it **from** you"

*mal* ("to hide" is the verb, and the other four forms are polypersonal affixes.

In the case below,

*xelebi ga-m-i-tziv-d-a* "My hands got cold".

*xelebi* means "hands". The *m* marker indicates genitive or "my". With intransitive verbs, Georgian often omits *my* before the subject and instead puts the genitive onto the verb to indicate possession.

Georgian verbs of motion focus on deixis, whether the goal of the motion is towards the speaker or the hearer. You use a particle to signify who the motion is heading towards. If it heading towards neither of you, you use no deixis marker. You specify the path taken to reach the goal through the use or prefixes called preverbs, similar to "verbal case." These come after the deixis marker:

<i>up</i>	<i>a-</i>
<i>out</i>	<i>ga-</i>
<i>in</i>	<i>sha-</i>
<i>down into</i>	<i>cha-</i>
<i>across/through</i>	<i>garda-</i>
<i>thither</i>	<i>mi-</i>
<i>away</i>	<i>c'a-</i>
<i>or down</i>	<i>da-</i>

Hence:

"up towards me" *amo-*. The deixis marker is *mo-* and "up" is *a-*

On the plus side, Georgian has borrowed a great deal of Latinate foreign vocabulary, so that will help anyone coming from a Latinate or Latinate-heavy language background.



Georgian is **rated 5**, extremely difficult.

## Northwest Caucasian

All NW Caucasian languages are characterized by a very small number of vowels (usually only two or three) combined with a vast consonant inventory, the largest consonant inventories on Earth. Almost any consonant can be plain, labialized or palatalized. This is apparently the result of an historical process whereby many vowels were lost and their various features became assigned to consonants. For instance, palatalized consonants may have come from *Ci* sequences and labialized consonants may have come from *Cu* sequences.

The grammars of these languages are complex. Unlike the NE Caucasian languages, they have simple noun systems, usually with only a handful of cases.

However, they have some of the complex verbal systems on Earth. These are some of the most synthetic languages in the Old World. Often the entire syntax of the sentence is contained within the verb. All verbs are marked with ergative, absolutive and direct object morphemes in addition to various applicative affixes. These are akin to what some might call "verbal case." For instance, in [applicative voice systems](#), applicatives may take forms such as comitative, locative, instrumental, benefactive and malefactive. These roles are similar to the case system in nouns - even the names are the same. So you can see why some call this "verbal case."

NW Caucasian verbs can be marked for aspect (whether something is momentous, continuous or habitual), mood (if something is certain, likely, desired, potential, or unreal). Other affixes can shape the verb in an adverbial sense, to express pity, excess or emphasis.

Like NE Caucasian, they are also ergative.

NW Caucasian makes it onto a lot of [craziest language](#) lists.

These are some of the strangest sounding languages on Earth. Of all of these languages, Abaza has the most consonants. [Here is a video](#) in the Abaza language.

# Ubykh

**Ubykh**, a Caucasian language of Turkey, is now extinct, but there is one second language speaker, a linguist who is said to have taught himself the language.

It has more consonants than any non-click language on Earth – 84 [consonant](#) sounds in all. Furthermore, the phonemic inventory allows some very strange consonant clusters. Ubykh has many rare consonant sounds.  $t^w$  is only also found in two of Ubykh's relatives, Abkhaz and Abaza and in two other languages, both in the Brazilian Amazon. The pharyngealized labiodental voiced fricative  $v^{\text{̣}}$  does not exist in any other language. It often makes it onto [weirdest phonologies](#) lists. Ubykh also got a very high score on a study of the [weirdest languages on Earth](#).

Combine that with only two vowel sounds and a highly complex grammar, and you have one tough language.

In addition, Ubykh is both agglutinative and polysynthetic, ergative and has polypersonal agreement:

*Aχ<sup>j</sup>azbatš<sup>ʷ</sup>awdət<sup>w</sup>aajlafaq<sup>ʷ</sup>ajt<sup>ʷ</sup>madaχ!*

"If only you had not been able to make him take it all out from under me again for them..."

There are an incredible 16 morphemes in that nine syllable word.

Ubykh has only four case systems on its nouns, but much case function has shifted over to the verb via preverbs and determinants. It is these preverbs and determinants that make Ubykh monstrously complex. The following are some of the directional preverbs:

- above and touching
- above and not touching
- below and touching
- below and not touching
- at the side of
- through a space

- through solid matter
- on a flat horizontal surface
- on a non-horizontal or vertical surface
- in a homogeneous mass
- towards
- in an upward direction
- in a downward direction
- into a tubular space
- into an enclosed space

There are also some preverbal forms that indicate deixis:

*j-* towards the speaker

Others can indicate ideas that would take up whole phrases in English:

*jtɕ<sup>w</sup>aa-* "on the Earth, in the Earth"

*ɤadja ajtɕ<sup>w</sup>aanaatq'a*

"They buried his body." Lit. "They put his body in the earth."

*faa-* "out of, into or with regard to a fire".

*Amdʒan zatʃətʃaɕ<sup>j</sup>a faastɕ<sup>w</sup>ən.*

"I take a brand out of the fire."

Morphemes may be as small as a single phoneme:

*want<sup>w</sup>aan* "They give you to him."

*w* - 2nd singular absolutive

*a* - 3rd singular dative

*n* - 3rd ergative

*t<sup>w</sup>* - to give

*aa* - ergative plural

*n* - present tense

Adverbial suffixes are attached to words to form meanings that are often formed by aspects or tenses in other languages:

*asfəpχa* "I need to drink it."  
*asfəfan* "I can drink it."  
*asfəgʲan* "I drink it all the time."  
*asfəlan* "I am drinking it all up."  
*asfətɕʷan* "I drink it too much."  
*asfaajən* "I drink it again."

Nouns and verbs can transform into each other. Any noun can turn into a stative verb:

*məzə* "child"  
*səməzəjtʰ* "I was a child." Lit. "I child-was." "child-was" is a verb "to be a child."

By the same token, many verbs can become nouns via the use of a nominal affix:

*qʰa* "to say"  
*səqʰa* "what I say" Lit. "That which I say - my speech, my words, my language, my orders, etc."

Number is marked on the verb via a verbal suffix and is only marked on the noun in the ergative case.

However, it does lack the convoluted case systems of the Caucasian languages next door and there is no grammatical gender.

Ubykh is **rated 6**, hardest of all.

## Abkhaz-Abazin

**Abkhaz** is an extremely difficult language to learn. Each basic consonant has eight different positions of articulation in the mouth. Imagine how difficult that would be for an Abkhaz child with a speech impediment. Abkhaz seems to put agreement markers on just about everything in the language. Abkhaz makes it onto many [craziest language](#) lists, and it recently got a very high score on a [weirdest language](#) study.

Abkhaz is **rated 6**, hardest of all.

## Burushaski

**Burushaski** is often thought to be a language isolate, related to no other languages, however, I think it is Dene-Caucasian. It is spoken in the Himalaya Mountains of far northern Pakistan in an area called the Hunza. It's verb conjugation is complex, it has a lot of inflections, there are complicated ways of making sentences depending on many factors, and it is an ergative language, which is hard to learn for speakers of non-ergative languages. In addition, there are very few to no cognates for the vocabulary.

Burushaski is **rated 6**, hardest of all.

## American Indian Languages

American Indian languages are also notoriously difficult, though few try to learn them in the US anyway. In the rest of the continent, they are still learned by millions in many different nations. You almost really need to learn these as a kid. It's going to be quite hard for an adult to get full competence in them.

One problem with these languages is the multiplicity of verb forms. For instance, the standard paradigm for the overwhelming number of regular English verbs is a maximum of five forms:

*steal*  
*steals*  
*stealing*  
*stole*  
*stolen*

Many Amerindian languages have over 1,000 forms of each verb in the language.

## **Kootenai**

Yet the Salishans (see below) always considered the neighboring language **Kootenai** to be too hard to learn. Kootenai also has a distinction between proximate/obviate along with direct/inverse alignment, probably from contact with Algonquian. However, the Kootenai direct/inverse system is less complex than Algonquian's, as it is present only in the 3rd person. Kootenai also has a [very strange feature](#) in that they have particles that look like subject pronouns, but these go outside of the full noun phrase. This is a very rare feature in the world's languages. Kootenai scored [very high](#) on a weirdest language survey.

Kootenai is an isolate spoken in Idaho by 100 people.

Kootenai is **rated 6**, hardest of all.

## **Yuchi**

Yuchi is a language isolate spoken in the Southern US. They were originally located in Eastern Tennessee and were part of the Creek Confederacy at one time. Yuchi is nearly extinct, with only five remaining speakers.

Yuchi has noun genders or classes based on three distinctions of position: standing, sitting or lying. All nouns are either standing,

sitting or lying. Trees are standing, and rivers are lying, for instance. If it is taller than it is wide, it is standing. If it is wider than it is tall, it is lying. If it is about as wide as it is tall, it is sitting. All nouns are one of these three genders, but you can change the gender for humorous or poetic effect. A linguist [once asked](#) a group of female speakers whether a penis was standing, sitting, or lying. After lots of giggles, they said the default was sitting, but you could say it was standing or lying for poetic effect.

Also all Yuchi pronouns must make a distinction between age (older or younger than the speaker) and ethnicity (Yuchi or non-Yuchi).

Yuchi gets a **6 rating**, hardest of all.

## **Dene-Yeniseian**

### **Na-Dene**

### **Athabaskan-Eyak**

### **Tlingit**

**Tlingit** is probably one of the hardest, if not the hardest, language in the world.

Tlingit is analyzed as partly synthetic, partly agglutinative, and sometimes polysynthetic. It has not only suffixes and prefixes, but it also has infixes or affixes in the middle of words.

-*'eech*- "to pick"

All prefixes must be in proper order for the word to work.

*tuyakaoonagadagaxayaecheen*.

"I am usually picking, on purpose, a long object through the hole

while standing on a table.”

*tuyakaoonagootxaya'eecheen.*

“I am usually being forced to pick a long object through the hole while standing on a table.”

*tuyaoonagootxawa'eecheen.*

“I am usually picking the edible long object through the hole while standing on a table.”

Tlingit has a pretty unusual phonology. For one thing, it is [the only language on Earth with no /l/](#). This despite the fact that it has five other laterals: *dl* ( $t^{\downarrow}$ ), *tl* ( $t^{\downarrow h}$ ), *tl'* ( $t^{\downarrow \uparrow}$ ), *l* ( $\downarrow$ ) and *l'* ( $\uparrow$ ). The  $\downarrow$  and  $\uparrow$  sounds are rare in the world's languages.  $\uparrow$  is only found in the wild NW Caucasian languages. It also has two labialized glottal consonants,  $\gamma^w$  and *hw* ( $h^w$ ).

Tlingit gets a **6 rating**, hardest of all.

## Athabaskan

### Southern

[Navajo](#) has long, short and nasal vowels, a tone system and a grammar totally unlike anything in Indo-European. A stem of only four letters or so can take enough affixes to fill a whole line of text.

Navajo is a polysynthetic language. In polysynthetic languages, very long words can denote an entire sentence, and it's quite hard to take the word apart into its parts and figure out exactly what they mean and how they go together. The long words are created because polysynthetic languages have an amazing amount of morphological richness. They put many morpheme together to create a word out of what might be a sentence in a non-polysynthetic language.

Some Navajo dictionaries have thousands of entries of verbs only, with no nouns. Many adjectives have no direct translation into Navajo. Instead, verbs are used as adjectives. A verb has no particular form like in English “to walk”. Instead, it assumes various forms depending



on whether or not the action is completed, incomplete, in progress, repeated, habitual, one time only, instantaneous, or simply desired. These are called aspects. Navajo must have one of the most complex aspect systems of any language:

The Primary aspects:

Momentaneous: punctually (takes place at one point in time)  
Continuative: an indefinite span of time & movement with a specified direction  
Durative: over an indefinite span of time, non-locomotive uninterrupted continuum  
Repetitive: a continuum of repeated acts or connected series of acts  
Conclusive: like durative but in perfective terminates with static sequel  
Semelfactive: a single act in a repetitive series of acts  
Distributive: a distributive manipulation of objects or performance of actions  
Diversative: a movement distributed among things (similar to distributive)  
Reversative: results in directional change  
Conative: an attempted action  
Transitional: a shift from one state to another  
Cursive: progression in a line through time/space (only progressive mode)

The subspects:

Completive: an event/action simply takes place (similar to the aorist tense)  
Terminative: a stopping of an action  
Stative: sequentially durative and static  
Inceptive: beginning of an action  
Terminal: an inherently terminal action  
Prolongative: an arrested beginning or ending of an action

Seriative: an interconnected series of successive separate & distinct acts  
 Inchoative: a focus on the beginning of a non-locomotion action  
 Reversionary: a return to a previous state/location  
 Semeliterative: a single repetition of an event/action

The tense system is almost as wild as the aspectual system.

For instance, the verb *ndideesh* means "to pick up" or "to lift up". But it varies depending on what you are picking up:

*ndideeshtil* "to pick up a slender stiff object (key, pole)"  
*ndideeshleel* "to pick up a slender flexible object (branch, rope)"  
*ndideesh'aal* "to pick up a roundish or bulky object (bottle, rock)"  
*ndideeshgheel* "to pick up a compact and heavy object (bundle, pack)"  
*ndideeshjol* "to pick up a non-compact or diffuse object (wool, hay)"  
*ndideeshteel* "to pick up something animate (child, dog)"  
*ndideeshnil* "to pick up a few small objects (a couple of berries, nuts)"  
*ndideeshjih* "to pick up a large number of small objects (a pile of berries, nuts)"  
*ndideeshtsos* "to pick up something flexible and flat (blanket, piece of paper)"  
*ndideeshjil* "to pick up something I carry on my back"  
*ndideeshkaal* "to pick up anything in a vessel"  
*ndideeshtloh* "to pick up mushy matter (mud)"

But picking up is only one way of handling the 12 different consistencies. One can also bring, take, hang up, keep, carry around, turn over, etc. objects. There are about 28 different verbs one can use for handling objects. If we multiply these verbs by the consistencies, there are over 300 different verbs used just for handling objects.

In Navajo textbooks, there are conjugation tables for inflecting words, but it's pretty hard to find a pattern there. One of the most frustrating things about Navajo is that every little morpheme you add to a word seems to [change everything else around it](#), even in both directions.

Navajo is said to have a very difficult system for counting numerals.

There is also a noun classifier system with more than a dozen classifiers that affect inflection. This is quite a few classifiers even for a noun classifier language and is similar to African languages like Zulu. In addition, it has the strange direct/inverse system.

To add insult to injury, Navajo is an ergative language.

Navajo also has an honorifics or politeness system similar to Japanese or Korean.

Navajo also has the odd feature where the word *niinaa* "because" can be analyzed as a verb.

*X áhóót'íjd biniinaa...*

"Because X happened..."

*Shiniinaa sits'il.*

"It broke into pieces because of me."

In the latter sentence, the only way we know that 1st singular was involved in because of the person marking on *niinaa*.

There are 25 different kinds of pronominal prefixes that can be piled onto one another before a verb base.

Navajo has a very strange feature called animacy, where nouns take certain verbs according to their rank in the hierarchy of animation which is a sort of a ranking based on how alive something is. Humans and lightning are at the top, children and large animals are next and abstractions are at the bottom.

All in all, Navajo, even compared to other polysynthetic languages, has some of the most incredibly complicated polysynthetic morphology of any language. On [craziest grammar and craziest language](#) lists, Navajo is typically listed.

It is even said that Navajo children have a hard time learning Navajo as compared to children learning other languages, but Navajo kids definitely learn the language. Similarly with Hopi below, [even linguists](#) find even the best Navajo grammars difficult or even impossible to

understand.

However, Navajo is quite regular, a common feature in Amerindian languages.

Navajo is **rated 6**, hardest of all.

## Northern

**Slavey**, a Na-Dene language of Canada, is [hard](#) to learn. It is similar to Navajo and Apache. Verbs take up to 15 different prefixes. All Athabaskan languages have wild verbal systems. It also uses a completely different alphabet, a syllabic one designed for Canadian Indians.

Slavey is **rated 6**, hardest of all.

## Haida

**Haida** is often thought to be a Na-Dene language, but proof of its status is lacking. If it is Na-Dene, it is the most distant member of the family. Haida is in the competition for the most complicated language on Earth, with 70 different suffixes.

Haida is **rated 6**, hardest of all.

## Salishan

The **Salishan languages** spoken in the Northwest have a long reputation for being hard to learn, in part because of long strings of consonants, in one case 11 consonants long. Salish languages are the only languages on Earth that allow words without sonorants. Many of the vowels and consonants are not present in most of the world's widely spoken languages. The Salish languages are, like Chukchi, polysynthetic. Some translations treat all Salish words as either

verbs or phrases. Some say that Salish languages do not contain nouns, though this is controversial. The verbal system of Salish languages is absurdly complex.

