

LANGUAGE FAMILIES

Olga Krasnoukhova

A language family is a set of languages that descended from one single ancestral language and are therefore genealogically related. The ancestral language is often referred to as the proto-language of that family. Within a family, languages can be arranged into branches and groups. These divisions are usually based on shared innovations: grammatical characteristics found in some related languages but not found in their common ancestor. Using the metaphor of a biological family tree, languages in a language family are sometimes called sister languages and daughter languages, dependent on their level of relatedness. This entry examines the linguistic diversity, and it addresses such questions as to why numbers of language families and languages spoken across the world can only be given as estimates.. The second part of the entry gives a detailed overview of language families of six geographic areas.

Linguistic Diversity

Among the processes of language formation is language divergence. That is a diversification through space and time, when an originally one speech community falls apart (often geographically) and gradually evolves into distinct languages. These would first form dialects of one language. Subsequently, the dialects would diverge to a degree that they are no longer mutually intelligible, thus starting to be separate languages. The distinction between a language and a dialect generally hinges on this—partly subjective—parameter of mutual intelligibility.

However, sociopolitical motivations also play a role. One example is Swedish and Norwegian, which are mutually intelligible, but are generally regarded as separate languages. Another example includes Serbian, Croatian, and Bosnian.

The number of existing languages is usually given as a rough estimate, since an exact number is impossible to state for two reasons. The fluid line between a language and a dialect is the first one. The second one is a simple lack of sufficient documentation for many languages that we theoretically know of, thus not allowing us to evaluate their status. Glottolog, a systematic catalog of the world's languages today, lists more than 7,000 languages. The number of top-level genealogical units, or language families that all the languages fall into, is likewise a question with no precise answer. Research on the membership in a family, as well as the level of relatedness that is involved, has been ongoing for many of the world's languages, and it regularly brings in adjustments in existing classifications. For the present-day view, Glottolog suggests around 240 top-level language families, also known as phyla. These language families are of a very different size, ranging from just two members to more than a thousand. For example, the largest language family, Atlantic-Congo, includes around 1,400 languages. The Austronesian language family counts around 1,200 languages, and the Indo-European has around 580 languages.

There are also languages that are not part of any language family. These are of two types. One is languages that are not documented sufficiently to be compared to other languages in a meaningful way, and for which any possible genealogical relatives remain unknown. These are referred to as unclassified languages. The other is languages that do not have any demonstrable genealogical relatedness to any other languages. These are language isolates. In principle, a language isolate can be regarded as a language family with just one living member, since it may

have had relatives in the past. It is not impossible that most present-day language isolates may have had sister or daughter languages that became extinct without leaving any traces or record. According to Glottolog, there are 188 language isolates and more than 100 unclassified languages.

A fluid borderline between a language family, a language, and a dialect is illustrated by Lyle Campbell for Basque, a language traditionally referred to as an isolate. Specialists of Basque keep working on the degree of genealogical relation between Basque and Aquitanian, an extinct language once spoken in the ancient Aquitaine. One hypothesis suggests that Aquitanian is the ancestral language of Basque. Another hypothesis puts forward that Aquitanian and Basque are related at the level of sister languages, having one common ancestor. If the latter hypothesis ever proves to be correct, then Basque could be considered as one surviving language of a family, not an isolate in the strict sense. Basque is tricky in another way, too: Should Basque be called a language, or is it already a small language family? By standard criteria, dialects that are not mutually intelligible are considered separate languages. Larry Trask in the 1990s argued that the dialects of Basque have diversified considerably and that at least Zuberoa (Suleitino, Souletin) is not mutually intelligible with the other dialects. Therefore, Basque could already be considered a small language family rather than a language. Open questions like the ones for Basque are pertinent for many if not most languages.

Based on the number of speakers who use a language as their mother tongue, languages can be roughly divided into healthy, endangered, moribund, and dead (or extinct). As of today, at least one fifth of the world's languages are either seriously endangered or moribund. The distribution of languages according to the number of speakers is extremely skewed. According to Ethnologue, 8 languages are spoken as their mother tongue by 40% of the world's population.

