

II. LAS LENGUAS DE LA TIERRA

1. NÚMERO Y DISTRIBUCIÓN GEOGRÁFICA DE LAS LENGUAS

Escribe George Steiner que los hombres son iguales en toda la Tierra. Todos respiran el mismo elemento químico y fallecen si carecen de él. Tienen el mismo número de dientes y de vértebras. Los etnógrafos dividen la especie humana en cuatro o siete “razas”,¹ dependiendo del tipo de clasificación utilizada. Existen seis tipos de sangre y tres tipologías de tamaños y estructuras óseas. Sería lógico, según Steiner, que las lenguas humanas también se dividieran en un número ínfimo, digamos media docena, de lenguas.²

Los pájaros de igual especie “hablan” el mismo lenguaje en todas las latitudes. Lo mismo ocurre con los lobos, los perros, las abejas y los demás animales. La realidad humana es radicalmente diferente. No se habla media docena ni medio centenar de lenguas. Hoy en el mundo se hablan aproximadamente 7 000 lenguas y dialectos³ (véase el cuadro II.1).

El número de hablantes de cada lengua es también muy diverso. Existen nada más 10 “megalenguas”, habladas por más de 100 millones de personas, y, en el polo inverso, 1 955 lenguas de menos de 1 000 hablantes. El 95% de las lenguas son habladas por grupos de menos de un millón de personas. El cuadro II.2 muestra la distribución de lenguas según el número de hablantes.

¹ Lo pongo entre comillas porque, en realidad, todos los humanos forman parte de una sola raza, la raza humana.

² George Steiner, *After Babel. Aspects of Language and Translation*, Oxford University Press, Oxford, 1998, p. 52. (Hay edición en español: *Después de Babel: aspectos del lenguaje y la traducción*, 3ª ed., FCE, México, 2001.)

³ La fuente es el Summer Institute of Linguistics (SIL), grupo misionero protestante interesado en las lenguas del mundo para la traducción y difusión de la Biblia en todos los países. Sus estadísticas generalmente son confiables.

CUADRO II.1. *Distribución geográfica de lenguas y dialectos en 2014*

Región	Número de lenguas vivas	Porcentaje del total	Número de hablantes (millones de personas)	Número de lenguas en extinción
Asia	2 294	32.3	4 359	191
África	2 144	30.2	1 178	134
Pacífico	1 313	18.5	39.3	204
América	1 061	14.9	975	341
Europa	287	4	739	50
TOTAL	7 099	100	7 291	920

FUENTE: Gary F. Simons y Charles D. Fennig (eds.), *Ethnologue: Languages of the World*, 20ª ed., SIL International, Dallas, Texas, 2017.

CUADRO II.2. *Distribución de lenguas según el número de hablantes nativos 2014*

Número de hablantes por lengua	Número de lenguas	Porcentaje de lenguas (en relación con el número total de lenguas habladas)	Número de hablantes
Más de 100 millones	10	0.14	3 158 millones
Entre 100 y 10 millones	80	1.1	2 809 millones
Entre 10 y un millones	307	4.3	949 millones
Entre un millón y 100 000	956	13.5	305 millones
Entre 100 000 y 10 000	1 811	25.5	61.8 millones
Entre 10 000 y 1 000	1 980	27.9	7.63 millones
Entre 1 000 y 150	1 064	15.0	470 000
Menos de 100 y (desconocidos)	891	12.6	12 900
TOTAL	7 099	100	7 291 millones

FUENTE: Gary F. Simons y Charles D. Fennig (eds.), *Ethnologue: Languages of the World*, 20ª ed., SIL International, Dallas, Texas, 2017.

2. LAS LENGUAS MÁS HABLADAS

Veamos primero la lista de lenguas más habladas, medidas por el número de hablantes nativos, o sea, las lenguas llamadas maternas o aprendidas en el seno familiar.

CUADRO II.3. *Los 20 idiomas más hablados*

Idioma	Hablantes nativos (millones de personas)	Porcentaje de la población mundial
Chino mandarín	898	12
Español	478	6.6
Inglés	372	5.1
Hindi/urdu	329	4.5
Árabe	295	4
Bengalí	242	3.3
Portugués	219	3
Ruso	154	2.1
Japonés	128	1.8
Lahnda (Pakistán)	119	1.6
Subtotal. Hablantes nativos de las 10 lenguas más habladas (millones)	3 234	44% de la población total
Javanés (Java)	84	1.2
Chino Wu	80.1	1.1
Coreano	77	1.1
Alemán	77	1.1
Francés	76	1
Telugu (India)	74	1
Chino Yue	72.9	1
Maratí (India)	72	1
Turco	71	1
Tamil (India y Pakistán)	68	0.9
Total. Habitantes nativos de las 20 lenguas más habladas (millones)	3 985	54.7% de la población total

FUENTE: Gary F. Simons y Charles D. Fennig (eds.), *Ethnologue: Languages of the World*, 20ª ed., SIL International, Dallas, Texas, 2017.

Las 10 lenguas más habladas representan 0.14% del total de lenguas, pero son habladas por 44.4% de la población mundial; 3 985 millones de personas, o sea, 54.7% de la población mundial, tienen como lengua materna alguna de las 20 lenguas más habladas de la tierra, 0.28% del total de lenguas.

Si incluyéramos entre hablantes de una lengua además de los “nativos” a aquellos que la han aprendido como segunda lengua —lengua vehicular o franca—, el orden cambiaría de manera importante.

CUADRO II.4. Las 10 lenguas más habladas o “megalenguas”, incluidas la lengua materna y la segunda lengua

Lengua	Millones de hablantes	Países o regiones del mundo
Chino mandarín	1 092	China, Taiwán, Singapur, Malasia
Inglés	984	Estados Unidos, Gran Bretaña, Canadá, Australia, Nueva Zelanda, países de Asia, África, India, etcétera
Español	552	Hispanoamérica, España, Estados Unidos, Filipinas, países de África, etcétera
Hindi / urdu	544	India, Pakistán, Mauricio
Árabe	422	Arabia Saudita, del este de África al Golfo Pérsico y al Medio Oriente (todas las variedades)
Ruso	268	Rusia, antiguas repúblicas soviéticas, Europa del Este
Bengalí	262	Bangladesh, India, Nepal
Francés	230	Francia, Bélgica, Suiza, Luxemburgo, Canadá, países de África, Caribe, islas del Pacífico, etcétera
Portugués	229	Portugal, Brasil, Angola, Mozambique, etcétera
Japonés	128	Japón, Brasil, Estados Unidos

FUENTE: Gary F. Simons y Charles D. Fennig (eds.), *Ethnologue: Languages of the World*, 20ª ed., SIL International, Dallas, Texas, 2017.