All Salishan languages are rated **rated 6**, hardest of all.

## Nuxálk (Bella Coola)

**Nuxálk** is a notoriously difficult Salishan Amerindian language spoken in British Columbia.

It is famous for having some wild words and even sentences that don't seem to have any vowels in them at all. For instance:

*xłpχ<sup>w</sup> łtłpłsk<sup>w</sup>c* (*xłp'χ<sup>w</sup>łtłpł:sk<sup>w</sup>t* s' in IPA)  
"He had a bunchberry plant."

*sxs*  
"seal fat"

Here are some more [odd words and sentences](#):

*smnmnmuuc*  
"mute"

*Nuyamłamkis timantx tisyuttx ʔultimnastx.*  
"The father sang the song to his son."

*Musis tɪimmlkītχ taqlsx<sup>w</sup>t aχ.*  
"The boy felt that rope."

The language sounds odd when spoken. It has been described as ["whispering while chewing on a granola bar"](#) (see the video sample under Montana Salish below).

These wild consonant clusters are even crazier than the ones in Ubykh and NW Caucasian. In fact, the nutty consonant clusters in Salish and

causing a debate in linguistics about whether or not the syllable is even a universal phenomenon in language as some Salish words and phrases appear to lack syllables. Some Berber dialects have raised similar questions about the syllable.

Nuxálk makes it onto lists of the [craziest phonologies](#) on Earth.

Nuxálk is **rated 6**, hardest of all.

## Interior Salish

### Southern

**Montana Salish** is said to be just as hard to learn as Nuxálk.

Spokane (Montana Salish) has combining and independent forms with the same meaning:

*spim'cn* "mouth"  
*-cin* "mouth"

Montana Salish makes it onto a lot of [craziest grammars](#) lists.

[This link](#) shows an elder on the Flathead Indian Reservation in Montana, Steven Smallsalmon, speaking Montana Salish. He also leads classes in the language. This is probably one of the strangest sounding languages on Earth.

Montana Salish is **rated 6**, hardest of all.

### Central

**Straits Salish** has an aspectual distinction between persistent and nonpersistent. Persistent means the activity continues after its inception as a state. The persistent morpheme is *-í*. The result is

similar to English:

<i>figure out</i>	nonpersistent
<i>know</i>	persistent
<i>look at</i>	nonpersistent
<i>watch</i>	persistent
<i>take</i>	nonpersistent
<i>hold</i>	persistent

*-í* is referred to as a "parasitic morpheme" and only occurs in stem that has an underlying *ə* which serves as a "host" for the *-í* morpheme.

How strange.

The Saanich dialect of Straits Salish is often listed in the [rogue's gallery](#) of craziest grammars on Earth. The [writing system](#) is often listed as one of the worst out there. In addition, Saanich makes it onto craziest grammars lists for the parasitic morphemes and for having no distinction between nouns and verbs!

Straits Salish gets a **6 rating**, hardest of all.

**Halkomelem**, spoken by 570 people around Vancouver, British Columbia, is widely considered to be one of the hardest languages on Earth to learn.

In Halkomelem, many verbs have an orientation towards water. You can't just say, "She went home." You have say how she was going home in relation to nearby bodies of water.

So depending on where she was walking home in relation to the nearest river, you would say:

"She was farther away from the water and going home."

"She was coming home in the direction away from the water."

"She was walking parallel to the flow of the water downstream."

"She was walking parallel to the flow of the water upstream."

Halkomelem gets a **6 rating**, hardest of all.

## Lushootseed

**Lushootseed** is said to be just as hard to learn as Nuxálk.

Lushootseed is one of the few languages on Earth that has no nasals at all, except in special registers like baby talk and the archaic speech of mythological figures. It also has laryngealized glides and nasals.

Lushootseed is **rated 6**, hardest of all.

## Iroquoian

All Iroquoian languages are extremely difficult, but Athabaskan is probably even harder. Siouan languages may be equal to Iroquoian in difficulty.

Compare the same phrases in Tlingit (Athabaskan) and Cherokee (Iroquoian).

Tlingit:

*kutíkusa'áat* "It's cold outside."

*kutíkuta'áat* "It's cold right now."

In Tlingit, you can add or modify affixes at the beginning as prefixes, in the middle as infixes and at the end as suffixes. In the above example, you changed a part of the word within the clause itself.

Cherokee:



*doyáditlv uyvtlv* "It is cold outside." Lit. "Outside it is cold."  
*ka uyvtlv* "It is cold now." Lit. "Now it is cold."

As you can see, Cherokee is easier.

## Cherokee

**Cherokee** is very hard to learn. In addition to everything else, it has a completely different alphabet. It's polysynthetic, to make matters worse. It is possible to write a Cherokee sentence that somehow lacks a verb. There are five categories of verb classifiers. Verbs needing classifiers must use one. Each regular verb can have an incredible 21,262 inflected forms! All verbs contain a verb root, a pronominal prefix, a modal suffix and an aspect suffix. In addition, verbs inflect for singular, plural and also dual. For instance:

DATꞆ *a'lv'íha*

You have 126 different forms:

EꞆꞆꞆꞆ	<i>gvyalv'íha</i>	<i>I tie you up</i>
SEꞆꞆꞆꞆ	<i>degvyalviha</i>	<i>I'm tying you up</i>
hꞆꞆꞆꞆ	<i>jiyalv'ha</i>	<i>I tie him up</i>
SꞆꞆꞆꞆ		<i>I tie it</i>
ꞆꞆꞆꞆꞆꞆ	<i>sdayalv'íha</i>	<i>I tie you (dual)</i>
TCꞆꞆꞆꞆ	<i>ijvyalv'íha</i>	<i>I tie you (pl)</i>
SHꞆꞆꞆꞆ	<i>gajiyalv'íha</i>	<i>I tie them (animate)</i>
SSꞆꞆꞆꞆ		<i>I tie them up (inanimate)</i>
ꞆꞆꞆꞆꞆ	<i>squahlv'íha</i>	<i>You tie me</i>
AꞆꞆꞆꞆ	<i>hiyalv'íha</i>	<i>You're tying him</i>
ꞆꞆꞆꞆꞆ	<i>hatlv'íha</i>	<i>You tie it</i>
ꞆꞆꞆꞆꞆꞆ	<i>skinalv'íha</i>	<i>You're tying me and him</i>
AVꞆꞆꞆꞆꞆ	<i>goginatlv'íha</i>	<i>They tie me and him etc.</i>

Let us look at another form:

*to see*

<i>I see myself</i>	<i>gadagotia</i>
<i>I see you</i>	<i>gvgohtia</i>
<i>I see him/her</i>	<i>tsigotia</i>
<i>I see it</i>	<i>tsigotia</i>
<i>I see you two</i>	<i>advgotia</i>
<i>I see you (plural)</i>	<i>istvgotia</i>

<i>I see them (live)</i>	<i>gatsigotia</i>
<i>I see them (things)</i>	<i>detsigotia</i>
<i>You see me</i>	<i>sgigotia</i>
<i>You see yourself</i>	<i>hadagotia</i>
<i>You see him/her</i>	<i>higo(h)tia</i>
<i>You see it</i>	<i>higotia</i>
<i>You see another and me</i>	<i>sginigotia</i>
<i>You see others and me</i>	<i>isgigotia</i>
<i>You see them (living)</i>	<i>dehigotia</i>
<i>You see them (living)</i>	<i>gahigotia</i>
<i>You see them (things)</i>	<i>detsigotia</i>
<i>He/she sees me</i>	<i>agigotia</i>
<i>He/she sees you</i>	<i>tsagotia</i>
<i>He/she sees you</i>	<i>atsigotia</i>
<i>He/she sees him/her</i>	<i>agotia</i>
<i>He/she sees himself/herself</i>	<i>adagotia</i>
<i>He/she sees you + me</i>	<i>ginigotia</i>
<i>He/she sees you two</i>	<i>sdigotia</i>
<i>He/she sees another + me</i>	<i>oginigotia</i>
<i>He she sees us (them + me)</i>	<i>otsigotia</i>
<i>He/she sees you (plural)</i>	<i>itsigotia</i>
<i>He/she sees them</i>	<i>dagotia</i>
<i>You and I see him/her/it</i>	<i>igigotia</i>
<i>You and I see ourselves</i>	<i>edadotia</i>
<i>You and I see one another</i>	<i>denadagotia/dosdadagotia</i>
<i>You and I see them (living)</i>	<i>genigotia</i>
<i>You and I see them (living or not)</i>	<i>denigotia</i>
<i>You two see me</i>	<i>sgninigotia</i>
<i>You two see him/her/it</i>	<i>esdigotia</i>
<i>You two see yourselves</i>	<i>sdadagotia</i>
<i>You two see us (another and me)</i>	<i>sginigotia</i>
<i>You two see them</i>	<i>desdigotia</i>
<i>Another and I see you</i>	<i>sdvgotia</i>
<i>Another and I see him/her</i>	<i>osdigotia</i>
<i>Another and I see it</i>	<i>osdigotia</i>
<i>Another and I see you-two</i>	<i>sdvgotia</i>
<i>Another and I see ourselves</i>	<i>dosdadagotia</i>
<i>Another and I see you (plural)</i>	<i>itsvgotia</i>
<i>Another and I see them</i>	<i>dosdigotia</i>
<i>You (plural) see me</i>	<i>isgigoti</i>
<i>You (plural) see him/her</i>	<i>etsigoti</i>
<i>They see me</i>	<i>gvgigotia</i>

<i>They see you</i>	<i>getsagotia</i>
<i>They see him/her</i>	<i>anigoti</i>
<i>They see you and me</i>	<i>geginigoti</i>
<i>They see you two</i>	<i>gesdigoti</i>
<i>They see another and me</i>	<i>gegigotia/gogenigoti</i>
<i>They see you (plural)</i>	<i>getsigoti</i>
<i>They see them</i>	<i>danagotia</i>
<i>They see themselves</i>	<i>anadagoti</i>
<i>I will see</i>	<i>datsigoi</i>
<i>I saw</i>	<i>agigohvi</i>
<i>He/she will see</i>	<i>dvgohi</i>
<i>He/she</i>	<i>sawugohvi</i>

Number is marked for inclusive vs. exclusive and there is a dual. 3rd person plural is marked for animate/inanimate. Verbs take different object forms depending on if the object is solid/alive/indefinite shape/flexible. This is similar to the Navajo system.

Cherokee also has lexical tone, with complex rules about how tones may combine with each other. Tone is not marked in the orthography. The phonology is noted for somehow not having any labial consonants.

However, Cherokee is very regular. It has only three irregular verbs. It is just that there are many complex rules.

Cherokee is **rated 5.5**, close to most difficult of all.

## **Iroquoian**

### **Northern Iroquoian**

#### **Five Nations-Huronian-Susquehannock**

##### **Huronian**

##### **Huron-Petun**

**Wyandot**, a dormant language that has been extinct for about 50 years, has some unbelievably complex structures. Let us look at [one of them](#). Wyandot is the only language on Earth that allows negative sentences that somehow do not contain a negative morpheme. Wyandot makes it onto [craziest grammars](#) lists. (To be continued).

## **Siouan-Catawban**

### **Siouan**

#### **Mississippi Valley-Ohio Valley Siouan**

#### **Mississippi Valley Siouan**

### **Dakota**

Lakota and other Siouan languages may well be as convoluted as Iroquoian.

In Lakota, all adjectives are expressed as verbs. Something similar is seen in Nahuatl.

*Ógle sápe kiŋ mak'ú.*

"The shirt **it is black** he gave it to me."

"He gave me the black shirt."

In the above, "it is black" is a stative verb and serves as an adjective.

*Ógle kiŋ sabyá mak'ú.*

"Shirt the blackly he gave it to me."

"He gave me the black shirt." Lit. "He gave me the shirt **blackly**."

"Blackly" is an adverb serving as an adjective above.

Lakota gets a **5.5 rating**, hardest of all.

## **Algic**

### **Algonquian**

All **Algonquian languages** have distinctions between animate/inanimate nouns, in addition to having proximate/obviate and direct/inverse distinctions. However, most languages that have proximate/obviate and direct/inverse distinctions are not as difficult as Algonquian. Proximate/obviative is a way of marking the 3rd person in discourse. It distinguishes between an important 3rd person (proximate) and a more peripheral 3rd person (obviative). Animate nouns and possessor nouns tend to be marked proximate while inanimate nouns and possessed nouns tend to be marked obviative.

Direct/inverse is a way of marking discourse in terms of saliency, topicality or animacy. Whether one noun ranks higher than another in terms of saliency, topicality or animacy means that that nouns ranks higher in terms of person hierarchy. It is used only in transitive clauses. When the subject has a higher ranking than the object, the direct form is used. When the object has a higher ranking than the object, the inverse form is used.

### **Central Algonquian**

#### **Cree-Montagnais**

**Cree** is very hard to learn. They are written in a variety of different ways with different alphabets and syllabic systems, complicating matters even further. The syllabic alphabet has many problems and is often listed as one of the [worst scripts](#) out there. They are both polysynthetic and have long, short and nasal vowels and aspirated and unaspirated voiceless consonants. Words are divided into metrical feet, the rules for determining stress placement in words are quite complex and there is lots of irregularity. Vowels fall out a lot, or

syncopate, within words.

Cree adds noun classifiers to the mix, and both nouns and verbs are marked as animate or inanimate. In addition, verbs are marked for transitive and intransitive. In addition, verbs get different affixes depending on whether they occur in main or subordinate clauses.

Cree is **rated 6**, hardest of all.

## Ojibwa-Patowatomi

**Ojibwa** is said to be about as hard to learn as Cree as it is very similar.

Ojibwa is **rated 6**, hardest of all.

## Plains Algonquian

### Cheyenne

**Cheyenne** is well-known for being a hard Amerindian language to learn.

Like many polysynthetic languages, it can have very long words.

*Náohkêsáa'oné'seómepehévetsêhésto'anéhe.*  
"I truly don't know Cheyenne very well."

However, Cheyenne is quite regular, but has so many complex rules that it is hard to figure them all out.

Cheyenne is **rated 6**, hardest of all.

## Arapahoan

**Arapaho** has a [strange phonology](#). It lacks phonemic low vowels. The vowel system consists of *i*, *ɨ*, *u*, *ɛ*, and *ɔ*, with no low phonemic vowels. Each vowel also has a corresponding long version. In addition, there are four diphthongs, *ei*, *ou*, *oe*, and *ie*, several triphthongs, *eii*, *oee*, and *ouu*, as well as extended sequences of vowels such as *eee* with stress on either the first or the last vowel in the combination.

Long vowels of various types are common:

*Héétbih'ínkúútiinoo.*

"I will turn out the lights."

*Honoosóó'.*

"It is raining."

There is a pitch accent system with normal, high, and allophonic falling tones. Arapaho words also undergo some very wild sound changes.

Arapaho is **rated 6**, hardest of all.

**Gros Ventre** has a similar phonological system and similar elaborate sound changes as Arapaho.

Gros Ventre is **rated 5**, hardest of all.

## Caddoan

## Northern

## Wichita

**Wichita** has many strange phonological traits. It has only one nasal. Labials are rare and appear in only two roots. It also may have only

three vowels, *i*, *e*, and *a*, with only height as a distinction. Such a restricted vertical vowel distribution is only found in NW Caucasian and the Papuan Ndu languages. There is apparently a three-way contrast in vowel length - regular, long and extra-long. This is only found in Mixe and Estonian.

There are some interesting tenses. Perfect tense means that an act has been carried out. The strange intensive tense means that one hopes or hoped to carry out an act. The habitual tense means one regularly engages in the activity, not that one is doing so at the moment.

Long consonant clusters are permitted.

*nahʔinckskih*  
"while sleeping"

There are many cases where a *CVʔ* sequence has been reduced to *Cʔ* due to loss of the vowel, resulting in odd words such as:

*ki·sʔ*  
"bone"

Word order is ordered in accordance with novelty or importance.

*hira:wisʔiha:s kiyari:ce:hire:*  
"Our ancestors God put us on this Earth."

*weʔe hira:rʔ tiʔi na:kirih*  
"God put our ancestors on this Earth."

In the sentence above, "our ancestors" is actually the subject, so it makes sense that it comes first.

Wichita has inclusive and exclusive 3rd person plural and has singular, dual and plural. There is an evidential system where if you say you



know something, you must say how you know it - whether it is personal knowledge or hearsay.

Wichita gets a **6 rating**, hardest of all.

