How do linguists categorize languages?

Linguists separate languages into language groupings using what is called the **comparative method**, a technique of linguistic analysis that compares lists of related words, called cognates, in a selection of languages. For example, the word for “father” in Latin is *pater*, in French *père*,and in Spanish, *padre*. These similarities suggest that these three words come from a common ancestor language (Latin), and are in the same language family (the Italic, or Romance, family). Similarly, the word for “father” in English is related to *Vater* in German, (“v” is pronounced as “f” here), suggesting that English and German are related in a similar way and are part of the Germanic language family.

Taking this comparison a step further, we find that other words that begin with “p” in Italic languages show up in Germanic languages beginning with “f”; so, for example, the words for “fish” in Latin, French, and Spanish (*pisces*, *poisson*,and *pescado*,respectively) all begin with “p”, and the word for “fish” in German and English begin with “f” -- *Fisch* and *fish* (and *fisk* in Danish, another Germanic language*).* We can surmise, therefore, that the Italic and Germanic language families themselves may descend from a common root language. And in fact, we know that scholars have shown that Italic and Germanic are members of the larger language family, **Indo-European,** and that languages in this family descend from a common root language, **Proto Indo-European.** The Indo-European language family is only one of about 100 language families, but more than half the world’s population speaks an Indo-European language as a first or second language.[[1]](#footnote-1)

**Proto-Indo-European Word Roots**

The following table gives examples of words in five Indo-European languages and then the older Proto-Indo-European (PIE) root words. From lists like these, regular sound relationships are established and a possible proto-form (hypothetical word) is proposed (traditionally indicated by the asterisk, \*=possible).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sanskrit** | **Greek** | **Latin** | **Gothic** | **English** | **PIE roots** | **meaning** |
| pita | pater | pater | fadar | father | \*pater- | father |
| padam | poda | pedem | fotu | foot | \*ped- | foot |
| bhratar | phrater | frater | brothor | brother | \*bhrater- | brother |
| bharami | phero | fero | βaira | bear | \*bher- | carry |
| jivah |  | wiwos | qius | quick (living) | \*gwei- | live |
| sanah | henee | senex | sinista | senile[[2]](#footnote-2) | \*sen- | old |
| virah |  | wir | wair | were (as in *were*wolf) | \*wiro- | man |

Sanskrit is a language that is used for ceremonial purposes now in India, but like Greek and Latin, has influenced many languages in the region. Gothic is an extinct Germanic language.

 **Major Language Families of the World**

|  |  |  |
| --- | --- | --- |
| **Family** | **Sampling of Members** | **Main Areas Where Spoken** |
| Uralic | Hungarian, Finnish, some Siberian languages like Mordvin | Europe |
| Altaic | Mongolian  | Mongolia |
| Sino-Tibetan | Chinese languages like Mandarin and Cantonese | Asia |
| Malayo-Polynesian | Malay, Indonesian, Maori, Hawaiian, Malagasy | Island nations of Southeast Asia and the Pacific Ocean, continental Asia, Madagascar |
| Iroquoian | Cherokee, Mohawk, Seneca, Huron | North America |
| Afro-Asiatic | Arabic and Hebrew | North Africa and Middle East |
| Caucasian | Georgian and Chechen | Caucasus Mountains |
| Dravidian | Tamil, Malayalam, and Kannada | India |
| Austro-Asiatic | Vietnamese and Khmer | India to Vietnam |
| Niger-Congo | Swahili, Shona, Xhosa, Yoruba, and Zulu | Africa, south of the Sahara Desert |
| Nilo-Saharan | Luo, Songhay, Dinka, Nubian | North Africa, Upper Nile region of Sudan |
| Khoisan <Khoi-San> | ‘Bushman’ languages, Nama, Sandawe, Hadza | Southern and eastern Africa |
| Eskimo-Aleut | Aleut, Eskimo languages like Inuit, Yupik, Atka, Iñupiaq | North American Arctic |
| Uto-Aztecan | Hopi, Comanche, O’odham | Western U.S., Mexico |
| Mayan | Ch’olan, Quiche, Yucatec | Northern Central America, Meso-America |

Source: Lewis, M. P. (ed.). 2009. *Ethnologue: Languages of the world*, 16th edn. Dallas, TX: SIL International. Online version: http://www.ethnologue.com/.

Exercise 1.1 — Numbers in IE and non-IE Languages

Can you determine which languages are Indo-European (as English is), and which are not?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **English (IE)** | **Language 1** | **Language 2** | **Language 3** | **Language 4** | **Language 5** |
| one | uno | nigen | mot | aeva | jedyn |
| two | due | khoyar | hai | dva | dwaj |
| three | tre | ghorban | ba | thrayo | tři |

Indo-European: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Non-European: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Faux Amis***

Sometimes words can look like they have a close relationship, when in fact they do not. The French expression for these sorts of look-alikes is *faux amis* or ‘false friends.’ As you will read in an early description of the language of the Nootka Jargon spoken by the Nuu-Chah-Nulth, José Mariano Moziño wrote, “In order to say ‘we go,’ the Mexicans [Nahuatl Indians] use the word *tlato,* and the Nootkans, in order to say ‘go away,’ use *tlatlehua.”* This looks to be an accidental overlap, or maybe even a borrowing, and likely does not suggest that the languages are related. Another example comes from the Lummi Salish language; the word for *father* in Lummi is /mæn/, looking a lot like, but having nothing to do with English *man.* Likewise, the Chinook Jargon word, *siyápuł,* and French *chapeau* both mean ‘hat.’ Sometimes the associations are created. Nootka Jargon *łutsma* ‘woman’ was changed by English speakers to be more like the word ‘woman,’ coming into Chinook Jargon as *łuchman*.

1. See the appendix for a partial list of non-Indo-European languages. [↑](#footnote-ref-1)
2. The word *senile* is not a Germanic-based word but rather a loan word from Latin via Old French. There is a reconstructed Proto-Germanic form \**sinaz.* [↑](#footnote-ref-2)