Como ya señalé, una *lengua materna* o *lengua nativa* es aquella que el individuo aprende en el seno familiar. *Lengua vehicular* o *lengua franca* es la que se aprende fuera del seno familiar y se utiliza para la intercomunicación allí donde conviven comunidades de lenguas diferentes. La principal es el inglés. El francés, el árabe, el hindi, el ruso y el suajili son también ejemplos de lenguas vehiculares, en tanto que no lo son lenguas tan importantes como el japonés o el alemán.

Las estadísticas acerca del número de lenguas y de sus correspondientes hablantes no son siempre confiables. Se prestan a confusión los enfoques con los que se estudian dialectos muy parecidos, el bilingüismo y lenguas casi idénticas pero habladas en países separados por grandes diferencias políticas o religiosas.⁴ Por otra parte, existen muchas lenguas insuficientemente conocidas y que carecen de escritura. Es por tanto imposible establecer con precisión su número de hablantes.

3. EXPANSIÓN Y CONTRACCIÓN EN EL NÚMERO DE LENGUAS: LA EXTINCIÓN DE LAS LENGUAS

Con la dispersión de los grupos humanos por la faz de la Tierra, la lengua primigenia de hace 130 000 años se fue transformando y paulatinamente se fue convirtiendo en los miles de idiomas que se hablan hoy. El número de lenguas se fue multiplicando durante los milenios pero, en algún momento, el proceso de multiplicación se detuvo y se inició un proceso de signo contrario. En nuestros días somos testigos de una veloz disminución del caudal lingüístico.

¿Cuándo y por qué se produjo tal inversión? Los idiomas se multiplican con el crecimiento y la dispersión de la población. Se extinguen a causa de la influencia de idiomas nacionales o imperiales, a cuya sombra numerosas lenguas minoritarias son paulatinamente absorbidas y acaban por desaparecer. El latín, por ejemplo, sustituyó al etrusco y a lenguas osco-umbras que se extinguieron en los primeros siglos de nuestra era; pero, al contrario de estas lenguas, el latín dio origen, con el paso de los siglos, a un buen número de lenguas romances y se extinguió como lengua viva. Se estima que una lengua

⁴ En las estadísticas se presentan como dos lenguas diferentes el urdu, hablado en Pakistán, y el hindi, hablado en India, a pesar de que son casi idénticos.

se transforma en lenguas *filiales* bien diferenciadas en un periodo de 500 a 1 000 años.

El proceso de creación y desaparición de lenguas es continuo. Resulta una verdad de Perogrullo o digna de Monsieur de La Palice⁵ el decir que el caudal lingüístico empieza a disminuir cuando el ritmo de extinción empieza a superar al de creación de idiomas. Este punto se dio a partir de mediados del siglo XVI. Es probable que el número de lenguas haya alcanzado entonces un máximo de alrededor de 15 000.⁶

La desaparición de lenguas fue ocasionada por múltiples factores, entre los que destacan las guerras de conquista, las enfermedades traídas por los conquistadores, que diezmaron las poblaciones aborígenes; el crecimiento del nacionalismo y la necesidad de hablar las nuevas lenguas de la administración, el comercio y la enseñanza. Durante el siglo XX se produjo un vertiginoso aumento en el número de hablantes de unas cuantas *megalenguas* a expensas de multitud de lenguas minoritarias. Esto fue resultado de las dos guerras mundiales, el surgimiento de la Unión Soviética en un territorio inmenso y plurilingüístico,⁷ la industrialización, el servicio militar, el desarrollo de los transportes carreteros y ferroviarios, la radio (1920), la televisión (1950) y, en una palabra, la globalización. El hecho es que el número de lenguas y dialectos ha disminuido de un máximo probable de 15 000 a aproximadamente 7 000 hacia 2015. Este fenómeno continúa a ritmo acelerado en nuestros días.

La extinción de lenguas recuerda la de las numerosas especies animales que desaparecen continuamente, aunque por distintas razones.

De la misma manera que se hace con las especies, podemos clasificar las lenguas en tres grupos:

- *Lenguas en extinción.* De los aproximadamente 7 000 idiomas actuales, poco menos de 2 000 —28% del total— son hablados por menos de 1 000 hablantes, como se vio en el cuadro II.2. En general, estas lenguas ya no son aprendidas por los niños y desaparecerán en no más de 50 años.

⁵ Bernard de La Palice, mariscal de Francia, fue objeto de una canción que, al parecer, terminaba con esta línea: "Una hora antes de su muerte, estaba todavía con vida". Ésta fue la primera "verdad de La Palice".

⁶ Daniel Nettle y Suzanne Romaine, *Vanishing Voices. The Extinction of the World's Languages*, Oxford University Press, Oxford, 2000.

⁷ Más de 70 lenguas se extinguieron en la URSS durante el periodo soviético, ante el esfuerzo de crear nuevas generaciones de ciudadanos rusohablantes para el gobierno central (*ibid.*, p. 196).

- *Lenguas en peligro.* Es probable que unas 2 000 lenguas adicionales ya no existan hacia el año 2100. En otras palabras, dentro de un siglo quedarán alrededor de 3 100 lenguas, es decir, menos de la mitad que en nuestros días.
- *Lenguas no amenazadas actualmente.* Las 3 100 restantes, encabezadas por las megalenguas.

Es muy rápido, pues, el ritmo de desaparición de lenguas. Aproximadamente 2 000 lenguas se extinguirán en los próximos 50 años, y 4 000 en un siglo. Ello representa, en promedio, 40 lenguas por año, o sea, 3.3 lenguas por mes. En cambio, unas cuantas, encabezadas por el chino mandarín, el inglés, el hindi y el español, crecen, prosperan y ocupan espacios cada vez mayores. Estas cuatro lenguas son habladas como lenguas nativas por casi 30% de la población mundial. Si incluimos a hablantes nativos y a quienes las han aprendido como segunda lengua, el porcentaje sube a más de 40%. Es irrefrenable la tendencia hacia la concentración de un creciente número de hablantes en un número cada vez menor de lenguas. En nuestros días, 90 lenguas son habladas por 82% de la población mundial. Dentro de un siglo, alrededor de 15 a 20 lenguas serán habladas por 95% de la población.

Cada lengua que se extingue representa un enfoque humano que se pierde. La extinción colectiva de lenguas es un drama cultural de importancia que rara vez valoramos en su magnitud real. Miles de idiomas han desaparecido a lo largo de la historia. De la mayoría desconocemos incluso el nombre. Pero también han desaparecido las lenguas de algunas de las más grandes civilizaciones, como la sumeria y la egipcia. Del etrusco, la principal lengua preindoeuropea de Italia, no sabemos prácticamente nada. Palabras y lenguas habladas hoy en buena parte de la Tierra caerán en el olvido y serán apenas descoloridos jeroglíficos dentro de algunos milenios.