## Hokan

### Tequiatecan

### Coastal Chantal

**Huamelutec** or Lowland Oaxaca Chantal has the odd glottalized fricatives *f'*, *s'*, *ʃ'* and *x'* as its only glottalized consonants. They alternate with plain *f*, *s*, *l* and *x*. *f'*, *ʃ'* and *x'* are extremely rare in the world's languages, usually only found in 2-3 other languages, often in NW Caucasian. *x'* occurs only in one other language - Tlingit. *s'* is slightly more common, occurring five other languages including Tlingit. In other languages, these odd sounds derived from sequences of consonant + *q*: *Cq* -> *Cʔ* -> glottalized fricative.

Sentence structure is odd:

"Hit the ball the man."

"Hit the man the ball."

"The man hit the ball."

All mean the same thing.

Huamelutec gets a **6 rating**, hardest of all.

## Karok

**Karok** is a language isolate spoken by a few dozen people in northern California. The last native speaker recently died, however, there are ~80 who have varying levels of L2 fluency.

In Karok, you can use a suffix for different types of containment - fire, water or a solid.

*pa:θ-kirih*

"throw into **a fire**"

*pa:θ-kurih*

"throw into **water**"

*pa:θ-ruprih*

"throw through **a solid**"

The suffixes are unrelated to the words for "fire", "water" and "solid".

Karok gets a **5 rating**, hardest of all.

## **Uto-Aztecan**

### **Northern**

**Hopi** is so difficult that even grammars describing the language are almost impossible to understand. For instance, Hopi has two different words for *and* depending on whether the noun phrase containing the word *and* is nominative or accusative.

Hopi is **rated 6**, hardest of all.

## **Southern Uto-Aztecan**

### **Corachol-Aztecan**

#### **Core Nahua**

#### **Nahuatl**

In Nahuatl, most adjectives are simply stative verbs.

Hence:

*Umntu omde waya eTenochtitlan.*

"The man **he is tall** went to Tenochtitlan."

"The tall man went to Tenochtitlan."

"He is tall" is a stative verb in the above.

Nahuatl gets a **6 rating**, hardest of all.

## **Numic**

### **Central Numic**

**Comanche** is legendary for being one of the hardest Indian languages of all to learn. Reasons are unknown, but all Amerindian languages are quite difficult. I doubt if Comanche is harder than other Numic languages.

Bizarrely enough, Comanche has very strange sounds called voiceless vowels, which seems to be an oxymoron, as vowels would seem to be inherently voiced. English has something akin to voiceless vowels in the words "p**articular**" and "p**eculiar**", where the bolded vowels act something akin to a voiceless vowel.

Comanche was used for a while by the codespeakers in World War 2 - not all codespeakers were Navajos. Comanche was specifically chosen

because it was hard to figure out. The Japanese were never able to break the Comanche code.

Comanche is **rated 6**, hardest of all.

## **Oto-Manguan**

### **Western Oto-Mangue**

### **Oto-Pame-Chinantecan**

### **Chinantecan**

**Chinantec**, an Indian language of southwest Mexico, is [very hard](#) for non-Chinantecs to learn. The tone system is maddeningly complex, and the syntax and morphology are very intricate.

Chinantec is **rated 6**, hardest of all.

## **Popolocan**

### **Mazatecan**

### **Lowland Valley**

### **Southern**

**Jalapa Mazatec** has distinctions between modal, creaky, breathy-voiced vowels along with nasal versions of those three. It also has creaky consonants and voiceless nasals. It has three tones, low, mid and high. Combining the tones results in various contour tones. In addition, it has a 3-way distinction in vowel length. Whistled speech is also possible.

It has a phonemic distinction between "ballistic" and "controlled" syllables which is only present on Oto-Manguean.

Ballistic (short)

*sū* "warm"  
*nī·ntū* "slippery"  
*tsā* "guava"

Controlled (half-long)

*sū·* "blue"  
*nī·ntū·* "needle"  
*tsā·* "full"  
*hɥ* "six"

Jalapa Mazatec is **rated 6**, hardest of all.

**Maipurean**

**Northern**

**Upper Amazon**

**Eastern Nawiki**

**Tariana** is a very difficult language mostly because of the unbelievable amount of information it crams into its morphology and syntax. This is mostly because it is an Arawakan language that has been heavily influenced by neighboring Tucanoan languages, with the result that it has many of the grammatical categories and particles present in both families.

This stems from the widespread bilingualism in the Vaupes Basin of Colombia, where many people grow up bilingual from childhood and often become multilingual by adulthood. Learning up to five different

languages is common. Code-switching was frowned upon and anyone using a word from Language Y while speaking Language X would get laughed at. Hence the various languages tended to borrow features from each other quite easily.

For instance, Tariana has both a noun classifier system and a gender system. Noun classifiers and gender are sometimes subsumed under the single category of "noun classifiers." Yet Tariana has both, presumably from its relationship to two completely different language families. So in Tariana is not unusual to get both demonstratives and verbs marked for both gender and noun classifier. Tariana borrowed such things as serialized perception verbs and the dubitative marker from Tucano.

In addition, Tariana has some very odd sounds, including aspirated nasals *mh* ( $m^h$ ), *nh* ( $n^h$ ) and *ñh* ( $\eta^h$ ) and an aspirated *w* ( $w^h$ ) of all things. They seem to be actually aspirated, not just partially devoiced as many voiceless nasals and liquids are.

Tariana **gets 6**, hardest of all.

## Huitotoan

### Proto-Bora-Muinane

**Bora**, a Wintotoan language spoken in Peru and Colombia near the border between the two countries, has a mind-boggling [350 different](#) noun classes. The noun classifier system is actually highly productive and is often used to create new nouns. New nouns can be created very easily, and their meanings are often semantically transparent. In some noun classifier systems, classifiers can be stacked one upon the other. In these cases, typically the last one is used for agreement purposes.

Bora also is a tonal language, but it has only two tones. In addition, nearly all consonantal phonemes have phonemic aspirated and palatalized counterparts. The agreement structure in the language is also quite convoluted. The classifier system effectively replaces much derivational morphology on the noun and noun compounding processes that other languages use to expand the meanings of

nominals.

Bora gets a **6 rating**, hardest of all.

## Tucanoan

### Eastern Tucanoan

#### Bará-Tuyuka

**Tuyuca** is a Tucanoan language spoken in by 450 people in the department of Vaupés in Colombia. An article in *The Economist* magazine concluded that it was the [hardest language on Earth to learn](#).

It has a simple sound system, but it's agglutinative, and agglutinative languages are pretty hard.

For instance, *hóabãsiriga* means "I don't know how to write." It has two forms of 1st person plural, "I and you" (inclusive) and "I and the others" (exclusive).

It has between 50-140 noun classes, including strange ones like "bark that does not cling closely to a tree", which can be extended to mean "baggy trousers or wet plywood that has begun to fall apart."

Like Yamana, a nearly extinct Amerindian language of Chile, Tuyuca marks for evidentiality, that is, how it is that you know something. For instance:

*Diga ape-wi.* "The boy played soccer." ("I saw him playing".)

*Diga ape-hiyi.* "The boy played soccer." ("I assume he was playing soccer, though I did not see it firsthand".)

Evidential marking is obligatory on all Tuyuca verbs, and it forces you to think about how you know whatever it is you know.

Tuyuca definitely gets a **6 rating**!

## Central Tucanoan

**Cubeo**, a language spoken in the Vaupes of Colombia, has a small closed class of adjective roots similar to Jul'hoan below:

<i>ira</i>	"big/large"
<i>kĩhĩ</i>	"small"
<i>bãbã</i>	"new/young"
<i>bĩki</i>	"old/great"
<i>běa</i>	"good/beautiful"
<i>ãbẽ</i>	"bad/ugly"

However, verbs can function as adjectives, and the adjective roots can either turn into nouns themselves or they can take the inflections of either nouns or verbs. Wild!

Similar to how the grammar of Tariana has been influenced by Tucano languages, the grammar of Tucanoan Cubeo has been influenced by neighboring Arawakan languages.

The grammar has been described as either SOV or OVS. That would mean that the following:

"The man the ball hit."

"The ball hit the man."

Mean the same things. OVS languages are quite rare.

Morphemes belong to one of four classes:

1. Nasal (many roots, as well as suffixes like *-xã* "Associative case")
2. Oral (many roots, as well as suffixes like *-pe* "similarity", *-du* "Frustrative case")
3. Unmarked (only suffixes, e.g. *-re* "In/Direct object")



4. Oral/Nasal (some roots and some suffixes /bã'kaxa-/mã'kaxa-  
"to defecate" and -kebã = "suppose")

Just by looking at any given consonant-initial suffix, it is impossible to determine which of the first three categories it belongs to. They must be learned one by one.

Cubeo has nasal assimilation, common to many Amazonian languages. In some of these, nasalization is best analyzed at the syllable level - some syllables are nasal and others are not.

*dĩ-bI-ko*

*/dĩ-bĩ-ko/*

**nĩmĩko**

"She recently went."

The underlying form *dĩ-bI-ko* is realized on the surface as *nĩmĩko*. The *ĩ* in *dĩ-bI-ko* nasalizes the *d*, the *b*, and the *I* on either side of it, so nasal spreading works in both directions. However, it is blocked from the third syllable because *k* is part of a class of non-nasalizable consonants.

Pretty difficult language.

Cuneo gets a **6 rating**, hardest of all.

## Carib

## Waiwai

[Hixkaryana](#) is famous for being the only language on Earth to have basic OVS (Object-Verb-Subject) word order.

The sentence *Toto yonoye kamara*, or "The man ate the jaguar", actually means "The jaguar ate the man."

*Toto yonoye kamara*

Lit. "The man ate the jaguar."

Gloss: "The jaguar ate the man."

Grammatical suffixes attached to the end of the verb mark not only number but also aspect, mood and tense.

Hixkaryána gets a **6 rating**, hardest of all.

## **Nambikwaran**

### **Mamaindê**

This is actually a series of closely related languages as opposed to one language, but the **Southern Nambikwara** language is the most well-known of the family, with 1,200 speakers in the Brazilian Amazon.

Phonology is complex. Consonants distinguish between aspirated, plain and glottalized, common in the Americas. There are strange sounds like prestopped nasals and glottalized fricatives. There are nasal vowels and three different tones. All vowels except one have both nasal, creaky-voiced and nasal-creaky counterparts, for a total of 19 vowels.

The grammar is polysynthetic with a complex evidential system.

Reportedly, Nambikwara children do not pick up the language fully until age 10 or so, one of the latest recorded ages for full competence. Nambikwara is sometimes said to be the hardest language on Earth to learn, but it has some competition.

Nambikwara definitely gets a **6 rating**, hardest of all!

## **Muran**

**Pirahã** is a language isolate spoken in the Brazilian Amazon. Recent writings by Daniel Everett indicate that not only is this one of the hardest languages on Earth to learn, but it is also one of the weirdest

languages on Earth. It is monumentally complex in nearly every way imaginable. It is commonly listed on the [rogue's gallery](#) of craziest languages and phonologies on Earth.

It has the smallest phonemic inventory on Earth with only seven consonants, three vowels and either two or three tones. Everett recently wrote a paper about it after spending many years with them. Previous missionaries who had spent time with the Pirahã generally failed to learn the language because it was too hard to learn. It took Everett a very long time, but he finally learned it well.

Many of Everett's claims about Pirahã are astounding: whistled speech, no system for counting, very few Portuguese loans (they deliberately refuse to use Portuguese loans) evidence for the Sapir-Whorf linguistic relativity hypothesis, and evidence that it violates some of Noam Chomsky's purported language universals such as embedding. It also has the  $t_{\text{ɓ}}$  sound - a bilabially trilled postdental affricate which is only found in two other languages, both in the Brazilian Amazon - Oro Win and Wari'.

Initially, Everett never heard the sound, but as they got to know him better, they started to make it more often. Everett believes that they were ridiculed by other groups when they made the odd sound.

Pirahã has the simplest kinship system in any language - there is only word for both mother and father, and the Pirahã do not have any words for anyone other than direct biological relatives.

Pirahã may have only two numerals, or it may lack a numeral system altogether.

Pirahã does not distinguish between singular and plural person. This is highly unusual. The language may have borrowed its entire pronoun set from the Tupian languages Nheengatu and Tenarim, groups the Pirahã had formerly been in contact with. This may be one of the only attested case of the borrowing of a complete pronoun set.

There are mandatory evidentiality markers that must be used in Pirahã discourse. Speakers must say how they know something, whether they saw it themselves, whether it was hearsay or whether they inferred it circumstantially.

There are various strange moods:

Desiderative: desire to perform an action

and two types of Frustrative:

Inchoative/Incomplete: frustration in starting an action

Causative/Incomplete: frustration in completing an action

There are others:

Immediate/Intensive: you are going to do something now/you intend to do it in the future

There are many verbal aspects:

Perfect/Imperfect: completed/incomplete

Telic/Atelic: reaching a goal/not reaching a goal

Continuative: continuing

Iterative: repetitive

Inchoative: beginning an action

Each [Pirahã verb has 262,144 possible forms](#), or possibly in the many millions, depending on which analysis you use.

The future tense is divided into future/somewhere and future/elsewhere. The past tense is divided into plain past and immediate past.

Pirahã has a closed class of only 90 verb roots, an incredibly small number. But these roots can be combined together to form compound verbs, a much larger category.

Here is [one example](#) of three verbs strung together to form a compound verb:

*xig ab op* "take turn go" means "bring back": "You take something away, you turn around, and you go back to where you got it to return it."

There are no abstract color terms in Pirahã. There are only two words for colors, one for "light" and one for "dark." The only other languages with this restricted of a color sense are in Papua New Guinea. The other color terms are not really color terms, but are more descriptive - "red" is translated as "like blood".

Pirahã can be whistled, hummed or encoded into music. Consonants and vowels can be omitted altogether and meaning conveyed instead via variations in stress, pitch, and rhythm. Mothers teach the language to children by repeating musical patterns.

Pirahã may well be one of the hardest languages on Earth to learn.

Pirahã gets a **6 rating**, hardest of all.

## Quechuan

**Quechua** (actually a large group of languages and not a single language at all) is one of the easiest Amerindian languages to learn. Quechua is a classic example of a highly regular grammar with few exceptions. Its agglutinative system is more straightforward than even that of Turkish. The phonology is dead simple.

On the down side, there is a lot of dialectal divergence (these are actually separate languages and not dialects) and a lack of learning materials. Some say that Quechua speakers spend their whole lives learning the language.

Quechua has inconsistent orthographies. There is a fight between those who prefer a Spanish-based orthography and those who prefer a more phonemic one. Also there is an argument over whether to use the Ayacucho language or the Cuzco language as a base.

Quechua has a difficult feature known as evidential marking. This

marker indicates the source of the speaker's knowledge and how sure they are about the statement.

*-mi* expresses personal knowledge:

*Tayta Wayllaqawaqa chufirmi.*

"Mr. Huayllacahua is a driver (I know it for a fact)."

*-si* expresses hearsay knowledge:

*Tayta Wayllaqawaqa chufirsi.*

"Mr. Huayllacahua is a driver (or so I've heard)."

*chá* expresses strong possibility:

*Tayta Wayllaqawaqa chufirchá.*

"Mr. Huayllacahua is a driver (most likely)."

Quechua is **rated 4**, very difficult.

## Aymaran

### Aymara

**Aymara** has some of the [wildest morphophonology](#) out there.

Morpheme-final vowel deletion is present in the language as a morphophonological process, and it is dependent on a set of highly complex phonological, morphological and syntactic rules (Kim 2013).

For instance, there are three types of suffixes: dominant, recessive and a 3rd class is neither dominant nor recessive. If a stem ends in a vowel, dominant suffixes delete the vowel but recessive suffixes allow the vowel to remain. The third class either deletes or retains the vowel on the stem depending on how many vowels are in the stem. If the root has two vowels, the vowel is retained. If it has three vowels, the vowel is deleted.

Although all of this seems quite odd, Finnish has something similar going on, if not a lot worse.

Nevertheless, Aymara is still said to be a very easy language to learn. The *Guinness Book of World Records* claims it is almost as easy to

learn as Esperanto.

Aymara gets a **2 rating**, very easy to learn.

## Australian

**Australian Aborigine languages** are some of the hardest languages on Earth to learn, like Amerindian or Caucasian languages. Some Australian languages have phonemic contrasts that few other languages have, such as apico-dental, lamino-dental, apico-post-alveolar, and lamino-postalveolar coronals.

Australian languages tend to be mixed ergative. Ordinary nouns are ergative-absolutive, but 1st and 2nd person pronouns are nominative-accusative. One language has a three way agent-patient-experiencer distinction in the 1st person pronoun. Australian pronouns typically have singular, plural and dual forms along with inclusive and exclusive 1st plural.