About 400 languages are the mother tongue of 94% of the world's population. This also means that the other 6,600 languages are spoken by only 6% of the world's population. Besides, the linguistic diversity is not distributed evenly in geographic terms. There are areas in the world that harbor an impressive number of languages. These are referred to as linguistic hotspots. Among such undisputed hotspots are the New Guinea and the East Melanesian Islands, which host 1,000–2,000 different languages. Among other linguistic hotspots are the Caucasus, eastern Siberia, north of Australia, the Northwest Pacific Plateau of the United States, the Macro-Sudan belt in northern Sub-Saharan Africa, and the Guaporé-Mamoré area in the Amazon region of South America. The continent of South America is in itself a macro-area of extreme linguistic diversity.

Below is an overview of language families found on six macro-areas, based on the data from Harald Hammarström and Glottolog. The number of languages given in the brackets should give the reader an indication of a family size, but should not be treated as final, as continually emerging data and ongoing research regularly bring in some adjustments.

Africa

There are more than 2,000 languages in Africa. The majority of these belong to the Atlantic-Congo language family, which is also the largest language family in the world (> 1,400 members). Another major language family in Africa is Afroasiatic (> 300 members), with North Africa as its core area. There are also middle-sized language families (< 100 members) such as Eastern Sudanic, Central Sudanic, Mande in West Africa, Omotic in Ethiopia, Dogon in West Africa, Saharan in Chad, and a few small languages families and isolates.

Eurasia

Eurasia refers here to the landmass north of Sinai; it includes Japan and islands to the north of it but does not include insular Southeast Asia. Eurasia counts around 1,400 languages. The largest language families (ranging between 580 and 160 languages) are Indo-European with Europe and India as its core areas, Sino-Tibetan in China and Nepal, and Austroasiatic in India and Southeast Asia. Midsized language families (< 100 members) are Tai-Kadai in Southeast Asia, Dravidian in southern India, Turkic with Central Asia as its core area, Uralic in Europe and northwest Russia, Hmong-Mien in southern China, Nakh-Dagestanian in Northeast Caucasus, Mongolian in Mongolia, Japonic in East Asia and Japan, and Tungusic in Siberia. There are several small language families (< 10 members), such as Great Andamanese and Jarawa-Onge in the Andaman Islands, Kartvelian and Abkhaz-Adyge in the Caucasus, and Chukotko-Kamchatkan, Yukaghir and Ket in Siberia. There are 8 one-member language families or isolates: Basque in Western Europe, Nivkh in Siberia, Ainu in northern Japan, Korean in Korea, Burushaski in North Pakistan, Kusunda in Nepal, Nihali in India, and Shom Pen in the Nicobar Islands.

Australia

There are around 250 languages in Australia. The largest language family present on the continent is Pama-Nyungan (> 200 members). Pama-Nyungan languages cover the vast majority of the continent. All other aboriginal languages of Australia are generally referred to as “non-Pama-Nyungan.” As the name suggests, this is not a genealogical classification, but a grouping of languages defined negatively. These non-Pama-Nyungan languages represent about 24 different genealogical units, ranging in size from about eleven to only one member.

Geographically, they are all found in the northern part of Australia, which makes it one of the most linguistically diverse areas in the world.

Oceania

Oceania is a label here for a geographic area that includes all islands of the tropical Pacific Ocean, between Sumatra and the Americas. There are more than 2,000 languages here. The largest language family is Austronesian (> 1,200). Austronesian languages are spoken on some of the coast of New Guinea, as well as most smaller islands. Most of the interior of New Guinea, part of its coast, and some island areas are occupied by languages labeled as “Papuan.” Papuan is another group defined negatively and is used in reference to languages spoken in this area that are not Austronesian. Genealogical relations between the Papuan languages have been a matter of ongoing research and debate among specialists of this area. At least two different views on classification have prevailed. According to Stephen Wurm, Papuan languages fall into five major phyla and six minor ones, plus several language isolates. According to William Foley, the available data suggest no fewer than 60 distinct top-level genealogical units with no evidence for higher level relationships. One of the problems is the lack of data on many of the languages, leaving existing proposals on classification at the level of hypotheses.