Lenguaje, idiomas, lenguas y dialectos

Existe una persistente confusión acerca de estos términos.

Lenguaje es el conjunto de sonidos articulados y de otras formas de comunicación que el hombre utiliza para manifestar lo que piensa o siente.

Se puede hablar del lenguaje de la música, por medio del cual el hombre expresa las más profundas e inefables emociones. Se puede, por extensión, hablar del lenguaje de las ballenas, pero sería incorrecto decir que en México y en Japón hablamos lenguajes diferentes.

Lengua o *idioma* es el sistema de comunicación verbal propio de un pueblo o nación, o común a varios. El español, portugués, catalán y ruso son lenguas.

Dialecto es un término con dos acepciones. Es, en primer lugar, una lengua que forma parte de una unidad lingüística mayor o que deriva de un tronco común. Así, podemos decir que el español, el portugués y el catalán son dialectos derivados del latín. En su segunda acepción, es “un sistema lingüístico derivado de otro pero sin diferenciación suficiente frente a otros de origen común”.⁸ El término no tiene nada de peyorativo pero ocasiona discusiones a veces violentas. Tal como yo lo empleo en este libro, el tzotzil y el tzeltal son dialectos mayas, como el asturiano o el leonés son dialectos del español. El catalán, el gallego y el español son dialectos latinos, pero, al mismo tiempo, son tres lenguas diferentes.

4. IDENTIFICACIÓN Y DESCUBRIMIENTO DE LAS GRANDES FAMILIAS LINGÜÍSTICAS

Las lenguas de la Tierra están agrupadas en familias lingüísticas, es decir, en grupos de lenguas que comparten orígenes comunes. Empezaré por el caso mejor estudiado, el de la familia indoeuropea.

La familia indoeuropea

En 1783 sir William Jones llegó a Calcuta para ocupar un puesto de juez en el Tribunal Superior de la Compañía Británica de la India Oriental. Sir William era un lingüista aficionado y se puso a estudiar la antigua lengua llamada sánscrito. Con asombro se percató de ciertas semejanzas entre el sánscrito y otras dos lenguas de la Antigüedad, el griego y el latín. La palabra *padre* es

⁸ *Diccionario de la lengua española*, Espasa-Calpe, Madrid, 2001.

pitar en sánscrito, *pater* en griego y en latín. *Madre* es *matar* en sánscrito y *mater* en griego y latín. Analizó otras numerosas palabras y en 1786 dictó en la Sociedad Asiática de Calcuta una conferencia en la que demostró la afinidad entre esas tres lenguas, “una afinidad mucho mayor que la que podría existir por mera casualidad; una afinidad tan fuerte que todo filólogo debe concluir que provienen de una rama común que quizá ya no exista”.⁹ Sugirió, por otra parte, la posibilidad de que el antiguo persa, el viejo germánico y el viejo celta pertenecieran también a la misma familia, como probaron años después los filólogos Franz Bopp, Rasmus Rask y Jacob Grimm.

Hoy sabemos que de ese tronco común provienen las lenguas latinas, germánicas, eslavas, celtas, el viejo persa y sus descendientes, el sánscrito y sus descendientes, el griego y muchas más.

El descubrimiento del origen común de las lenguas llamadas indoeuropeas a partir de entonces fue uno de los más trascendentales avances de la lingüística. El siguiente paso fue la búsqueda de su lugar de origen común. Si persas, indios, italianos, rusos, ingleses y demás “indoeuropeos” hablaban lenguas derivadas del mismo tronco, lógico era inferir que sus antepasados provenían de una misma zona geográfica.¹⁰ Si analizamos una serie de palabras de diversas lenguas indoeuropeas y les seguimos la pista hacia el pasado remoto, se puede hacer una reconstrucción de la lengua de los primeros indoeuropeos y formarnos una idea del estilo de vida de sus habitantes. Se ha dicho que mucho se puede saber de la historia de un pueblo a través de su vocabulario. Por las palabras que usaban —tales como *caballo* o *invierno*— se puede deducir que aquellos primeros indoeuropeos llevaban una vida seminómada, que tenían animales domésticos y cultivaban la tierra. Ninguna de las lenguas protoindoeuropeas contaba en su vocabulario con la palabra *mar*. El filólogo Friedrich Max Müller (1823-1900), quien desarrolló la paleontología lingüística, se dio a la tarea de identificar los términos geográficos, botánicos, zoológicos y climáticos comunes a esas lenguas. Si en las diferentes lenguas las palabras *oso*, *pino* y *nieve* son parecidas y, por el contrario, difieren los vocablos *león*, *palmera* y *cocodrilo*, se puede deducir

⁹ Conferencia de W. Jones en Calcuta, citada en Robert McCrum, Robert MacNeil y William Cran, *The Story of English*, Elisabeth Sifton Books, Viking, Nueva York, 1986, p. 51.

¹⁰ Lo de “misma zona” es relativo. Si nos remontamos 100 milenios más, la zona geográfica de donde provinieran sería probablemente África, tal como vimos en el capítulo anterior.

que la patria común no era tropical, sino templada o fría; que la fauna incluía osos, y la flora, pinos.

Este tipo de análisis llevó a muchos indoeuropeístas a ubicar la patria común en las estepas del sur de Rusia y de Ucrania. Allí vivían los indoeuropeos hace 5 000-7 000 años, y de allí emigraron en muy diversas direcciones.¹¹

El mapa II.1 muestra el recorrido de los pueblos indoeuropeos desde su probable lugar de origen. Dos innovaciones fundamentales contribuyeron a las migraciones de estos pueblos: la rueda y la domesticación del caballo. Oleadas de tribus se esparcieron hacia el oeste, este, norte y sur y llegaron a Grecia (los aqueos), al Asia Menor (los hititas), a Mesopotamia (los mitanos), a India (los arios) y a Europa (los celtas, itálicos, eslavos, bálticos, germánicos y otros). Tribus indoeuropeas que se habían detenido en el noroeste de Europa descendieron siglos después hacia el sur. Eran los bárbaros germánicos invasores del Imperio romano: francos, alamanes, suevos, visigodos, ostrogodos, vándalos y otros.

La familia indoeuropea es la mayor del mundo tanto por el número de hablantes como por su dispersión geográfica y por la importancia de sus distintos idiomas.

En el diagrama II.1 se advierte la integración de la familia lingüística indoeuropea, subdividida en 12 grupos lingüísticos, que representan en cifras redondas medio centenar de lenguas contemporáneas y varias docenas de dialectos, hablados por más de la mitad de la población mundial.

Muy distintos destinos han conocido los diversos grupos; unos han crecido vertiginosamente en tanto que otros se han reducido a una mínima expresión.