In some sentences, they have what is known as double case agreement which is rare in the world's languages:

"I gave a spear to my father."

"I gave a spear **mine-to father's-to.**"

Both elements of the phrase *my father* are in both dative and genitive.

However, Aboriginal languages do have the plus of being very regular. All Australian languages are **rated 6**, most difficult of all.

## Tor-Kwerba

### Orya-Tor

#### Tor

**Berik** is a Tor-Orya language spoken in the Indonesian colony of Irian Jaya in New Guinea.

Verbs take [many strange endings](#), in many cases mandatory ones, that indicate what time of day something happened, among other things.

*Telbener* "He drinks in the evening."

Where a verb takes an object, it will not only be marked for time of day but for the size of the object.

*Kitobana* "He gives three large objects to a man in the sunlight."

Verbs may also be marked for where the action takes place in reference to the speaker.

*Gwerantena* "To place a large object in a low place nearby."

Berik is **rated 6**, hardest of all.



## Trans New Guinea

**Madang**

**Croisilles**

**Gum**

**Amele**, spoken in Papua New Guinea, is the world's most complex language as far as verb forms go, with 69,000 finite and 860 infinitive forms.

Amele is **rated 6**, hardest of all.

**Torricelli**

**Wapei**

**Valman**

**Valman**, spoken in Papua New Guinea, is a bizarre case where the word for "and" that connects two nouns is actually a verb of all things, and is marked with the first noun as subject and the second noun as object.

"'John' (subject) and 'Mary' (object) went to the store."

"John" is marked as subject for some reason and "Mary" is marked as object, and the "and" word shows subject agreement with John and object agreement with "Mary".

Valman gets a **6 rating**, hardest of all.

## **Afroasiatic**

### **Semitic**

Semitic languages such as Arabic and Hebrew are notoriously difficult to learn, and Arabic (especially MSA) tops many language learners' lists as the hardest language they have ever attempted to learn. Although Semitic verbs are notoriously complex, the verbal system does have some advantages especially as compared to IE languages like Slavic. Unlike Slavic, Semitic verbs are not inflected for mood and there is no perfect or imperfect.

### **Central**

### **South**

### **Arabic**

**Arabic** has some very irregular manners of noun declension, even in the plural. For instance, the word "girls" changes in an unpredictable way when you say "one girl", "two girls" and "three girls", and there are two different ways to say "two girls" depending on context. "Two girls" is marked with the dual, but different dual forms can be used. All languages with duals are relatively difficult for most speakers that lack a dual in their native language. However, the dual is predictable from the singular, so one might argue that you only need to learn how to say "one girl" and "three girls".

Further, it is full of irregular plurals similar to "octopus" and "octopi" in English, whereas these forms are rare in English. With any given word, there might be 20 different ways to pluralize it, and there is no way to generalize a plural pattern from a singular pattern. In addition, many words have 2-3 ways of pluralizing them.

Some messy Arab plurals:

*kalb* -> *kilaab*  
*qalb* -> *quluub*  
*maktab* -> *makaatib*  
*taalib* -> *tullaab*  
*balad* -> *buldaan*

When you say "I love you" to a man, you say it one way, and when you say it to a woman, you say it another way. On and on.

The Arabic writing system is exceeding difficult, and is more of the hardest to use of any on Earth. Soft vowels are omitted. You have to learn where to insert missing vowels, where to double consonants and which vowels to skip in the script. There are 28 different symbols in the alphabet and four different ways to write each symbol depending on its place in the word.

Consonants are written in [different ways](#) depending on where they appear in a word. An *h* is written differently at the beginning of a word than you would write it at the end of a word. However, one simple aspect of it is that the medial form is always the same as the initial form. You need to learn not only Arabic words but also the grammar into to read Arabic.

Pronouns attach themselves to roots, and there are many different verb conjugation paradigms which simply have to be memorized. The system for measuring quantities is extremely confusing. The grammar has many odd rules that seem senseless. Unfortunately, most rules have exceptions, and it seems that the exceptions are more common than the rules themselves. Many people, including native speakers, complain about Arabic grammar.

Arabic does have case, but the system is rather simple.

The laryngeals, uvulars and glottalized sounds are hard for many foreigners to make and nearly impossible for them to get right. The *ha'* and *qa* sounds and the glottal stop in initial position give a lot of learners headaches.

Arabic is [at least as idiomatic](#) as French or English, so in order to speak it right you have to learn all of the expressionistic nuances.

One of the worst problems with Arabic is the dialects, which in many cases are separate languages altogether. If you learn Arabic, you

often have to learn one of the dialects along with Classical Arabic. All Arabic speakers speak both an Arabic dialect and Classical Arabic.

In some Arabic as a foreign language classes, even after 1 1/2 years, [not one student](#) could yet make a complete and proper sentence that was not memorized.

Adding weight to the commonly held belief that Arabic is hard to learn is research done in Germany in 2005 which showed that Turkish children learn their language at age 2-3, German children at age 4-5, but Arabic kids did not get Arabic [until age 12](#).

Arabic has complex verbal agreement with the subject, masculine and feminine gender in nouns and adjectives, head-initial syntax and a serious restriction to forming compounds. If you come from a language that has similar nature, Arabic may be easier for you than it is for so many others. Its 3 vowel system makes for easy vowels.

MSA Arabic is **rated 5**, extremely difficult.

**Arabic dialects** are often somewhat easier to learn than MSA Arabic. At least in Lebanese and Egyptian Arabic, the very difficult q' sound has been turned into a hamza or glottal stop which is an easier sound to make. Compared to MSA Arabic, the dialectal words tend to be shorter and easier to pronounce.

To attain anywhere near native speaker competency in **Egyptian Arabic**, you probably need to live in Egypt for 10 years, but Arabic speakers say that few if any second language learners ever come close to native competency. There is a huge vocabulary, and most words have a wealth of possible meanings.

Egyptian Arabic is **rated 4.5**, very to extremely difficult.

**Moroccan Arabic** is said to be particularly difficult, with much vowel elision in triconsonantal stems. In addition, all dialectal Arabic is plagued by irrational writing systems.

Moroccan Arabic is **rated 4.5**, very to extremely difficult.

**Maltese** is a strange language, basically a Maghrebi Arabic language (similar to Moroccan or Tunisian Arabic) that has very heavy influence

from non-Arabic tongues. It shares the problem of Gaelic that often words look one way and are pronounced another. It has the common Semitic problem of difficult plurals. Although many plurals use common plural endings (*-i, -iet, -ijiet, -at*), others simply form the plural by having their last vowel dropped or adding an *s* (English borrowing). There's no pattern, and you simply have to memorize which ones act which way. Maltese permits the consonant cluster *spt*, which is surely hard to pronounce.

On the other hand, Maltese has quite a bit of IE loans from Italian, Sicilian, Spanish, French and increasingly English. If you have knowledge of Romance languages, Maltese is going to be easier than most Arabic dialects.

Maltese is **rated 4**, very difficult.

## South

### Canaanite

**Hebrew** is hard to [learn](#) according to a number of Israelis. Part of the problem may be the abjad writing system, which often leaves out vowels which must simply be remembered. Also, other than borrowings, the vocabulary is Afroasiatic, hence mostly unknown to speakers of IE languages. There are also difficult consonants as in Arabic such as pharyngeals and uvulars.

The *het* or glottal *h* is particularly hard to make. However, most modern Israelis no longer make the *het* sound or *a'ain* sounds. Instead, they pronounce the *het* like the *chaf* sound and the *a'ain* like an *alef*. Almost all Ashkenazi Israeli Jews no longer use the *het* or *a'ain* sounds. But most Jews who came from Arab countries (often older people) still use the sound, and some of their children do (Dorani 2013).

Hebrew has complex morphophonological rules. The letters *p, b, t, d, k* and *g* change to *v, f, dh, th, kh* and *gh* in certain situations. In some environments, pharyngeals change the nature of the vowels around them. The prefix *ve-*, which means "and", is pronounced differently when it precedes certain letters. Hebrew is also quite

irregular.

Hebrew has quite a few voices, including active, passive, intensive, intensive passive, etc. It also has a number of tenses such as present, past and the odd jussive.

Hebrew also has two different noun classes. There are also many suffixes and quite a few prefixes that can be attached to verbs and nouns.

Even most native Hebrew speakers do not speak Hebrew correctly by a long shot.

Quite a few say Hebrew is [as hard to learn](#) as MSA or perhaps even harder, but this is controversial.

Hebrew gets a **5 rating** for extremely difficult.

## **Berber**

### **Northern**

#### **Atlas**

Berber languages are considered to be very hard to learn. Worse, there are very few language learning resources available.

**Tamazight** allows doubled consonants at the beginning of a word! How can you possibly make that sound?

Tamazight gets a **6 rating**, hardest of all.

In **Tachelhit**, words like this are possible:

*tkkststt*

"You took it off."

*tfkststt*

"You gave it."

In addition, there are words which contain only one or two consonants:

g  
"be"

ks  
"feed on"

Tachelhit gets a **6 rating**, hardest of all.

**South**

**Ethiopian**

**South**

**Transversal**

**Amharic–Argobba**

**Amharic**

**Amharic** is said to be a very hard language to learn. It is quite complex and its sentence structures seem strange even to speakers of other Semitic languages. Hebrew speakers say they have a hard time with this language. There are a multitude of rules which almost seem ridiculous in their complexity, there are numerous conjugation patterns, objects are suffixed to the verb, the alphabet has 274 letters, and the pronunciation seems strange. However, if you already know Hebrew or Arabic, it will be a lot easier. The hardest part of all is the verbal system, as with any Semitic language. It is easier than Arabic.

Amharic gets a **4.5 rating**, very hard to extremely hard.

# Cushitic

## East Cushitic

**Dahalo** is legendary for having some of the wildest consonant phonology on Earth. It has all four airstream mechanisms found in languages: ejectives, implosives, clicks and normal pulmonic sounds. There are both glottal and epiglottal stops and fricatives and laminal and apical stops.

There is also a strange series of nasal clicks and are both glottalized and plain. Some of these clicks are also labialized. It has both voiced and unvoiced prenasalized stops and affricates, and some of the stops are also labialized. There is a weird palatal lateral ejective. There are three different lateral fricatives, including a labialized and palatalized one, and one lateral approximant. It contrasts alveolar and palatal lateral affricates and fricatives, the only language on Earth to do this.

The Dahalo are former elephant hunting hunter-gatherers who live in northern Kenya. It is believed that at one time they spoke a language like Sandawe or Hadza, but they switched over to Cushitic at some point. The clicks are thought to be substratum from a time when Dahalo was a Sandawe-Hadza type language.

Dahalo gets a **6 rating**, hardest of all.

## Somali

**Somali** has one of the strangest preposition systems on Earth. It actually has no real prepositions at all. Instead it has [preverbal particles and possessives](#) that serve as prepositions.

Here is how possessives serve as prepositions:

*habeennimada horteeda*  
"the night her front"  
"before nightfall"



*kulaylka dartiisa*

"the heat his reason"

"because of the heat"

Here we have the use of a preverbal particle serving as a preposition:

*kú ríd shandádda*

"Into put the suitcase."

"Put it into the suitcase."

Somali combines four "prepositions" with four deictic particles to form its prepositions. There are four basic "prepositions": "to", "in", "from" and "with".

These combine with a four different deictic particles:

"toward the speaker"

"away from the speaker"

"toward each other"

"away from each other"

Hence you put the "prepositions" and the deictic particles together in various ways. Both tend to go in front of and close to the verb:

*Nínkíi bàan cèelka xádhig kagá sóo saaray.*

"...well-the rope with-from towards-me I-raised."

"I pulled the man out of the well with a rope."

*Way inoogá warrámi jireen.*

"They us-to-about news gave."

"They used to give us news about it."

Prepositions are the hardest part of the Somali language for the learner.

Somali deals with verbs of motion via deixis in a similar way that

Georgian does. One reference point is the speaker, and the other is any other entities discussed. Verbs of motion are formed using adverbs. Entities may move:

*wada* "towards each other"

*kala* "away from each other"

*so* "towards the speaker"

*si* "away from the speaker"

Hence:

*kala durka* "separate"

*si gal* "go in (away from the speaker)"

*so gal* "come in (toward the speaker)"

Somali lacks orthographic consistency. There are four different orthographic systems in use - the [Wadaad Arabic script](#), the [Osmanya](#) Ethiopic script, the [Borama](#) script and the Latin [Somali alphabet](#), the current system.

All of the difficult sounds of Arabic are also present in Somali, another Semitic language - the *alef*, the *ha*, the *qaf* and the *kha*. There are long and short vowels. There is a retroflex *d*, the same sound found in South Indian languages. Somali also has 2 tones - high and low. For some reason, Somali tends to make it onto craziest phonologies [lists](#).

Somali pluralization makes no sense and must be memorized. There are seven different plurals, and there is no clue in the singular that tells you what form to use in the plural. See here:

Republication:

*áf* "language" -> *afaf* "languages"

Suffixation:

*hoóyo* "mother" -> *hoyoóyin* "mothers"

Note the tone shifts in all three of the plurals above.

There are [four cases](#), absolutive, nominative, genitive, and vocative. Despite the presences of absolutive and nominative cases, Somali is not an ergative language. Absolutive case is the basic case of the noun, and nominative is the case given to the noun when a verb follows in the sentence. There are different articles depending on whether the noun was mentioned previously or not (similar to the articles *a* and *the* in English). The absolutive and nominative are marked not only on the noun but also on the article that precedes it.

In terms of [difficulty](#), Somali is much harder than Persian and probably about as difficult as Arabic.

Somali gets a **5 rating**, extremely hard to learn.

## **Dravidian**

### **Southern**

#### **Tamil-Kannada**

#### **Tamil-Kodagu**

#### **Tamil-Malayalam**

### **Malayalam**

**Malayalam**, a Dravidian language of India, has been cited as the hardest language to learn by a language foundation, but the citation is obscure and hard to verify.

Malayalam words are often even hard to look up in a Malayalam dictionary.

For instance, *adiyAnkaLAKkikkoNDirikkukayumANello* is a word in Malayalam. It means something like "I, your servant, am sitting and mixing something (which is why I cannot do what you are asking of me)." The part in parentheses is an example of the type of sentence where it might be used.

The above word is composed of many different morphemes, including conjunctions and other affixes, with sandhi going on with some of them so they are eroded away from their basic form. There doesn't seem to be any way to look that word up or to write a Malayalam dictionary that lists all the possible forms, including forms like the word above. It would probably be way too huge of a book. However, all agglutinative languages are made up of affixes, and if you know the affixes, it is not particularly hard to parse the word apart.

Malayalam is said to be very hard to pronounce correctly.

Further, few foreigners even try to learn Malayalam, so Malayalam speakers, like the French, [might not listen to you and might make fun of you](#) if your Malayalam is not native sounding.

However, Malayalam has the advantage of having many pedagogic materials available for language learning such as audio-visual material and subtitled videos.

Malayalam is **rated 5**, extremely difficult.

## Tamil

**Tamil**, a Dravidian language is hard, but probably not as difficult as Malayalam is. Tamil has an incredible 247 characters in its alphabet. Nevertheless, most of those are consonant-vowel combinations, so it is almost more of a syllabary than an alphabet. Going by what would traditionally be considered alphabetic symbols, there are probably only 72 real symbols in the alphabet. Nevertheless, Tamil probably has one of the easier Indic scripts as Tamil has fewer characters than other scripts due to its lack of aspiration. Compare to Devanagari's over 1,000 characters.

But no Indic script is easy. A problem with Tamil is that all of the characters seem to look alike. It is even worse than Devanagari in

that regard. However, the more rounded scripts such as Kannada, Sinhala, Telegu, and Malayalam have that problem to a worse degree. Tamil has a few sharp corners in the characters that helps to disambiguate them.

In addition, as with other languages, words are written one way and pronounced another. However, there are claims that the difficulty of Tamil's diglossia is overrated.

Tamil has two different registers for written and spoken speech, but the differences are not large, so this problem is exaggerated. Both Tamil and Malayalam are spoken very fast and have extremely complicated, nearly impenetrable scripts. If Westerners try to speak a Dravidian language in south India, more often than not the Dravidian speaker will simply address them in English rather than try to accommodate them.

Tamil has the odd evidential mood, similar to Bulgarian.

However, on the plus side, the language does seem to be very logical and regular, almost like German in that regard. In addition, there are a lot of language learning materials for Tamil.

Tamil is **rated 4**, very difficult.

## **Altaic**

### **Korean**

Most agree that **Korean** is a hard language to learn.