North America

North America harbors around 320 languages, falling into around 30 language families. The largest ones, ranging between 70 and 45 members, are Uto-Aztecan, Algic, and Athapaskan-Eyak-Tlingit. Other are smaller in size, such as Salishan, Siouan, Eskimo-Aleut, Iroquoian, and Miwok-Costanoan. There are many small language families (< 10 members), such as Cochimi-

Yuman, Pomoan, Kiowa-Tanoan, Muskogean, Wakashan, Caddoan, Sahaptian, Maiduan, Tsimshian, Chinook, Haida, Keresan, Palaihnihan, and Wintuan. And there are 13 language isolates: Coosan, Kalapuyan, Karuk, Klamath-Modoc, Kutenai, Takelma, Wappo, Washo, Yana, Yokutsan, Yuchi, Yuki, and Zuni.

Central and South America

There are around 300 languages in Central America. The largest language family here is Otomanguean with around 180 languages spoken in central and southern Mexico (for the purposes of this entry, this part of Mexico has been grouped with Central and South American countries). Other languages families are much smaller in size (< 30). These include Mayan, Mixe-Zoque, Totonacan, Huavean, Tarascan, and Tequistlatecan (all with Mexico as the core area), Misumalpan in Honduras and Nicaragua, and Jicaquean in Honduras. There are two living language isolates, such as Seri in Mexico, and Xincan in Guatemala.

According to Lyle Campbell, South America has about 420 languages still spoken today, and these fall in no less than 108 distinct genealogical units. The largest language families found here range between 70 and 25 members: Tupian, Arawakan, Quechuan, Cariban, Panoan, and Tucanoan. There are also a few ‘middle-sized’ families (ranging between 25 and 10 members): Chibchan, Ge, and Chocoan. However, the great majority of language families in South America are small, and often comprise less than 6 members. There is one more intriguing fact about the genealogical diversity of the South American continent. According to Frank Seifart and Harald Hammarström, here one finds the highest number (and proportion) of language isolates in the world, with as much as 60 % of the lineages being isolates.

Olga Krasnoukhova

See also Australian Aboriginal Languages; Descriptive Linguistics; Indigenous Languages of Central and South America; Indigenous Languages of North America; Indigenous Languages of the Pacific; Languages of Africa; Languages of Europe; Languages of East Asia; Languages of North Asia; Languages of South Asia; Languages of Southeast Asia; Languages of Central Asia; Linguistic Typology

FURTHER READINGS

Campbell, L. (2012). Classification of the indigenous languages of South America. In L. Campbell & V. Grondona (Eds.), *The indigenous languages of South America: A comprehensive guide* (59–166). Berlin, Germany: De Gruyter Mouton.

Campbell, L. (2016). Language isolates and their history, or, What’s weird, anyway? In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 36, 16–31.
<http://dx.doi.org/10.3765/bls.v36i1.3900>

Cysouw, M, & Good, J. (2013). Languoid, doculect, and glossonym: Formalizing the notion “language.” *Language Documentation & Conservation*, 7, 331–359.

Simons, G. F., & Fennig, C. D. (Eds.). (2017). *Ethnologue: Languages of the world* (20th ed.). Dallas, TX: SIL International. Retrieved from <http://www.ethnologue.com>

Foley, W. A. 1986. *The Papuan languages of New Guinea*. Cambridge, United Kingdom: Cambridge University Press.

Hammarström, H., & Forkel, R., & Haspelmath, M. (2017). *Glottolog 3.0*. Jena, Germany: Max Planck Institute for the Science of Human History. Retrieved April 2017 from

<http://glottolog.org>

Hammarström, H. (2010). Online appendix for a full-scale test of the language farming dispersal hypothesis. *Diachronica*, 27(2), 197–213.

Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago, Illinois: University of Chicago Press.

Seifart, F & H. Hammarström (2018 forthc.) Language isolates in South America. In L. Campbell (Ed.), *Language isolates* (260-286). London and New York: Routledge. Trask, R. L. (1997). *History of Basque*. London, United Kingdom: Routledge.

Wurm, S. A. (1982). *Papuan languages of Oceania* (Ars Linguistica, 7). Tübingen, Germany: Gunter Narr.