Grupo itálico. Se divide en el subgrupo osco-umbro, que se extinguió, y el subgrupo latino. Más adelante analizaré en detalle el caso de las lenguas latinas. Aquí me limitaré a señalar que este grupo está integrado por las siguientes lenguas:

¹¹ Otras teorías, tales como la de los indoeuropeístas soviéticos Gamkrelidze e Ivanov (1984), indican que los indoeuropeos estarían establecidos hace 7 000-8 000 años en Armenia. Para el inglés Renfrew, la patria común era Anatolia (actual Turquía) y de allí iniciaron sus migraciones (Enrique Bernárdez, *¿Qué son las lenguas?*, Alianza Editorial, Madrid, 2001, pp. 86 y ss.). La lengua de los que se quedaron sería el anatolio, que incluye lenguas como el hitita, que sería así lengua hermana y no hija del indoeuropeo.

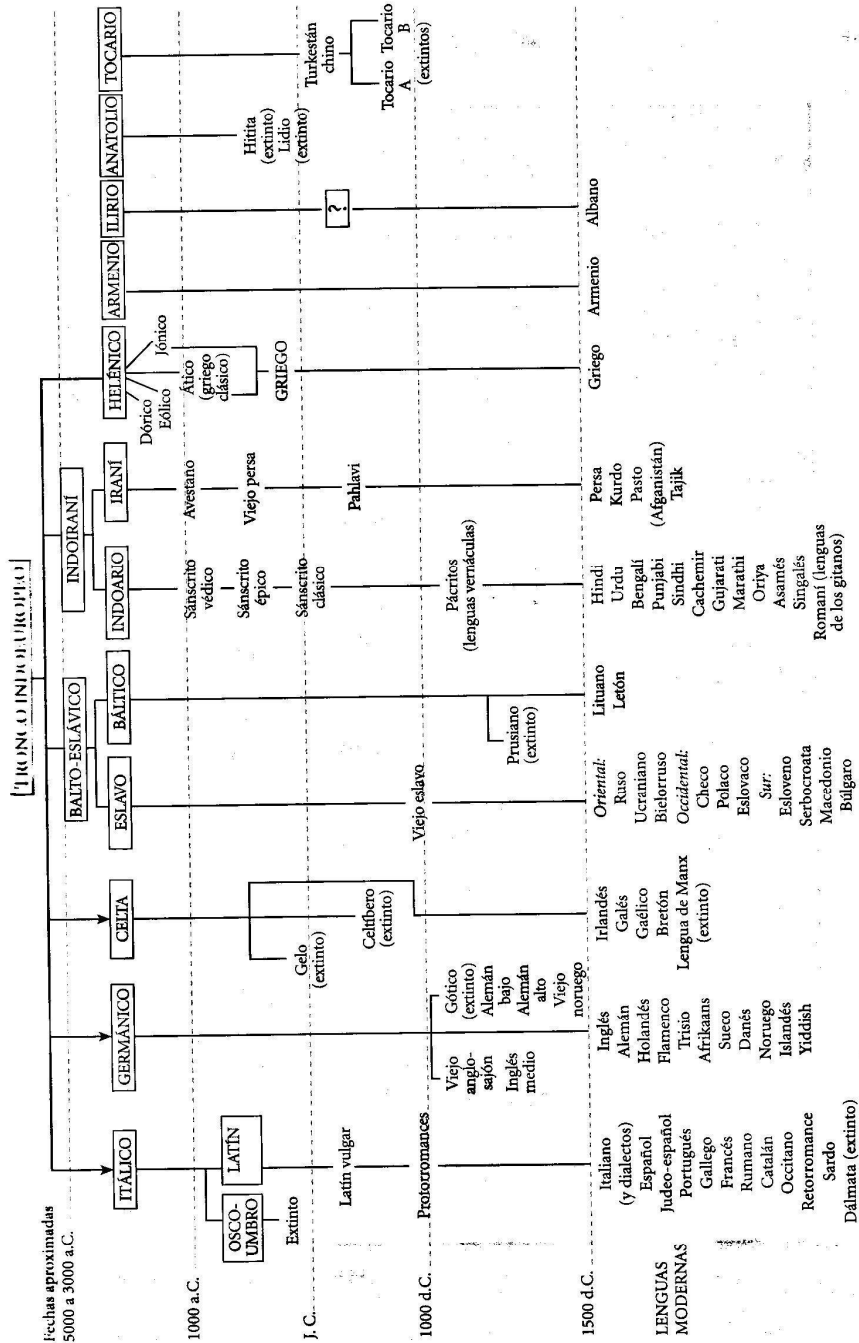
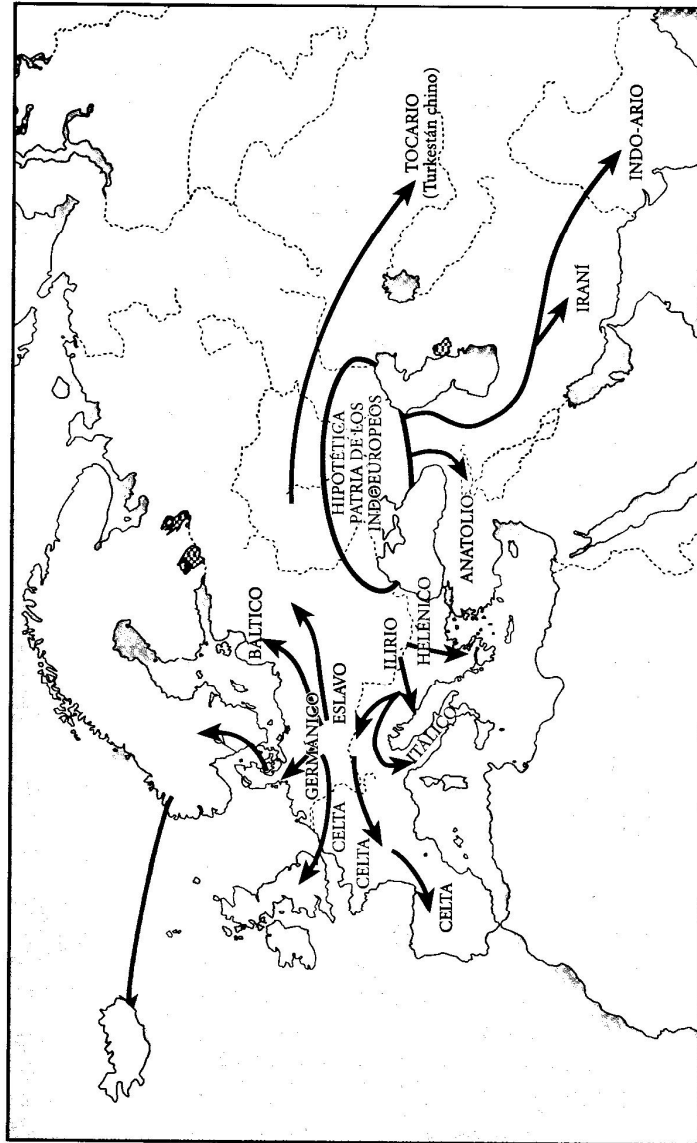


DIAGRAMA II.1. La familia indoeuropea de lenguas.