The alphabet, Hangul at least is reasonable; in fact, it is quite elegant. But there are four different Romanizations- Lukoff, Yale, Horne, and McCune-Reischauer - which is preposterous. It's best to just blow off the Romanizations and dive straight into Hangul. This way you can learn a Romanization later, and you won't mess up your Hangul with spelling errors, as can occur if you go from Romanization to Hangul.

Hangul can be learned very quickly, but learning to read Korean books and newspapers fast is another matter altogether because you really need to know the hanja or Chinese character that are used in addition

to the Hangeul. After World War 2, the Koreans decided to officially [get rid of](#) their Chinese characters, but in practice this was not successful. With the use of Chinese characters in Korean, you can be a lot more precise in terms what you are trying to communicate.

Bizarrely, there are two different numeral sets used, but one is derived from Chinese so it should be familiar to Chinese, Japanese or Thai speakers who use similar or identical systems.

Korean has a wealth of homonyms, and this is one of the tricky aspects of the language. Any given combination of a couple of characters can have multiple meanings. Japanese has a similar problem with homonyms, but at least with Japanese you have the benefit of kanji to help you tell the homonyms apart. With Korean Hangeul, you get no such advantage.

Similarly, there seem to be many ways to say the same thing in Korean. The learner will feel when people are using all of these different ways of saying the same thing that they are actually saying something different each time, but that is not the case.

One problem is that the *b*, *p*, *j*, *ch*, *t*, and *d* are pronounced differently than their English counterparts. The consonants, the *pachim* system, and the morphing consonants at the end of the word that slide into the next word make Korean harder to pronounce than any major European language. Korean has a similar problem with Japanese, that is, if you mess up one vowel in sentence, you render it incomprehensible.

The vocabulary is very difficult for an English speaker who does not have knowledge of either Japanese or Chinese. On the other hand, Japanese or Chinese will help you a lot with Korean.

Korean is agglutinative and has a subject-topic discourse structure, and the logic of these systems is difficult for English speakers to understand. In addition, there are hundreds of ways of conjugating any given verb based on tense, mood, age or seniority. Adjectives also decline and take hundreds of different suffixes.

Meanwhile, Korean has an honorific system that is even wackier than that of Japanese. A single sentence can be said in three different ways depending on the relationship between the speaker and the listener. However, the younger generation is not using the honorifics so much, and a foreigner isn't expected to know the honorific system anyway.

Maybe 60% of the words are based on Chinese words, but unfortunately, much of this Chinese-based vocabulary intersects with Japanese versions of Chinese words in a confusing way.

Speakers of Korean can learn Japanese fairly easily. Korean seems to be a more difficult language to learn than Japanese. There are maybe twice as many particles as in Japanese, the grammar is dramatically more difficult and the verbs are quite a bit harder. The phonemic inventory in Korean is also larger and includes such oddities as double consonants.

Korean is rated by language professors as being one of the hardest languages to learn.

Korean is **rated 5**, extremely hard.

## Japonic

**Japanese** also uses a symbolic alphabet, but the symbols themselves are sometime undecipherable in that even Japanese speakers will sometimes encounter written Japanese and will say that they don't know how to pronounce it. I don't mean that they mispronounce it; that would make sense. I mean they don't have the slightest clue how to say the word! This problem is essentially nonexistent in a language like English.

The Japanese orthography is one of the most difficult to use of any orthography.

There are over 2,000 frequently used characters in three different symbolic alphabets that are frequently mixed together in confusing ways. Due to the large number of frequently used symbols, it's said that even Japanese adults learn a new symbol a day a ways into adulthood.

The Japanese writing system is probably crazier than the Chinese writing system and it often makes it onto lists of worst orthographies. The very idea of writing an agglutinative language in a combination of two syllabaries and an ideography seems wacky right off the bat. Japanese borrowed Chinese characters.

But then they gave each character several pronunciations, and in

some cases as many as 24. Next they made two syllabaries using another set of characters, then over the next millennia came up with all sorts of contradictory and often senseless rules about when to use the syllabaries and when to use the character set. Later on they added a Romanization to make things even worse.

Chinese uses 5-6,000 characters regularly, while Japanese only uses around 2,000. But in Chinese, each character has only one or maybe two pronunciations. In Japanese, there are complicated rules about when and how to combine the hiragana with the characters. These rules are so hard that many native speakers still have problems with them. There are also personal and place names (proper nouns) which are given completely arbitrary pronunciations often totally at odds with the usual pronunciation of the character.

There are some writers, typically of literature, who deliberately choose to use kanji that even Japanese people cannot read. For instance, Ryu Murakami uses the odd symbols 撥る、, 轢く、and 憑け.

The Japanese system is made up of three different systems: the katakana and hiragana (the kana) and the kanji, similar to the hanzi used in Chinese. Chinese has at least 85,000 hanzi. The number of kanji is much less than that, but kanji often have more than one meaning in contrast to hanzi.

After WW2, Japan decided to [simplify](#) its language. They both simplified and reduced the number of Chinese characters used, and they unified the written and spoken language, which previously had been different.

Speaking Japanese is not as difficult as everyone says, and many say it's fairly easy. However, there is a problem similar to English in that one word can be pronounced in multiple ways, like *read* and *read* in English.

A common problem is that a perfectly grammatically correct sentence uttered by a Japanese language learner, while perfectly correct, is still not acceptable by Japanese speakers because "we just don't say it that way." The Japanese speaker often cannot tell why the unacceptable sentence you uttered is not ok. On the other hand, this problem may be common to more languages than Japanese.

There is also a class of Japanese called "honorifics" or "keigo" that is quite hard to master. Honorifics are meant to show respect and to



indicate one's place or status in the social hierarchy. These typically affect verbs but can also affect particles and prefixes. They are usually formed by archaic or highly irregular verbs. However, there are both regular and irregular honorific forms. Furthermore, there are five different levels of honorifics. Honorifics vary depending on who you are and who you are talking to. In addition, gender comes into play.

Although it is true the Japanese young people are said to not understand the intricacies of keigo, it is still expected that they know how to speak this well. Consequently, many young Japanese will opt out of certain conversations because they feel that their keigo is not very good. Books explaining how to use keigo properly have been big sellers among young people in Japan in recent years, as young people try to appear classy, refined, or cultured.

In addition, Japanese born overseas (especially in the US), while often learning Japanese pretty well, typically have a very poor understanding of keigo. Instead of embarrassing themselves by not using keigo or using it wrong, these Japanese speakers often prefer to speak in English to Japanese people rather than bother with keigo-less Japanese. Overcorrection in keigo is also a problem when hypercorrection leads to someone making errors in keigo due to "trying too hard." This looks like phony or insincere politeness and is often worse than not using keigo at all.

One wild thing about Japanese is counting forms. You actually use different numeral sets depending on what it is you are counting! There are dozens of different ways of counting things which involve the use of a complex numerical noun classifier system.

Japanese grammar is often said to be simple, but that does not appear to be the case on closer examination. Particles are especially vexing. Verbs engage in all sorts of wild behavior, and adverbs often act like verbs. Nouns can act like adjectives and adverbs. Meanwhile, honorifics change the behavior of all words. There are [particles](#) like *ha* and *ga* that have many different meanings. One problem is that all noun modifiers, even phrases, must precede the nouns they are modifying.

It's often said that Japanese has no case, but this is not true. Actually, there are [seven cases](#) in Japanese. The aforementioned *ga* is a clitic meaning nominative, *made* is terminative case, *-no* is genitive and *-o*

is accusative.

In this sentence:

"The plane that was supposed to arrive at midnight, but which had been delayed by bad weather, finally arrived at 1 AM."

Everything underlined must precede the noun "plane":

"Was supposed to arrive at midnight, but had been delayed by bad weather, the plane finally arrived at 1 AM."

One of the main problems with Japanese grammar is that it is going to seem so different from the sort of grammar and English speaker is likely to be used to.

Speaking Japanese is one thing, but reading and writing it is a whole new ballgame. It's perfectly possible to know the meaning of every kanji and the meaning of every word in a sentence, but you still can't figure out the meaning of the sentence because you can't figure out how the sentence is stuck together in such a way as to create meaning.

The real problem is that the Japanese you learn in class is one thing, and the Japanese of the street is another. One problem is that in street Japanese, the subject is typically not stated in a sentence. Instead it is inferred through such things as honorific terms or the choice of words you used in the sentence. Probably no one goes crazier on negatives than the Japanese. Particularly in academic writing, triple and quadruple negatives are common, and can be quite confusing.

Yet there are problems with the agglutinative nature of Japanese. It's a completely different syntactic structure than English. Often if you translate a sentence from Japanese to English it will just look like a meaningless jumble of words.

However, Japanese grammar has the advantage of being quite regular. For instance, there are only four frequently used irregular verbs.

Like Chinese, the nouns are not marked for number or gender.

However, while Chinese is forgiving of errors, if you mess up one vowel in a Japanese sentence, you may end up with incomprehension. Although many Japanese learners feel it's fairly easy to learn, surveys of language professors continue to rate Japanese as one of the [hardest languages](#) to learn. A [study by the US Navy](#) concluded that the hardest language the corpsmen had to learn in the course of service was Japanese. However, it's generally agreed that Japanese is easier to learn than Korean. Japanese speakers are able to learn Korean pretty easily.

Japanese is **rated 5**, extremely hard.

**Classical Japanese** is much harder to read than Modern Japanese. Though you can get by with much less kanji when reading the modern language, you will need a minimum knowledge of 3,000 kanji for reading Classical Japanese, and that's using a dictionary. There are only about 500-1,000 frequently used characters, but there are countless other words that will come up in your reading, for instance, special words used in the Imperial Court. Many words have more than one meaning, and unless you know this, you will be lost. 東宮 for instance means "Eastern Palace". However, it also means "Crown Prince" because his residence was to the east of the Emperor's.

The movie *The Seven Samurai* (set in the late 1500's) seems to use some sort of Classical Japanese, or at least [Classical vocabulary and syntax with modern pronunciation](#). Japanese language learners say they can't understand a word of the archaic Japanese used in this movie.

Classical Japanese **gets 5.5**, nearly hardest of all.

## Turkic

### Oghuz

#### Western Oghuz

**Turkish** is often considered to be [hard to learn](#), and it's rated one of the hardest in surveys of language teachers, however, it's probably

easier than its reputation made it out to be.

It is agglutinative, so you can have one long word where in English you might have a sentence of shorter words. One word is:

*Çekoslovakyalılastıramadıklarımızdanmissiniz?*

"Were you one of those people whom we could not turn into a Czechoslovakian?"

Many words have more than one meaning.

However, the agglutination is very regular in that each particle of meaning has its own morpheme and falls into an exact place in the word. See here:

<i>göz</i>	"eye"
<i>göz-lük</i>	"glasses"
<i>göz-lük-çü</i>	"optician"
<i>göz-lük-çü-lük</i>	"the business of an optician"

Nevertheless, agglutination means that you can always create new words or add new parts to words, and for this reason even a lot of Turkish adults have problems with their language.

There is no verb "to be", which is hard for many foreigners. Instead, the concept is wrapped onto the subject of the sentence as a *-dim* or *-im* suffix. Turkish is an imagery-heavy language, and if you try to translate straight from a dictionary, it often won't make sense.

However, the suffixation in Turkish, along with the vowel harmony, are both precise. Nevertheless, many words have irregular vowel harmony. The rules for making plurals are very regular, with no exceptions (the only exceptions are in foreign loans). In Turkish, incredible as it sounds, you can make a plural out of anything, even a word like "what", "who", or "blood". However, there is some irregularity in the strengthening of adjectives, and the forms are not predictable and must be memorized.

Turkish is a language of precision in other ways. For instance, there are eight different forms of subjunctive mood that describe various degrees of uncertainty that one has about what one is talking about. This relates to the evidentiality discussed under Tuyuca above, and Turkish has an evidential form similar to Tamil and Bulgarian. On Turkish news, verbs are generally marked with *miş*, which means that the announcer believes it to be true though he has not seen it firsthand. The particle *miş* is interesting because this evidential form is [coded into the tense system](#), which is an unusual use of evidentiality.

The Roman alphabet and almost mathematically precise grammar really help out. Turkish lacks gender and has but a single irregular verb - *olmak*. Nevertheless, there are many verbal forms. However, this is controversial and it depends on how you define grammatical irregularity. There is strangeness in some of the verb paradigms, but it is argued that these oddities are rule-based. The aorist tense is [said to have irregularity](#).

There is some irregular morphophonology, but not much. The oblique relative clauses have complex morphosyntax. Turkish has two completely different ways of making relative clauses, one of which may have been borrowed from Persian. There are many gerunds for verbs, and these have many different uses. At the end of the day, Turkish grammar is not as regular or as simple as it is made out to be.

Words are pronounced nearly the same as they are written. A suggestion that Turkish may be easier to learn than many think is the research that shows that Turkish children learn [attain basic grammatical mastery](#) of Turkish at age 2-3, as compared to 4-5 for German and 12 for Arabic. The research was conducted in Germany in 2005.

In addition, Turkish has a phonetic orthography.

However, Turkish is hard for an English speaker to learn for a variety of reasons.

It is agglutinative like Japanese, and all agglutinative languages are difficult for English speakers to learn. As in Japanese, you start your Turkish sentence the way you would end your English sentence. As in the Japanese example above, the subordinate clause must precede

the subject, whereas in English, the subordinate clause must follow the subject.

The italicized phrase below is a subordinate clause.

In English, we say, "**I hope** that he will be on time."

In Turkish, the sentence would read, "That he will be on time **I hope.**"

Turkish vowels are unusual to speakers of IE languages, and Turkish learners say the vowels are hard to make or even tell apart from one another.

Turkish is **rated 3.5**, harder than average to learn.

## Uralic

### Finno-Ugric

One test of the difficulty of any language is how much of the grammar you must know in order to express yourself on a basic level. On this basis, Finno-Ugric languages are complicated because you need to know quite a bit more grammar to communicate on a basic level in them than in say, German.

## Finnic

### Northern

**Finnish** is very hard to learn, and even long-time learners often still have problems with it. Famous polyglot Barry Farber said it was [one of the hardest](#) languages he learned. You have to know exactly which grammatical forms to use where in a sentence.

In addition, Finnish has 15 cases in the singular and 16 in the plural. This is hard to learn for speakers coming from a language with little or

no case.

For instance:

*talo* - "the house"

Cases:

*talon* "house's"  
*taloasome* "of the house"  
*taloksiinto* "as the house"  
*talostafrom* "inside the house"  
*talollaon* "to the house"  
*taloltafrom* "beside the house"  
*taloistafrom* "the houses"  
*taloissa* "in the houses"

It gets much worse than that. This [web page](#) shows that the noun *kauppa* "shop" can have 2,253 forms.

A simple adjective + noun type of noun phrase of two words can be conjugated in up to 100 different ways.

Adjectives and nouns belong to 20 different classes. The rules governing their case declension depend on what class the substantive is in.

As with Hungarian, words can be very long. [For instance:](#)

*lentokonesuihkuratbiinimoottoriapumekaanikkoaliupseerioppilas*  
"non-commissioned officer cadet learning to be an assistant mechanic for airplane jet engines"

Like Turkish, Finnish agglutination is very regular. Each bit of information has its own morpheme and has an exact place in the word.

Like Turkish, Finnish has vowel harmony, but the vowel harmony is very regular like that of Turkish. Unlike Turkish or Hungarian, consonant gradation forms a major part of Finnish morphology. In order to form a sentence in Finnish, you will need to learn about verb types, cases and consonant gradation, and it can take a while to get your mind around those things.

Finnish, oddly enough, always puts the stress on the first syllable. Finnish vowels will be hard to pronounce for most foreigners.

However, Finnish has the advantage of being pronounced precisely as it is written. This is also part of the problem though, because if you don't say it just right, the meaning changes. So, similarly with Polish, when you mangle their language, you will only achieve incomprehension. Whereas with say English, if a foreigner mangles the language, you can often winnow some sense out of it.

However, despite that fact that written Finnish can be easily pronounced, when learning Finnish, as in Korean, it is as if you must learn [two different languages](#) - the written language and the spoken language. A better way to put it is that there is "one language for writing and another for speaking." You use different forms whether conversing or putting something on paper.

Some pronunciation is difficult. The [contrast](#) between short and long vowels and consonants is particularly troublesome. Check out these minimal pairs:

*sydä**m**ellä*  
*sydä**mm**ellä*

*jolle**k**in*  
*jolle**kk**in*

A problem for the English speaker coming to Finnish would be the vocabulary, which is alien to the speaker of an IE language. Finnish language learners often find themselves looking up over half the words they encounter. Obviously, this slows down reading quite a bit!