MAPA II.1.1. La expansión indoeuropea según la teoría tradicional y la primera distribución de grupos lingüísticos indoeuropeos.

Italiano	Rumano
Español	Catalán
Judeo-español ¹²	Occitano
Portugués	Sardo
Francés	Retorromance o retofriulano

Grupo germánico. De oscura existencia en los primeros 15 siglos de nuestra era, empezó a cobrar gran importancia a partir de 1500. Lo integran principalmente:

Inglés	Noruego
Alemán	Danés
Holandés	Islandés
Afrikaans ¹³	Yiddish ¹⁴
Sueco	Otras

Grupo eslávico. Al igual que el grupo germánico, partió de una oscura existencia hasta convertirse en uno de los principales grupos lingüísticos de nuestra época, sobre todo por la gran difusión del ruso. El grupo eslávico, que examinaré en mayor detalle en el capítulo XI, está constituido por las siguientes lenguas:

Ruso	Búlgaro
Polaco	Macedonio
Bielorruso	Serbocroata
Ucraniano	Eslovaco
Checo	Esloveno

Grupo celta. Los celtas fueron de los primeros indoeuropeos en poblar Europa. Hacia el siglo III a.C. se extendían desde el Mar del Norte al Mediterrá-

¹² Lengua hablada por los judíos sefarditas, es decir, los descendientes de los judíos expulsados de España en 1492.

¹³ Lengua cercana al holandés, hablada sobre todo por los descendientes de los colonizadores holandeses en Sudáfrica.

¹⁴ Alemán mezclado sobre todo con hebreo. *Yiddish* proviene de *jüdisch-deutsch sprache*, o lengua judío-alemana. *Jüdisch* se convirtió en *yiddish*.

neo, y desde España hasta el Mar Negro. La Galia de los romanos era inicialmente el conjunto de regiones habitadas por los celtas (los *galli*). A partir del siglo I a.C., el nombre Galia se limitó a algo más que el territorio actual de Francia. Testimonio de su antigua extensión son los nombres que dejaron en Europa y Asia: Galia en Francia, Galicia en España, el país de Gales británico, la Galitzia en Polonia y la Galacia en Turquía.¹⁵

Las lenguas celtas que sobreviven hoy son las siguientes:

Irlandés. Hablada por 20% de la población de la República de Irlanda —donde es lengua oficial, junto con el inglés— y por un porcentaje menor en Irlanda del Norte.

Galés. Lengua oficial del País de Gales, junto con el inglés, y la más floreciente de las lenguas celtas.

Gaélico. Variedad del celta, llevada por los irlandeses al norte de Escocia, a partir del siglo V de nuestra era. En 1961 no quedaban sino 81 000 hablantes del gaélico, pero sólo para 1 000 personas era lengua única.¹⁶

Bretón. Hablada en diversos dialectos en la región francesa de Bretaña.¹⁷

La lengua de Manx (de la isla de Man). Se extinguió en 1974, con la muerte de Ned Maddrell, el último hablante de esa lengua.¹⁸

Grupo báltico. Está integrado únicamente por el lituano y el letón.¹⁹

El griego. Lengua clásica por excelencia de la Antigüedad, el griego era además *lingua franca* en numerosas regiones de Grecia, Italia y el Asia Menor. Los primeros documentos en griego, descubiertos en Creta, datan del siglo XIV a.C.

Cuatro dialectos fueron los principales: el *jónico* —utilizado por Homero en la *Ilíada*, la primera obra literaria de la cultura occidental—, el *dórico* de

¹⁵ El nombre de *celtas* (*Keltoi*: celta) aparece por primera vez en los autores griegos del siglo V a.C. En el siglo III a.C. se llama *gálatas* (*galatae*) a los celtas que en 279 llegaron a amenazar el santuario de Delfos dedicado a Apolo. El historiador Polibio emplea los nombres *celta* y *gálatas* para designar indistintamente a los celtas localizados en Occidente u Oriente.

¹⁶ Henriette Walter, *L'Aventure des langues en Occident*, Robert Laffont, París, 1994, p. 100.

¹⁷ El bretón no es heredero directo del galo, que se extinguió, sino que fue introducido en la Bretaña por galeses y cónicos que huían de sus tierras ante la presión de las invasiones sajonas y anglos.

¹⁸ Daniel Nettle y Suzanne Romaine, *op. cit.*, p. 2.

¹⁹ Como veremos después, el estonio y el finlandés pertenecen a un grupo diferente, el urálico-yukaghir.

Esparta, el *eólico* de Tesalónica y Lesbos, y el *ático*, la lengua de Atenas, vehículo de una obra literaria, filosófica y científica sin paralelo en la historia.

El griego alcanzó su máxima difusión en la época de Alejandro Magno. Los distintos dialectos se unificaron en la época helenística y formaron una lengua común o *koiné* que se utilizó durante 1 000 años. El griego continuó siendo la lengua oficial del Imperio romano de Oriente pero, a diferencia del latín, no procreó lenguas en sus territorios y se fue reduciendo paulatinamente hasta quedar restringido, en nuestros días, a Grecia y Chipre y a lo poco que queda de las antiguas colonias griegas del sur de Italia, Calabria y Córcega. En la región calabresa de Lecce existen aún nueve poblaciones de habla griega. La principal es Calimera ("Buen día", en griego). El griego actual, la lengua oficial de Grecia y cooficial de Chipre, es el *demótico*, que proviene de la *koiné* helénica.

El sur de Italia se llamaba Magna Grecia porque había allí más griegos que en la propia Grecia. La lengua de esa región siguió siendo el griego hasta hace seis o siete siglos. Quedan cerca de 20 000 hablantes de griego entre Italia y Córcega, en su mayoría ancianos. Hablan dialectos diferentes (el *salento* y el *aspromonte*) que están en peligro de extinción.

Es riquísimo el aporte lingüístico y cultural que hizo el griego a numerosas lenguas, empezando por el latín, como veremos más adelante.

El albaniano. La lengua de Albania y Kosovo está aislada, sin hermanos, dentro de la familia indoeuropea.

El armenio. Es un caso parecido al albaniano: miembro único de su grupo.

El hitita. Lengua desaparecida, que se habló en Anatolia (Turquía y norte de Siria). Su descubrimiento tuvo lugar a fines del siglo XIX, cuando se encontró un archivo de antiquísima correspondencia diplomática egipcia. Dos cartas en un idioma desconocido quedaron en el misterio hasta que excavaciones en Turquía en 1905-1907 arrojaron miles de tabletas inscritas en la misma lengua, probablemente la de los hititas mencionados en la Biblia. La lengua hitita²⁰ fue descifrada poco después y se piensa que podría formar parte de un grupo más amplio llamado anatolio.

²⁰ Merritt Ruhlen, *A Guide to the World's Languages*, Stanford University Press, Stanford, CA, 1991, p. 55.

Grupo iraní. Se advertirá en el diagrama II.1 que tanto las lenguas iraníes como las indoarias derivan de la misma rama indoiraní. Las iraníes tienen por tanto un parentesco más cercano con las de India que con los otros idiomas indoeuropeos. Notable similitud tienen las antiguas lenguas de Persia e India, el persa avestano y el sánscrito de los Vedas.

Los principales idiomas iraníes contemporáneos son:

Persa o farsi, en Irán.

Kurdo, en Irán, Turquía, Irak y Siria.

Afgano (o pasto), en Afganistán.

Tajik, en Tajikistán.²¹

Grupo indoario. Las tradiciones religiosas de los arios, conquistadores de la India, se preservaron durante siglos mediante una rigurosa e increíble tradición oral, la de los Vedas o *Libros del conocimiento*. El más antiguo e importante es el *Rig-Veda* o *Versos del conocimiento*, y consta de 1017 poemas en sánscrito. Datan aproximadamente de 1500 a.C. —varios siglos antes que la *Iliada* y la *Odisea*— y no fueron plasmados por escrito sino hacia 600 a.C. Este libro constituye el más antiguo texto de la literatura indoeuropea. El sánscrito es el padre de las lenguas indoarias, al igual que el latín lo es de las lenguas romances.

El idioma de las dos grandes obras épicas, el *Mahabharata* y el *Ramayana*, compuestas algunos siglos antes de Cristo, se conoce precisamente como sánscrito “épico”.

El sánscrito “clásico” alcanzó su apogeo en los siglos IV a VI de nuestra era, en la época de los guptas. Grandes obras se escribieron entonces en el sánscrito clásico, que ha quedado desde entonces como lengua literaria de poca evolución. Sigue hoy usándose como lengua literaria y es uno de los idiomas nacionales de la India.

²¹ El sufijo *-stan* significaba “tierra”, “país”, en viejo persa. Es el origen de los nombres de los países de Asia Central: Afganistán (país de los afganos), Tajikistán (de los tajikos), Turkmenistán (de los turkmenos), Kirguistán (de los kirguizes), Uzbekistán (de los uzbekos), Kazajistán (de los kazajos), y de regiones tales como Kurdistán (de los kurdos), Baluchistán (de los baluchis), etc. Pakistán es un curioso acrónimo compuesto con letras de las diversas regiones: P(unjab), A(fgania), K(ashmir), I(rán), S(ind), T(ujaristán), Af(ganistán) y (balochistán)N. Al mismo tiempo significa “el país de los ‘paks”, los “espiritualmente puros” (véase Rahmat Ali en la bibliografía).

Las principales lenguas indoarias contemporáneas y las regiones donde se hablan son las siguientes:

Hindi (India)

Urdu²² (India y Pakistán)

Bengalí (Bangladesh e India)

Punjabi (India)

Sindhi (India)

Cachemir (Cachemira)

Gujarati (India)

Marathi (India)

Oriya (India)

Asamés (India)

Romaní (lenguas de los gitanos esparcidos por el mundo)²³

Singalés (Sri Lanka)

El tocario. Lengua desaparecida de China sudoccidental y descubierta a principios del siglo XX en dos variantes, llamadas, a falta de mejor nombre, tocario A y tocario B. Su descubrimiento en unos antiguos manuscritos fue una triple sorpresa. Primero, por tratarse de una lengua indoeuropea totalmente desconocida. Segundo, porque los manuscritos, escritos entre 600 y 800 d.C., se encontraron en las cuevas de un monasterio budista en el Turkestan chino, a más de 1500 kilómetros de la región más próxima de lengua indoeuropea, India, y, finalmente, porque la lengua resultó más cercana a las lenguas de Europa que a las indoarias. Fue una sorpresa análoga a la que produciría el descubrimiento en España de una lengua emparentada con el chino.

Algunos cuadros comparativos de las lenguas indoeuropeas. En los cuadros II.5 y II.6 aparece una serie de palabras que demuestran el parentesco de diversos idiomas indoeuropeos.

Existen numerosas palabras del sánscrito que tienen una semejanza bastante clara con su equivalente en lenguas latinas y otras lenguas indo-

²² El urdu es prácticamente igual al hindi pero, al ser hablado en zonas predominantemente musulmanas, utiliza el alfabeto arabopersa y no el *devanagari* del hindi.

²³ Se podrá leer una breve explicación acerca del romaní en el capítulo IV.

CUADRO II.5. Nombres de los números del 2 al 10 en diversas lenguas indoeuropeas

Español	dos	tres	cuatro	cinco	seis	siete	ocho	nueve	diez
Sánscrito	dva	tri	catur	paca	sat	sapta	asta	nava	dasa
Griego	duo	treis	tettares	pente	hex	hepta	okto	ennea	deka
Latín	duo	tres	quattuor	quinque	sex	septem	octo	novem	decem
Francés	deux	trois	quatre	cinq	six	sept	huit	neuf	dix
Ruso	dva	tri	chetyre	pyat	shest	siem	vosiem	dieviat	diesiat
Inglés	two	three	four	five	six	seven	eight	nine	ten
Alemán	zwei	drei	vier	fünf	sechs	sieben	acht	neun	zehn
Hindi	do	tin	char	panch	chhe	sat	ath	nau	das

CUADRO II.6. Algunas palabras usuales en diversas lenguas indoeuropeas

Español	padre	madre	fray (hermano) ^a	sor (hermana) ^b	día	noche	nieve	nuevo	joven
Sánscrito	pitar	matar	brhatar	svasar	divasah (ratrau) ^d	hima	newas	yuvan (o nava)	
Griego	pater	mater	phrater	adelfi	mera	nuktos	nipha	neos	neos
Latín	pater	mater	frater	soror	dies	noctis	nivis	novus	juvenis
Francés	père	mère	frère	soeur	jour ^c	nuit	neige	nouveau	jeune
Ruso	(otiets)	mat'	brat	sestra	dien	noch'	sneg	novy	yunni
Inglés	father	mother	brother	sister	day	night	snow	new	young
Alemán	Vater	Mutter	Bruder	Schwester	Tag	Nacht	Schnee	neu	jung
Hindi	pita	man	bai	bahin	din	(rat)	(barf) ^e	naya	javan

^a y ^b Hermano y hermana tienen también una etimología latina, pero diferente. Por eso utilicé los sustantivos *fray* y *sor*, cuyo origen es idéntico al de las otras lenguas.