In the grammar, the partitive case and potential tense can be difficult.

Here is an example of how Finnish verb tenses combine with various cases to form words:



### I A-Infinitive

Base form            *mennä*

### II E-Infinitive

Active inessive     *mennessä*

Active instructive *mennen*

Passive inessive    *mentäessä*

### III MA-Infinitive

Inessive             *menemässä*

Elative              *menemästä*

Illative              *menemään*

Adessive             *menemällä*

Abessive             *menemättä*

Active instructive *menemän*

Passive instructive *mentämän*

## Verbs in Finnish

Finnish verbs are [very regular](#). The irregular verbs can almost be counted on one hand:

*juosta*

*käydä*

*olla*

*nähdä*

*tehdä*

and a few others. In fact, on the plus side, Finnish in general is [very regular](#).

One easy aspect of Finnish is the way you can build [many forms](#) from a base root:

*kirj-*

*kirja*            "book"

*kirje* "letter"  
*kirjoittaa* "to write"  
*kirjailija* "writer"

As in many Asian languages, there are no masculine or feminine pronouns, and there is no grammatical gender. The numeral system is quite simple compared to other languages. Finnish has a [complete lack](#) of consonant clusters. In addition, the phonology is fairly simple. Finnish is **rated 5**, extremely hard to learn.

## Southern

**Estonian** has similar difficulties as Finnish, since they are closely related. However, Estonian is more irregular than Finnish. In particular, the very regular agglutination system described in Finnish seems to have gone awry in Estonian. Estonian has [14 cases](#), including strange cases such as the abessive, adessive, elative and inessive. On the other hand, all of these cases can simply be analyzed as the genitive case plus a single unvarying suffix for each case. In addition, there is no gender, so the only things you have to worry about when forming cases are singular and plural.

Estonian has a strange mood form called the [quotative](#), often translated as "reported speech."

*tema on* "he/she/it is"

*tema olevat* "it's rumored that he/she/it is or he/she/it is said to be"

This mood is often used in newspaper reporting and is also used for gossip.

Estonian has an astounding 25 diphthongs. It also has three different varieties of vowel length, which is strange in the world's languages.

There are short and extra-long vowels and consonants.

*lina* "linen" short *n*

*linna* "the town's" long *n*, written as *nn*

*`linna* "into the town" extra-long *n*, not written out!

There are differences in the pronunciation of the three forms above, but in rapid speech, they are hard to hear, though native speakers can make them out.

Difficulties are further compounded in that extra-long sonorants (*m*, *n*, *ng*, *l*, and *r*) and vowels are not written out. All in all, phonemic length can be a [problem](#) in Estonian, and foreigners never seem to get it completely down.

Estonian pronunciation is not very difficult, though the *õ* sound can cause problems. However, Estonian has completely lost the vowel harmony system it inherited from Finnish, resulting in words that seem very hard to pronounce.

At least in written form, Estonian is not as complex as Finnish. Estonian can be seen as an abbreviated and modernized form of Finnish. The grammar is also like a simplified version of Finnish grammar and may be much easier to learn.

Estonian is **rated 4.5**, very to extremely difficult.

## **Sami**

### **Eastern**

**Skolt Sami's** Latinization is often listed as one of the [worst Latinizations](#) around. The rest of the language is quite similar to, and as difficult as, Finnish.

Skolt Sami gets a **5 rating**, extremely hard to learn.

# Ugric

## Hungarian

It's widely agreed that **Hungarian** is one of the hardest languages on Earth to learn. Even language professors agree. The British Diplomatic Corps did a study of the languages that its diplomats commonly had to learn and concluded that [Hungarian was the hardest](#). Hungarian grammar is maddeningly complex, and Hungarian is [often listed](#) on craziest grammar lists. For one thing, there are many different forms for a single word via word modification. This enables the speaker to make his intended meaning very precise. Looking at nouns, there are about 257 different forms per noun.

Hungarian is said to have from 24-35 different cases (there are charts available showing 31 cases), but the actual number may only be 18. Nearly everything in Hungarian is inflected, similar to Lithuanian or Czech. Similar to Georgian and Basque, Hungarian has the polypersonal agreement, albeit [to a lesser degree](#) than those two languages. There are many irregularities in inflections, and even Hungarians have to learn how to spell all of these in school and have a hard time learning this.

The case distinctions alone can create many different words out of one base form.

For the word "house", we end up with 31 different words using case forms:

<i>házba</i>	"into the house"
<i>házban</i>	"in the house"
<i>házból</i>	"from [within] the house"
<i>házra</i>	"onto the house"
<i>házon</i>	"on the house"
<i>háziról</i>	"off [from] the house"
<i>házhoz</i>	"to the house"
<i>házig</i>	"until/up to the house"
<i>háznál</i>	"at the house"

*háztól* "[away] from the house"  
*házzá* Translative case, where the house is the end product of a transformation, such as "They turned the cave into a house."  
*házként* "as the house", which could be used if you acted in your capacity as a house or disguised yourself as one. "He dressed up as a house for Halloween."  
*házáért* "for the house", specifically things done on its behalf or done to get the house. "They spent a lot of time fixing things up for the house".  
*házul* Essive-modal case. Something like "house-ly" or "in the way/manner of a house". "The tent served as a house (in a house-ly fashion)".

And we do have some basic cases:

*ház* Nominative "The house is down the street."  
*házat* Accusative "The ball hit the house."  
*háznak* Dative "The man gave the house to Mary."  
*házzal* Instrumental Similar to the standard Instrumental Case used by many languages, but more similar to English "with". Refers to both instruments and companions.

The genitive takes 12 different declensions, depending on person and number:

*házam* "my house"  
*házaim* "my houses"  
*házad* "your house"  
*házaid* "your houses"  
*háza* "his/her/its house"  
*házai* "his/her/its houses"  
*házunk* "our house"  
*házaink* "our houses"  
*házatok* "your house"  
*házaitok* "your house"  
*házuk* "their house"  
*házaik* "their houses"

*egyház* "church", as in "the Catholic Church". (Literally "one-house")

In addition, the genitive suffixes to the possession, which is not how the genitive works in IE.

*ember* "man/person"

*ház* "house"

*a(z)* "the"

*az ember háza* "the man's house" Lit. "the man house-his"

*a házam* "my house" Lit. "the house-my"

*a házad* "your house" Lit. "the house-your"

There are also very long words such as this:

*megszentségteleníthetetleniségeskedéseitekért...*

"for your (you all possessive) repeated pretensions at being impossible to desecrate..."

Being an agglutinative language, that word is made up of many small parts of words, or morphemes. That word means something like

The preposition is stuck onto the word in this language, and this will seem strange to speakers of languages with free prepositions.

Hungarian is full of synonyms, similar to English.

For instance, there are 108 different words that mean *to move*:

*halad, jár, megy, dülöngél, lépdél, botorkál, kódorog, sétál, andalog, rohan, csörtet, üget, lohol, fut, átvág, vágat, tipeg, libeg, biceg, poroszkál, vágtazik, somfordál, bóklászik, szedi a lábát, kitér, elszökken, betér, botladozik, ügyeleg, slattyog, bandukol, lófrál, szalad, vánszorog, kószál, kullog, baktat, koslat, kaptat, császkaál, totyog, suhan, robog, rohan, kocog, cselleng, csatangol, beszlisszol,*

*elinal, elillan, bitangol, lopakodik, sompolyog, lapul, elkotródik, settenkedik, sündörög, eltérül, elődalog, kóborol, lézeng, ődöng, csavarog, lődörög, elvándorol, tekereg, kóvályog, ténfereg, özönlik, tódul, vonul, hömpölyög, ömlik, surran, oson, lépeget, mozog, ballag, sprintel, galoppol, nyargal, trappol, cammog, caplat, kutyagol, viharzik, ólálkodik, sunnyog, slattyog, kolbászol, siet, iszkol, száguld, iramodik, mászkál, spurizik, jár-ke, kószál, tévelyeg, csámborog, sétafikál, kujtorog, barangol, kóricál, mendegél, flangál, mászkál, császkál and mozgolódik.*

Only about five of those terms are archaic and seldom used, the rest are in current use.

In addition, while most languages have names for countries that are pretty easy to figure out, in Hungarian even languages of nations are hard because they have changed the names so much. "Italy" becomes *Olazország*, "Germany" becomes *Nemetország*, etc.

As in Russian and Serbo-Croatian, word order is relatively free in Hungarian. It is not completely free as some say but rather is it governed by a set of rules. The problem is that as you reorder the word order in a sentence, you say the same thing but the meaning changes slightly in terms of nuance. Further, there are quite a few dialects in Hungarian. Native speakers can pretty much understand them, but foreigners often have a lot of problems. Accent is very difficult in Hungarian due to the bewildering number of rules used to determine accent. In addition, there are exceptions to all of these rules. Nevertheless, Hungarian is probably more regular than Polish.

Hungarian spelling is also very strange for non-Hungarians, but at least the orthography is phonetic. Nevertheless, the orthography often makes it onto [worst orthographies](#) lists.

Hungarian phonetics is also strange. One of the problems with Hungarian phonetics is vowel harmony. Since you stick morphemes together to make a word, the vowels that you have used in the first part of the word will influence the vowels that you will use to make up the morphemes that occur later in the word. The vowel harmony gives Hungarian a "singing effect" when it is spoken. The *ty, ny, sz, zs, dzs, dz, ly, cs* and *gy* sounds are hard for many foreigners to make. The *á, é, ó, ö, ő, ú, ü, ű, and í* vowel sounds are not found in English.

Verbs are marked for object (indefinite, definite and person/number), subject (person and number) tense (past, present and future), mood (indicative, conditional and imperative), and aspect (frequency, potentiality, factitiveness, and reflexiveness).

*Elmentegegethetnék.*

"I could make others save you occasionally (on a disk)."

Verbs change depending on whether the object is definite or indefinite.

*Olvasok könyvet.*

"I read **a** book." (indefinite object)

*Olvasom a könyvet.*

"I read **the** book." (definite object)

As noted in the introduction to the Finno-Ugric section, you need to know quite a bit of Hungarian grammar to be able to express yourself on a basic level. For instance, in order to say:

"I like your sister."

you will need to understand the following Hungarian forms:

1. verb conjugation and definite or indefinite forms
2. possessive suffixes
3. case
4. how to combine possessive suffixes with case
5. word order
6. explicit pronouns
7. articles

It's hard to say, but Hungarian is probably [harder to learn](#) than even the hardest Slavic languages like Czech, Serbo-Croatian and Polish. At any rate, it is generally agreed that Hungarian grammar is more complicated than Slavic grammar, which is pretty impressive as Slavic



grammar is quite a beast.

Hungarian is **rated 5**, extremely hard to learn.

## Sino-Tibetan

### Chinese

It's fairly easy to learn to speak **Mandarin** at a basic level, though the tones can be tough. This is because the grammar is very simple - short words, no case, gender, verb inflections or tense. But with Japanese, you can keep learning, and with Chinese, you often tend to hit a wall, often because the syntactic structure is so strangely different from English (isolating).

Actually, the grammar is [harder](#) than it seems. At first it seems simple, like a simplified English. No word is capable of declension, and there is no tense, case, and number, nor are there articles. But the simplicity makes it difficult. No tense means there is no easy way to mark time in a sentence. Furthermore, tense is not as easy as it seems. Sure, there are no verb conjugations, but instead you must learn some particles and special word orders that are [used to mark tense](#). Mandarin has 12 different adverbs for which there are no good English translations.

Once you start digging into Chinese, there is a complex layer under all the surface simplicity. There is such things as aspect, serial verbs, a complex classifier system, syntax marked by something called topic-prominence, a strange form called the detrimental passive, preposed relative clauses, use of verbs rather than adverbs to mark direction, and all sorts of strange stuff. Verb complements can be baffling, especially potential and directional complements. The 把, 是 and 的 constructions can be very hard to understand.

The topic-prominence is interesting in that only a few major languages have topic-comment syntax, and most of those are Oriental languages with a lot of Chinese borrowing. Topicalization is not marked morphologically.

There are sentences where the entire meaning changes with the

addition of a single character. Chinese sentences are SVO (Subject -Verb - Object) at their base, but that is a bit of an illusion. A sentence that causes you to discuss time duration makes you repeat the verb after the direct object - SVOVT (T= time phrase). In the case of topicalization, sentences can have the structure of OSV (Object - Subject - Verb).

Relative clauses and all subordinate clauses come before the noun they modify. In other words:

English: "The man **who always wore red** walked into the room."

Chinese: "**Who always wore red** the man walked into the room."

The relative clause in the sentences above is marked in bold.

In Chinese, the prepositional phrase comes between the subject and the verb:

English: "The man hit the ball **into the yard.**"

Chinese: "The man **into the yard** hit the ball."

The prepositional phrase is bolded in the sentences above.

In Chinese, adjectives are actually stative verbs as in Nahuatl and Lakota.

的菜很好吃。

*Nàgè rède cài hěnhǎochī.*

"The **it is hot** food is good to eat."

"The hot food is delicious."

The 的 symbol turns "food hot" into "food it is hot", an attributive verb. 的 means something like "to be".

There are dozens of words called particles which shade the meaning of a sentence ever so slightly.

Chinese phonology is not as easy as some say. There are way too

many instances of the *zh, ch, sh, j, q,* and *x* sounds in the language such that many of the words seem to sound the same. There is a distinction between aspirated and nonaspirated consonants. There is also the presence of uncommon retroflex consonants.

Chinese orthography is probably the hardest orthography of any language. The alphabet uses symbols, so it's not even a real alphabet. There are at least 85,000 symbols and actually many more, but you only need to know about 3-5,000 of them, and many Chinese don't even know 1,000. To be highly proficient in Chinese, you need to know 10,000 characters, and probably less than 5% of Chinese know that many.

In addition, the characters have not been changed in 3,000 years, and the alphabet is at least somewhat phonetic, so we run into a serious problem of lack of a spelling reform.

The Communists tried to simplify the system (simplified Mandarin), but instead of making the connections between the phonetic aspects of character more sensible by decreasing their number and increasing their regularity (they did do this somewhat but not enough), they simply decreased the number of strokes needed for each symbol typically without dealing with the phonetic aspect of all. The simplification did not work well, so now you have a mixture of two different types of written Chinese - simplified and traditional.

In addition to all of this, Chinese borrowed a lot from the Japanese symbolic alphabet a full 1,000 years after it had already been developed and had not undergone a spelling reform, adding insult to injury.

Even leaving the characters aside, the stylistic and literary constraints required to write Chinese in an eloquent or formal (literary) manner would make your head swim. And just because you can read Chinese does not mean that you can read Classical Chinese prose. It's as if it's written in a different language - actually, it is technically a different language similar to Middle English or Old English. However, few Middle English or Old English texts are read anymore, and Classical Chinese is still widely read.

However, the orthography is at least consistent. 90% of characters have only one reading. Once you learn the character, you know the meaning in most but not all contexts.

Writing the characters is even harder than reading them. One wrong dot or wrong line either completely changes the meaning or turns the symbol into nonsense.

It's a real problem when you encounter a symbol you don't know because there is no way to sound out the word. You are really and truly lost and screwed. There is a clue at the right side of the symbol, but it is not always accurate. You need to learn quite a bit of vocabulary just to speak simple sentences.

Similarly, a dictionary is not necessarily helpful when trying to read Chinese. You can have a Chinese sentence in front of you along with a dictionary, and the sentence still might not make sense even after looking it up in the dictionary.

Some Chinese Muslims write Chinese using an Arabic script. This is often considered to be one of the [worst orthographies](#) of all.

The tones are often quite difficult for a Westerner to pick up. If you mess up the tones, you have said a completely different word. Often foreigners who know their tones well nevertheless do not say them correctly, and hence, they say one word when they mean another. However, compared to other tone systems around the world, the tonal system in Chinese is comparatively easy.

A major problem with Chinese is homonyms. To some extent, this is true in many tonal languages. Since Chinese uses short words and is disyllabic, there is a limited repertoire of sounds that can be used. At a certain point, all of the sounds are used up, and you are into the realm of homophones.

Tonal distinctions are one way that monosyllabic and disyllabic languages attempt to deal with the homophone problem, but it's not good enough, since Chinese still has many homophones, and meaning is often discerned by context, stress, rhythm and intonation. Chinese, like French and English, is heavily [idiomatic](#).

It's little known, but Chinese also uses different forms (classifiers) to count different things, like Japanese.

There is zero common vocabulary between English and Chinese, so you need to learn a whole new set of lexical forms.

In addition, nouns often show relatedness or hierarchy. For instance,

in English, you can simply say "my brother" or "my sister", but in Chinese, you cannot do this. You have to indicate whether you are speaking of an older or younger sibling.