^c La palabra francesa *jour* también proviene del latín *dies* (día) pero del adjetivo *diurno* y no del sustantivo *dies*.

^d Están entre paréntesis palabras de origen diferente del de la mayoría de las lenguas indoeuropeas.

^e Probablemente de origen árabe.

europas. El sánscrito *raja* significa *rex* en latín o *rey* en español. Son dos términos que provienen de la antigua raíz indoeuropea *reg-*, de la cual descienden también el inglés *right* (derecho) y el alemán *Reich* (reino). El sánscrito *maharajá* proviene de *maha* (magno) y *raja* (rey). El sánscrito *maha*, el griego *mega*, el latín *magnus*, los términos españoles *majestad*, *mayor*, *más*, *maestro* y muchos más tienen un antiquísimo ancestro común: la raíz indoeuropea *meg*.

Las familias lingüísticas. Vista de conjunto²⁴

Eminentes lingüistas como Ruhlen y otros agrupan las 6000 lenguas del mundo en 17 familias. Es una clasificación controvertida, pero ha venido ganando adeptos entre los lingüistas. Veámosla a ojo de pájaro, por orden decreciente de número de hablantes.

1. *Familia indoeuropea*: 3 450 millones de hablantes. Ya fue objeto de comentarios detallados. Lenguas habladas hoy en toda Europa, América, India, Pakistán, Bangladesh, Irán y Afganistán.

2. *Familia sinotibetana*: 1 270 millones de hablantes. Lenguas de China, Tíbet, Nepal, India, Birmania (hoy Myanmar), Tailandia y Laos.

La familia sinotibetana se divide en dos grupos: el grupo chino —mandarín, wu, yue (cantonés), xiang, gan-kejia, min— y el grupo tibetano —tibetano y birmano—.

El chino mandarín es la lengua más hablada del mundo y la principal de China, donde se llama *pǔto-nghuà* (lengua oficial). Algunas de las numerosas lenguas de China difieren tanto del chino mandarín que no permiten la intercomprensión. Ahora bien, el medio que permite la comunicación de todos los chinos es la escritura —la más antigua de la Tierra aún en uso—, basada en ideogramas y no en sistemas fonéticos o alfabéticos. Los ideogramas significan lo mismo, aunque en el habla sean radicalmente diferentes. Así, la literatura china es patrimonio de todas las lenguas chinas aunque éstas no sean inteligibles entre sí. Dos chinos de lenguas distintas se pueden entender

²⁴ Basado en Merritt Ruhlen, *op. cit.*, pp. 286-289. Las cifras de hablantes de lenguas indoeuropeas, sinotibetanas, afroasiáticas y altaicas provienen, por estar más actualizadas, de Rafael del Moral, "Las lenguas del mundo", en *Diccionario Espasa*, Espasa-Calpe, Madrid, 2002.

perfectamente por escrito. Un buen número de ideogramas chinos se entienden incluso en países que lingüísticamente pertenecen a otras familias —como Japón y Corea— y en los que la cultura china ejerció una poderosa influencia.

3. *Familia afroasiática o hamito-semítica*: 400 millones de hablantes. Abarca la mayor parte de las lenguas del Oriente Medio y varias del norte de África (árabe, hebreo, arameo, amárico, egipcio antiguo, chádico, cushítico, bereber, etc.). La familia afroasiática será tema del capítulo XIII.

4. *Familia altaica*: 350 millones de hablantes. Generalmente se divide en dos grupos: *a*) grupo altaico (turco,²⁵ turcomano, kirguizo, azerí, kazajo, mongol) y *b*) grupo coreano-japonés-ainú.

La homogeneidad de esta familia es motivo de grandes discrepancias entre lingüistas. Algunos ubican el coreano, el japonés y el ainú en tres familias diferentes.

5. *Familia áustrica*: 293 millones de hablantes. Sur de China, Vietnam, Camboya, Tailandia, Laos, Indonesia y Filipinas.

6. *Familia niger-kordofana*: 181 millones de hablantes. África Central y del Sur (suajili, zulú, fula, mandinka, yoruba, etcétera).

7. *Familia elamo-dravídica*: 145 millones de hablantes. Sur y oriente de India y sur de Pakistán.²⁶

8. *Familia urálico-yukaghir*: 22 millones de hablantes. Abarca las lenguas fino-ugrias como el húngaro —islote rodeado de lenguas indoeuropeas—, finlandés, estonio, lapón y carelio, y el samoyedo y yukaghir de Rusia.

9. *Familia amerindia*: 18 millones de hablantes. Abarca todas las lenguas aborígenes de América Latina y algunas de los Estados Unidos, pero su homogeneidad es objeto de controversia. El capítulo XIV trata en detalle esta familia.

10-17. Todas las demás familias tienen en su conjunto 19 millones de hablantes. Incluyen las familias khoisán (sur de África), nilosahariana (África Central), australiana, indopacífica (Nueva Guinea, Islas Salomón, Timor,

²⁵ Después de la caída de Constantinopla en 1453, el turco empezó a suplantarse y finalmente a eliminar el griego, hasta entonces predominante en lo que es hoy Turquía.

²⁶ Los idiomas dravidianos se hablaban probablemente en toda la India y en Irán, y quizá fueron introducidos por pastores neolíticos, desplazados más tarde por los invasores arios de lenguas indoeuropeas. La Biblia menciona a los elamitas, cuya lengua, emparentada con los idiomas dravídeos, se hablaba hace 4500 años y se extinguió hace aproximadamente 2000 años.

etc.), na-dené (Alaska, Canadá, los Estados Unidos), esquimo-aleutiana (Alaska, Canadá, Groenlandia, noreste de Rusia), chukchi-kamchatka y caucásica (Georgia).

El mapa II.2 muestra la ubicación geográfica de las 17 familias lingüísticas mencionadas.²⁷

Lenguas genéticamente aisladas

Merecen comentario aparte las lenguas genéticamente aisladas, sin relación probada alguna con ninguna otra lengua. Se trata del etrusco (extinto), el sumerio (extinto), el vasco o euskera, el buruchasquí (norte de India), el nahalí (India central), el gilyak (isla de Sajalín) y el ket (Siberia). En el capítulo IV retornaremos al tema del vasco.