*mei mei* "younger sister"  
*jie jie* "older sister"  
*ge ge* "older brother"  
*di di* "younger brother"

Mandarin scored very high on a [weirdest languages](#) study.

On the positive side, Chinese grammar is fairly regular and word derivation, compound words are sensible and the meaning can be determined by looking at the word. In other languages, compound words are not necessarily so obvious.

Many agree that Chinese is the hardest to learn of all of the major languages. A recent survey of language professors rated Chinese as the hardest language on Earth to learn.

Mandarin gets a **5 rating**, extremely hard to learn.

However, **Cantonese** is [even harder](#) to learn than [Mandarin](#). Cantonese has [eight tones](#) to Mandarin's four, and in addition, they continue to use a lot of the older traditional Chinese characters that were superseded when China moved to a simplified script in 1949. Furthermore, since non-Mandarin characters are not standardized, Cantonese cannot be written down as it is spoken.

In addition, Cantonese has verbal aspect, possibly up to 20 different varieties. Modal particles are difficult in Cantonese. Clusters of up to the 3 sentence final particles are very common. 我食咗飯 and 我食咗飯架喇嗰 are both grammatical for "I have had a meal", but the particles add the meaning of "I have already had a meal", answering a question or even to imply "I have had a meal, so I don't need to eat anymore".

Cantonese gets a **5.5 rating**, nearly hardest of all.

**Min Nan** is also said to be harder to learn than Mandarin, as it has a more complex tone system, with five tones on three different levels. Even many Taiwanese natives don't seem to get Min Nan right these days, as it is falling out of favor, and many fewer children are being raised speaking it than before.

Min Nan gets a **6 rating**, nearly hardest of all.

A [recent 15 year survey](#) out of Fudan University utilizing both the departments of Linguistics and Anthropology looked at 579 different languages in 91 linguistic families in order to try to find the most complicated language in the world. The [result](#) was that a Wu dialect (or perhaps a separate language) in the Fengxian district of southern Shanghai (**Dôndän Wu**) was the most phonologically complex language of all, with 20 separate vowels (Wang 2012). The nearest competitor was Norwegian with 16 vowels.

Dôndän Wu gets a **5.5 rating**, nearly hardest of all.

**Classical Chinese** is still read by many Chinese people and Chinese language learners. Unless you have a very good grasp on modern Chinese, classical Chinese will be completely wasted on you. Classical Chinese is much harder to read than reading modern Chinese.

Classical Chinese covers an era extending over 3,000 years, and to attain a reading fluency in this language, you need to be familiar with all of the characters used during this period along with all of the literature of the period so you can understand all the allusions. Even with a knowledge of Classical Chinese, you need to read it in context. If you are good at Classical Chinese and someone throws you a random section of it, it will take you a good amount of time to figure it out unless you know context.

The language is much more to the point than Modern Chinese, but this is not as good as it sounds. This simplicity leaves a room for ambiguity, and context plays an important role. A joke about some obscure historical or literary anecdote will be lost on you unless you know what it refers to. For reading modern Chinese, you will need at least 5,000 characters, but even then, you will still need a dictionary. With Classical Chinese, there are no lower limits on the number of characters you need to know. The sky is the limit.

Classical Chinese gets a **6 rating**, hardest of all.

## Tibeto-Burman

### Qiangic

### Northern

### Qiang

In **Qiang**, a language of Sichuan Province in China, not only are there rhotic vowels, which are present in only 1% of the world's languages, but there is also rhoticity harmony, where a non-rhotic vowel in a morpheme becomes rhotic when it is followed by a morpheme with a rhotic vowel.

$RUa + k^h e \rightarrow RUak^h e$   
 $me + we \rightarrow mewe$

Rhotic vowels are found in US English - Unstressed ə: "standard", "dinner", "Lincolnshire", "editor", "measure", and "martyr".

Qiang also has a very bad romanization, so bad that the Qiang will not even use it. Voiced consonants are written by adding a vowel to the symbol for the voiceless consonant. It has long and short vowels, but these are not represented in the system.

Qiang gets a **5 rating**, extremely hard to learn.

# Western Tibeto-Burman

## Bodish

### Central Bodish

#### Central

**Tibetan** probably has one of the least rational orthographies of any language. The orthography has not changed in ~1,000 years while the language has gone through all sorts of changes. A language learner in Tibet can get by using phonetic spelling. The problem comes when you try to spell using the Classical Alphabet. For instance:

*Srong rtsan Sgam po* (written)

*son̄tsɛn̄ gam̄po* (spoken)

*bsgrubs* (written)

*d`up* (spoken)

While the orthography is etymological and completely outdated, it is quite predictable.

Tibetan gets a **5 rating**, extremely hard to learn.

## Southern

**Dzongka**, the official language of Bhutan, has some pretty wild phonology, in addition to having the Tibetan writing system, this time using Bhutanese forms of the Tibetan script.

It contrasts all of the following:  $s$ ,  $s^h$ ,  $^hs$ ,  $^hs^h$ ,  $ts$ ,  $^hts$ ,  $ts^h$ ,  $z$ ,  $^hz$ ,  $dz$ ,  $^hdz$ ,  $ns^h$ ,  $^m̄ts^h$ ,  $^n̄ts^h$ ,  $^ndz$ ,  $^p̄ts$ ,  $^p̄ts^h$ ,  $^p̄ts^{wh}$ , and  $^ϕs$ , and in addition it has four tones, but there is no single word that is distinguished by tone only. On top of that, there are 22 different vowels.



Dzongka gets a **5 rating**, extremely hard to learn.

## **Austroasiatic**

### **Mon-Khmer**

#### **Viet-Muong**

**Vietnamese** is also hard to learn because to an outsider, the tones seem hard to tell apart. Therefore, foreigners often make themselves difficult to understand by not getting the tone precisely correct. It also has ["creaky-voiced" tones](#), which are very hard for foreigners to get a grasp on.

Vietnamese grammar is fairly simple, and reading Vietnamese is pretty easy once you figure out the tone marks. Words are short as in Chinese. However, the simple grammar is relative, as you can have [25 or more forms](#) just for "I", the 1st person singular pronoun. In addition, the Latin orthography is said to be [quite bad](#). It was invented by missionaries a few centuries ago, and it has never made much sense.

Vietnamese gets **5 rating**, extremely hard to learn.

## **Eastern Mon-Khmer**

### **Khmer**

**Khmer** has a reputation for being [hard to learn](#). I understand that it has one of the most complex honorifics systems of any language on Earth. Over a dozen different words mean "to carry" depending on what one is carrying. There are several different words for "slave" depending on who owned the slave and what the slave did. There are 28-30 different vowels, including sets of long and short vowels and long and short diphthongs. The vowel system is so complicated that there isn't even agreement on exactly what it looks like. Khmer

learners, especially speakers of IE languages, often have a hard time producing or even distinguishing these vowels.

Speaking it is not so bad, but reading and writing it is pretty difficult. For instance, you can put up to five different symbols together in one complex symbol. The orthographic script is even worse than the Thai one. There are actually rules to this mess, but no one seems to know who they are.

Khmer gets a **4.5 rating**, very to extremely hard.

## **Bahnaric**

### **North Bahnaric**

#### **West**

### **Sedang-Todrah**

#### **Sedang**

**Sedang**, a language of Vietnam, has the highest number of vowel sounds of any language on Earth, at 55 distinct vowel sounds.

Sedang gets a **5 rating**, extremely hard to learn.

## **Hmong-Mien**

### **Hmongic**

### **Chuanqiandian**

**Hmong** is widely spoken in this part of California, but it's not easy to learn. There are eight tones, and they are not easy to figure out. It's not obviously related to any other major language but the obscure Mien.

It has some very strange consonants called voiceless nasals. We have them in English as allophones - the *m* in "small" is voiceless, but in Hmong, they put them at the front of words - the *m* in the word "Hmong" is voiceless. These can be very hard to pronounce.

The romanization is widely criticized for being a [lousy one](#), but the Hmong use it anyway.

Hmong gets a **5 rating**, extremely hard to learn.

## Austro-Tai

### Austronesian

### Tsouic

**Tsou** is a Taiwanese aborigine language spoken by about 2,000 people in Taiwan.

Tsou is also ergative like most Formosan languages. Tsou is the only language in the world that has no prepositions or anything that looks like a preposition. Instead it uses nouns and verbs in the place of prepositions. Tsou allows more potential consonant clusters than most other languages.

About 1/2 of all possible CC clusters are allowed. Tsou has an inclusive/exclusive distinction in the 1st person plural and a very strange visible and non-visible distinction in the 3rd person singular and plural. Both adjectives and adverbs can turn into verbs as they are marked for voice in the same way that verbs are. Verbs are extensively marked for voice.

Nouns are marked for a variety of odd cases, often referring to perception (visible/invisible) and person and [place deixis](#).

<i>'e</i>	visible and near speaker
<i>si/ta</i>	visible and near hearer
<i>ta</i>	visible but away from speaker

'o/to invisible and far away, or newly introduced to discourse  
na/no ~ ne non-identifiable and non-referential (often when scanning a class of elements)

Tsou gets a **5 rating**, extremely hard to learn.

## **Malayo-Polynesian**

### **Malayo-Chamic**

#### **Malayic**

#### **Malay**

**Bahasa Indonesia** is an easy language to learn. For one thing, the grammar is dead simple. There are only a handful of prefixes, only two of which might be seen as inflectional. There are also several suffixes. Verbs are not marked for tense at all. And the sound system of these languages, in common with Austronesian in general, is one of the [simplest](#) on Earth, with only two dozen phonemes. Bahasa Indonesia has few homonyms, homophones, homographs, heteronyms, etc. Words in general have only one meaning.

Though the orthography is not completely phonetic, it only has a small number of exceptions. The orthography, nevertheless, is one of the easiest on Earth to use.

The system for converting words into either nouns or verbs is regular. To make a plural, you simply repeat a word, so instead of saying "pencils", you say "pencil pencil."

Bahasa Indonesia gets a **1.5 rating**, extremely easy to learn.

**Malay** is only easy if you learn the standard spoken form or one of the creoles. Learning the literary language is quite a bit more difficult. However, the Jawi script, which is Malay written in Arabic script, is often considered to be [perfectly awful](#).

Malay get a **2 rating** for moderately easy.

**Philippine**  
**Greater Central Philippine**  
**Central Philippine**  
**Tagalog**

However, **Tagalog** is much harder than Malay or Indonesian. Compared to many European languages, Tagalog syntax, morphology and semantics are often quite different. Also, Tagalog is typically spoken very fast. Unlike Malay, verbs conjugate quite a bit in Tagalog. The main idea of Tagalog grammar is something called focus. Once you figure that out, the language gets pretty easy, but until you understand that concept, you are going to have a hard time.

Everything is affixed in Tagalog.

However, articles and creation of adjectives from nouns is very easy.

Compare:

*ganda* "beauty" (noun)  
*maganda* "beautiful" (adjective)

Tagalog gets a **4 rating**, very difficult.

# Central-Eastern Malayo-Polynesian

## Eastern Malayo-Polynesian

### Oceanic

#### Central-Eastern Oceanic

#### Remote Oceanic

#### Central Pacific

#### East Fijian-Polynesian

#### Polynesian

#### Nuclear

#### East

#### Central

#### Tahitic

**Maori** and other Polynesian languages have a reputation for being quite easy to learn. The main problem for English speakers is that the sentence structure is backwards compared to English. In addition, macrons can cause problems.

One problem with Maori is dialects. The dialects are so diverse that this means that there are multiple words for the same thing. Swiss German has a similar issue, with up to 50 words for each common household item (nearly every major dialect has its own word for common objects):

*ngongi, noni, koki, wai*

“water”

<i>whiri, rarangi, hiri</i>	"to plait, to twist, to weave"
<i>pai, maitai</i>	"good"
<i>tu, tū, tutehu, mātika</i>	"to stand"
<i>mau, mou</i>	"to hold"
<i>pau, pou</i>	"to be exhausted"
<i>ika, tohorā</i>	"whale"
<i>ika, ngohi</i>	"fish"
<i>kāwei, kāwai</i>	"line"
<i>ori, kori, keukeu, koukou, neke, nuku</i>	"to move"
<i>haere, hara, here, horo, whano</i>	"to go, to come"
<i>hara, hapa, hē</i>	"to be wrong"
<i>kōrerorero, wānanga, rūnanga</i>	"to discuss"
<i>tohunga, tahunga</i>	"priest"
<i>matikuku, maikuku</i>	"fingernail"
<i>kanohi, konohi, mata, whatu, kamo, karu</i>	"eye, face"

Entire Maori sentences can be written with vowels only.

*E uu aau?*

"Are yours firm?"

*I uaa ai.*

"It rained as usual."

*I ui au 'i auau aau?'*

*E uaua!*

"It will be difficult/hard/heavy!"

On the plus side, the pronunciation is simple, and there is no gender. The language is as regular as Japanese. No Polynesian language has more than 16 sounds, and they all lack tones. They all have five vowels, which can be either long or short. A consonant must be followed by a vowel, so there are no consonant clusters. All consonants are easy to pronounce.

Maori gets a **3 rating**, average difficulty.

## Marquesic

**Hawaiian** is a pretty easy language to learn. It is easy to pronounce, has a simple alphabet, lacks complex morphology and has a fairly simple syntax.

Hawaiian gets a **2 rating**, very easy to learn.

## North and Central Vanuatu

### East Santo

#### North

**Sakao** is a very strange language spoken by 4,000 people in Vanuatu. It is very strange. It is a polysynthetic Austronesian language, which is very weird. It allows extreme consonant clusters. Sakao has an incredible seven degrees of deixis.

The language has an amazing four persons: singular, dual, paucal and plural. The neighboring language Tomoko has singular, dual, trial and plural. The trial form is very odd. Sakao's paucal derived from Tomoko's trial:

*jørðæɭ*

"they, from three to ten"

*jørðæɭ løn*

"the five of them" (Literally, "they three, five")

All nouns are always in the singular except for kinship forms and demonstratives, which only display the plural:

*ðjæɣ* "my mother/aunt" -> *rðjæɣ* "my aunts"

*walðɣɣ* "my child" -> *raalðɣɣ* "my children"



It has a number of nouns that are said to be "inalienably possessed", that is, whenever they occur, they must be possessed by some possessor. These often take [highly irregular](#) inflections:

Sakao	English
<i>æsiŋæ-γ</i>	"my mouth"
<i>æsiŋæ-m</i>	"thy mouth"
<i>ɔsiŋɔ-n</i>	"his/her/its mouth"
<i>æscæŋ-...</i>	"... 's mouth"
<i>uly-γ</i>	"my hair"
<i>uly-m</i>	"thy hair"
<i>ulæ-n</i>	"his/her/its hair"
<i>nøl-...</i>	"... 's hair"

Here, "mouth" is either *æsiŋæ-*, *ɔsiŋɔ-* or *æscæŋ-*, and "hair" is either *uly-*, *ulæ-* or *nøl-*

Sakao, strangely enough, may not even have syllables in the way that we normally think of them. If it does have syllables at all, they would appear to be at least a vowel optionally surrounded by any number of consonants.

*i* (V)  
"thou"

*Mhertpr.* (CCVCCCC)  
"Having sung and stopped singing thou kept silent."

Sakao has a suffix *-in* that makes an intransitive verb transitive and makes a transitive verb ditransitive. Ditransitive verbs can take two

arguments - a direct object and an instrumental.

*Mɨjilɨn amas ara./Mɨjilɨn ara amas.*

"He kills the pig with the club/He kills with the club the pig."

Sakao polysynthesis allows compound verbs, each one having its own instrument or object:

*Mɔssɔnɛshɔβrɨn aða ɛðɛ.*

*He-shooting-fish-kept-on-walking with-a-bow the-sea.*

"He walked along the sea shooting the fish with a bow."

Sakao gets a **5 rating**, extremely hard to learn.

## **Central-Eastern Oceanic**

### **Southeast Solomonic**

#### **Malaita–San Cristobal**

##### **Malaita**

##### **Northern Malaita**

**Kwaio** is an Austronesian language spoken in the Solomon Islands.

It has [four different forms of number](#) to mark pronouns - not only the usual singular and plural, but also the rarer dual and the very rare paucal. In addition, there is an inclusive/exclusive contrast in the non-singular forms.

For instance:

1 dual inclusive (you and I)

1 dual exclusive (I and someone else, not you)

1 paucal inclusive (you, I and a few others)

1 paucal exclusive (I and a few others)

1 plural inclusive (I, you and many others)

1 plural exclusive (I and many others)

Pretty wild!

Kwaio gets a **5 rating**, extremely hard to learn.

## **Greater Barito**

### **East Barito**

### **Malagasy**

**Malagasy**, the official language of Madagascar, has a reputation for being even easier to learn than Indonesian or Malay.

Malagasy gets a **1 rating**, easiest of all to learn.

## **Tai-Kadai**

### **Kam-Tai**

### **Tai**

## **Southwestern**

**Thai** is a pretty hard language to learn. There are 75 symbols in the strange script, there are no spaces between words in the script, and vowels can come before, after, above or below consonants in any given syllable. There seem to be many different glyphs for every consonant, but the different glyphs for the same consonant will

sometimes change the sound of the neighboring vowel. The orthography is as insensible as that of English since centuries have gone by with no spelling reforms, in fact, Thai has not changed its system in 1000 years. The wild card of having tone thrown in adds to the insanity.

Consonant pronunciations vary depending on the location of the syllable in the word - for instance, *s* can change to *t*. There are many vowels which are spoken but not written. There are many consonants that are pronounced the same - for instance, there are six different *t*'s, not counting the *s*'s that turn into *t*'s. The Thai script is definitely one of the most difficult phonetic scripts. Nevertheless, the Thai script is easier to learn than the Japanese or Chinese character sets. In spite of all of that, the syntax is simple, like Chinese.

There are five tones, including a neutral tone. Tones are determined by a variety of complex things, including a combination of tone marks, the class of consonants, if the syllable ends in a sonorant or a stop and what the tone of the preceding syllable was. Tone marking in the orthography is quite complex.

The vowels are different than in many languages, and there are some unusual diphthongs: *eua*, *euai*, *au* and *uu*. There is a contrast between aspirated and unaspirated consonants.

There is a system of noun classifiers for counting various things, similar to Japanese. In addition, common to many Asian languages, there is a complicated honorifics system.

On the plus side, Thai is a regular language, with few exceptions to the rules. However, the rules are quite complex. The syntax is about as complex as that of Chinese, and the grammar is dead simple.

Thai gets a **5 rating**, extremely hard to learn.

**Lao** is very similar to Thai, in fact it is identical to a Thai language spoken by 16 million people in northeast Thailand called Northeastern Thai. The Lao script is similar to Thai, but it has fewer letters so there is somewhat less confusion.

Lao gets a **4.5 rating**, very to extremely hard to learn.

## Kam-Sui

The **Kam languages** of the Dong people in southwest China were rated by the Fudan University study referenced above under Wu as the 2nd most phonologically complex on Earth (Wang 2012). There are 32 stem initial consonants, including oddities like  $t_{\text{ɕ}}$ ,  $t_{\text{ɕ}}^h$ ,  $p^j$ ,  $p^{jh}$ ,  $\text{ɕ}$ ,  $k^w$ ,  $k^{wh}$ ,  $\eta^w$ ,  $tj^h$ ,  $ts^h$ . Note the many contrasts between aspirated and unaspirated voiceless consonants, including bilabial palatalized stops, labialized velar stops, and alveolar affricates. There are an incredible 64 different syllable finals, and 14 others that occur only in Chinese loans.

There are an astounding [15 different tones](#), nine in open syllables and six in checked syllables (entering tones). Main tones are high, high rising, high falling, low, low rising, low falling, mid, dipping and peaking. When they speak, it sounds as if they are singing.

Kam gets a **5 rating**, extremely hard to learn.

## Kra

## Paha

According to the Fudan University study quoted above, **Buyang** in the 3rd most phonologically complex language in the world. Buyang is a cluster of 4 related languages spoken by 1,900 people in Yunnan Province, China. Buyang has a [completely wild](#) consonant inventory.

It has a full set of both voiced and voiceless plain and aspirated stops, including voiceless uvulars. The contrast between aspirated and plain voiced stops is peculiar. The stop series also has distinctions between palatalized and rounded stops throughout the series. It has a labialized voiceless palatal fricative and a voiceless dental aspirated lateral, unusual sounds. It has four different voiceless aspirated nasals. It has voiceless y and w, more odd sounds. It also has plain and labialized palatal glides.

That is one heck of a wild phonology.

Buyang gets a **5 rating**, extremely hard to learn.

## **Niger-Kordofanian**

### **Niger-Congo**

#### **Atlantic-Congo**

##### **Kwa**

##### **Nyo**

#### **Ga-Dangme**

The African Bantu language **Ga** has a bad reputation for being a tough nut to crack. It is spoken in Ghana by about 600,000 people. It has two tones and engages in a strange behavior called [tone terracing](#) that is common to many West African languages. There is a phonemic distinction between three different types of vowel length. All vowels have 3 different lengths - short, long and extra long. It also has many sounds that are not in any Western languages.

Ga gets a **5 rating**, extremely hard to learn.

## **Potou-Tano**

### **Tano**

#### **Central Bia**

#### **Northern**

**Anyi** is a language spoken by 610,000 people in Côte d'Ivoire. It is relatively straightforward as far as African languages go. Probably the

hardest part about the language is that it is tonal, and it does have two tones. The phonology does have a +-ATR contrast which will seem very odd. ATR stands for advanced tongue root, so the language has a contrast between vowels with an advanced tongue root and without them. However, the grammar is pretty regular. There are few confusing phonological processes.

Anyi has a simple tense system, with only present, past and future. There is no aspect, mood or voice marking and it lacks the noun class systems so common in many African languages. It has a plural marker, but it is often optional.

The syntax does have serial verbs, which will seem odd to Westerners. It distinguishes between relative clauses marked with *b* and subordinate clauses marked with *kε*.

Anyi gets a **4 rating**, very hard to learn.

## **Volta-Congo**

### **Benue-Congo**

#### **Bantoid**

#### **Southern**

#### **Narrow Bantu**

#### **Central**

#### **M**

#### **Nyika-Safwa**

**Ndali** is a Bantu language with 150,000 speakers spoken in Malawi and Tanzania.

It has [many strange tense forms](#). For instance, in the past tense:

Past tense A: "He went just now."

Past tense B: "He went sometime earlier today."

Past tense C: "He went yesterday."

Past tense D: "He went sometime before yesterday."

Future tense is marked similarly:

Future tense A: "He's going to go right away."

Future tense B: "He's going to go sometime later today. "

Future tense C: "He's going to go tomorrow."

Future tense D: "He's going to go sometime after tomorrow."

Ndali gets a **5 rating**, extremely hard to learn.

## S

### Nguni

**Xhosa**, a language of South Africa, is quite difficult, with up to nine click sounds. Clicks only exist in one language outside of Africa - the Australian language Damin - and are extremely difficult to learn. Even native speakers mess up the clicks sometimes. Nelson Mandela said he had problems making some of the click sounds in Xhosa. The phonemics in general of Xhosa are pretty wild.

Xhosa gets a **5 rating**, extremely hard to learn.

Zulu and **Ndebele** also have these impossible click sounds. However, outside of click sounds, the phonology of Nguni languages is straightforward. All Nguni languages are agglutinative.

These languages also make plurals by changing the prefix of the noun, and the manner varies according the noun class. If you want to look up a word in the dictionary, first of all you need to discard the prefix. For instance, [in Ndebele](#),



"river"	<i>umfula</i>
"rivers"	<i>imifula</i> , <b>but</b>
"stone"	<i>ilitshe</i>
"stones"	<i>amatshe</i> , <b>yet</b>
"tree"	<i>isihlahla</i>
"trees"	<i>izihlahla</i>

Ndebele gets a **5 rating**, hardest of all.

**Zulu** has pitch accent, tones and clicks. There are nine different pitch accents, four tones and three clicks, but each click can be pronounced in five different ways. However, tones are not marked in writing, so it's hard to figure out when to use them. Zulu also has depressor consonants, which lower the tone in the vowel in the following syllable. In addition, Zulu has multiple gender - 15 different genders. And some nouns behave like verbs. It also has 12 different noun classes, but 90% of words are part of a group of only three of those classes.

Zulu gets a **5 rating**, extremely hard to learn.

## G

### Swahili

For unknown reasons, **Swahili** is generally considered to be an easy language to learn. The US military ranks it 1, with the easiest of all languages to learn. This seems to be the typical perception. Why Swahili is so easy to learn, I am not sure. It's a trade language, and trade languages are often fairly easy to learn. There's also a lot of controversy about whether or not Swahili can be considered a creole, but that has not been proven. For the moment, the reasons why Swahili is so easy to learn will have to remain mysterious.

On the down side, Swahili has many noun classes, but they have the

benefit of being more or less logical.  
Swahili gets a **2 rating**, moderately easy.

## **Khoisan**

### **Southern Africa**

#### **Southern**

#### **Hua**

**!Xóõ (Taa)**, spoken by only 4,200 Bushmen in Botswana and Namibia, is a notoriously difficult Khoisan language replete with the notoriously impossible to comprehend click sounds. Taa has anywhere from 130 to 164 consonants, the largest phonemic inventory of any language. Of this vast wealth of sounds, there are anywhere from 30-64 different click sounds. There are five basic clicks and 17 accompanying ones. Speakers develop a lump on their larynx from making the click sounds.

In addition, there are [four types](#) of vowels: plain, pharyngealized, breathy-voiced and strident. On top of that, there are four tones. Taa appears on [many lists](#) of the wildest phonologies and craziest languages period on Earth.

Taa gets a **5 rating**, extremely hard to learn.

#### **Northern**

**Ju|'hoan**, a Khoisan language spoken by 5,000 people in Botswana, has one of the [wildest phonological inventories](#) on Earth. Some question whether these segments actually exist and say that they are instead spoken with a "breathy-voice." However, voiced aspirated consonants do appear to be real.

In addition, Ju|'hoan has a closed class of only 17 adjectives since descriptive functions are done by verbs. They are the following:

*female*

*male*

*other* (those remaining)

*other* (strange)

*true*

*old*

*new*

*a certain*

*each*

*all*

*some*

the numbers one through four

Ju|'hoan scored very high on a [study](#) of the weirdest languages on Earth.

Ju|'hoan gets a **5 rating**, extremely hard to learn.

## **Eskimo-Aleut**

### **Eskimo**

### **Inuit-Inupiaq**

**Inuktitut** is extremely hard to learn. Inuktitut is polysynthetic-agglutinative, and roots can take many suffixes, in some cases up to 700. Verbs have 63 forms of the present indicative, and conjugation involves 252 different inflections. Inuktitut has the complicated polypersonal agreement system discussed under Georgian above and Basque below. In a typical long Inuktitut text, 92% of words will occur only once. This is quite different from English and many other languages where certain words occur very frequently or at least frequently. Certain fully inflected verbs can be analyzed both as verbs

and as nouns.

Words can be very long.

*Inuktituusuungutsialaarungnanngittuaraaluuvunga.*

"I truly don't know how to speak Inuktitut very well."

You may need to analyze up to 10 different bits of information in order to figure out a single word. However, the affixation is all via suffixes (there are no prefixes or infixes) and the suffixation is extremely regular.

Inuktitut is also rated one by linguists one of the hardest languages on Earth to [pronounce](#). Inuktitut may be as hard to learn as Navajo.

Inuktitut is **rated 6**, hardest of all.

**Kalaallisut (Western Greenlandic)** is very closely related to Inuktitut.

Look at this sentence:

*Aliikusersuillammassuaanerartassagalarpaalli...*

"However, they will say that he is a great entertainer, but..."

That word is composed of 12 separate morphemes. A single word can conceptualize what could be an entire sentence in a non-polysynthetic language.

Kalaallisut is **rated 6**, hardest of all.

# Chukotko-Kamchatkan

## Northern

### Chukot

**Chukchi** is a [polysynthetic](#), agglutinating and incorporating language and is often listed as one of the hardest languages on Earth to learn.

*Təmeyŋəlevtpəytərkən.*

"I have a fierce headache."

There are five morphemes in that word, and there are three lexical morphemes (nouns or adjectives) incorporated in that word: *meyŋ* "great", *levt* "head", and *pəyt* "ache".

Chukchi gets a **6 rating**, hardest of all.

## Basque

**Basque**, of course, is just a wild language altogether. There is an old saying that the Devil tried to learn Basque, but after seven years, he only learned how to say *Hello* and *Goodbye*. Many Basques, including some of the most ardent Basque nationalists, tried to learn Basque as adults. Some of them succeeded, but a very large number of them failed. Based on the number that failed, it does seem that Basque is harder for an adult to learn as an L2 than many other languages are. Basque grammar is maddeningly complex and it often makes it onto [craziest grammars](#) and craziest language lists.

There are 11 cases, and each one takes [four different forms](#). The verbs are quite complex. This is because it is an ergative language, so verbs vary according to the number of subjects and the number of objects and if any third person is involved.

This is the same polypersonal agreement system that Georgian has

above. Basque's polypersonal system is a polysynthetic system consisting of two verb types - synthetic and analytical. Only a few verbs use the synthetic form.

Three of Basque's cases - the absolutive (intransitive verb case), the ergative (intransitive verb case) and the dative - can be marked via affixes to the verb. In Basque, only present simple and past simple synthetic tenses take polypersonal affixes.

The analytical forms are composed of more than one word, while the synthetic forms are all one word. The analytic verbs are built via the synthetic verbs *izan* "be", *ukan* "have" and *egin* "do".

Synthetic:

*d-akar-ki-o-gu*  
*ekarri* "bring".

"We bring it to him/her." The verb is

*z-erama-zki-gu-te-n*  
*eraman* "take".

"They took them to us." The verb is

Analytic:

*Ekarriko d-i-o-gu =*  
"We'll bring it to him/her." Literally: "We  
will have-bring it to him/her." The analytic verb is built from *ukan*  
"have".

"We'll bring it to him/her." Literally: "We  
will have-bring it to him/her." The analytic verb is built from *ukan*  
"have".

*Eraman d-ieza-zki-gu-ke-te* "They can take them to us." Literally:  
"They can be taking them to us." The analytic verb is built from *izan*  
"be".

Most of the analytic verbs require an auxiliary which carries all sorts of information that is often carried on verbs in other languages - tense, mood, sometimes gender and person for subject, object and indirect object.

*Jaten naiz.*

"Eat I-am-doing."

"I am eating."

*Jaten nintekeen.*

"Eat I-was-able-to."

"I could eat."

*Eman geniezazkiake.*

"Give we-might-have-them-to-you-male."

"We might have given them to you."

In the above, *naiz*, *nintekeen* and *geniezazkiake* are auxiliaries. There are actually 2,640 different forms of these auxiliaries!

A language with ergative morphosyntax in Europe is quite a strange thing, and Basque is the only one of its kind.

The ergative itself is quite unusual:

*Gizona etorri da.* "The man has arrived."

*Gizonak mutila ikusi du.* "The man saw the boy."

*gizon* "man"

*mutil* "boy"

*-a* "the"

The noun *gizon* takes a different form whether it is the subject of a transitive or intransitive verb. The first sentence is in absolutive case (unmarked) while the second sentence is in the ergative case (marked by the morpheme *-k*).

If you come from a non-ergative IE language, the concept of ergativity itself is difficult enough to conceptualize, much less trying to actually learn an ergative language. Consequently, any ergative language will automatically be more difficult than a non-ergative one for all speakers of IE languages.

Ergativity also works with pronouns. There are four basic systems:

*Nor*: verb has subject only

*Nor-Nork*: verb has subject + direct complement

*Nor-Nori*: verb has subject + indirect complement.

*Nor-Nori-Nork*: verb has subject + indirect + direct complement

Some call Basque the most consistently ergative language on Earth.

If you don't grow up speaking Basque, it's hard to attain native speaker competence. It's quite a bit easier to write in Basque than to speak it.

Nevertheless, Basque verbs are quite regular. There are only a few irregularities in conjugations and they have phonetic explanations. In fact, the entire language is quite regular. In addition, most words above the intermediate level are borrowings from large languages, so once you reach intermediate Basque, the rest is not that hard. In addition, pronunciation is straightforward.

Basque is **rated 5.5**, nearly hardest of all.

## References

Dorani, Yakir. Hebrew speaker. Israel. August 2013. Personal communication.

Hewitt, B. G.. 2005. Georgian: A Learner's Grammar, p. 29.

Kim, Yuni. December 16, 2003. [Vowel Elision and the Morphophonology of Dominance in Aymara](#). UC Berkeley.

Kirk, John William Carnegie. 1905. A Grammar of the Somali Language: With Examples in Prose and Verse and an Account of the Yibir and Midgan Dialects, pp. 73-74.

Rogers, Jean H. 1978. Differential Focusing in Ojibwa Conjunct Verbs: On Circumstances, Participants, and Events. *International Journal of American Linguistics* 44: 167-179.

Wang, Chuan-Chao et al. 2012. Comment on "Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa." *Science* 335: 657.