5. LA GENÉTICA HUMANA Y LAS LENGUAS

Ha provocado gran repercusión en la lingüística el reciente descubrimiento hecho por geneticistas sobre la coincidencia de la clasificación de las lenguas y la de los grupos humanos según su parecido genético.²⁸

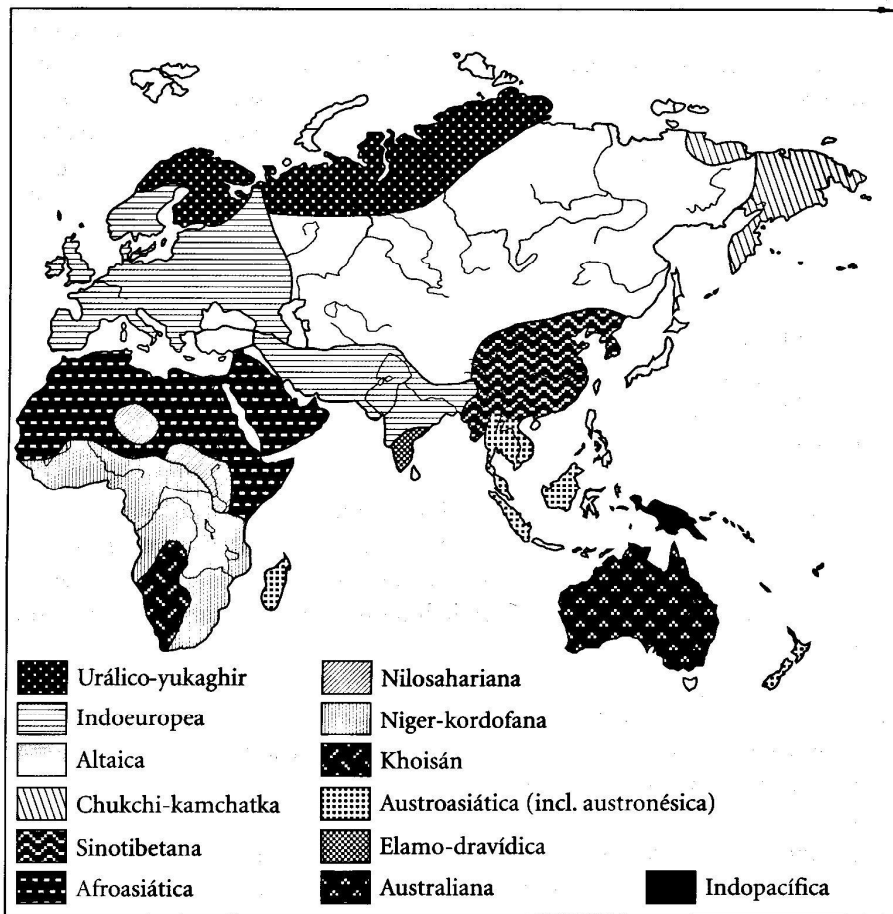
En 1987, Greenberg publicó su clasificación de las lenguas del continente americano en sólo tres macrofamilias: la amerindia, la na-dené y la esquimo-aleutiana. La publicación fue recibida con una tempestad de críticas y burlas pero, a medida que ha pasado el tiempo, se ha roto la unanimidad crítica. Lo notable es que la clasificación de Greenberg en tres macro-familias coincide con las conclusiones de Luigi Luca Cavalli-Sforza,²⁹ que clasificó la población americana en tres grupos humanos, de acuerdo con su tipología genética.³⁰

²⁷ Aún más controvertidas son las teorías que agrupan todas las lenguas de la Tierra en 12 familias —teoría que recibió recientemente el apoyo de genetistas— y la clasificación más extrema de las lenguas en sólo dos macrofamilias, la eurasiática y la nostrática (según el estadounidense Greenberg y los lingüistas rusos y soviéticos, encabezados por Illich Svitych).

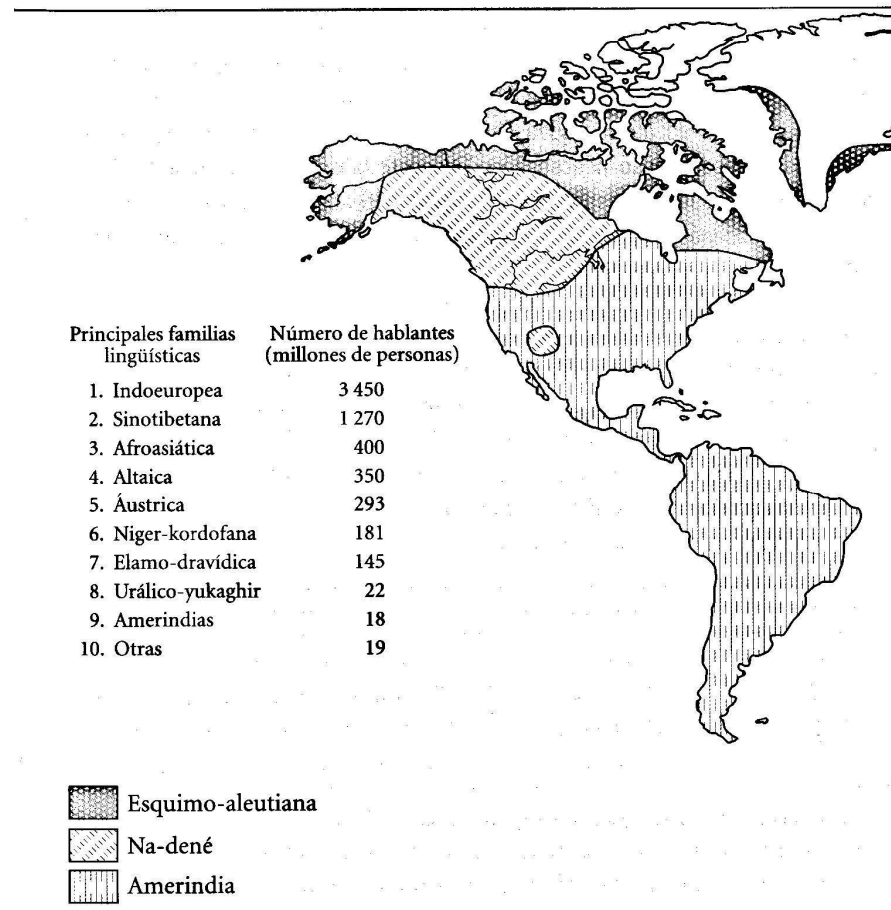
²⁸ Luigi Luca Cavalli-Sforza, *Genes, Peoples, and Languages*, University of California Press, Los Ángeles, 2000.

²⁹ *Idem*.

³⁰ Las clasificaciones lingüísticas y genéticas en cuestión se aplican sólo, por supuesto, a los idiomas y a las poblaciones aborígenes americanas. Las migraciones europeas y africanas distorsionarían por completo el panorama.



MAPA II.2. Distribución de las macrofamilias lingüísticas (según Greenberg/Ruhlen).



6. UN CONTACTO CON LAS LENGUAS MÁS ANTIGUAS

En octubre de 1999 hice una gira de conciertos por Sudáfrica, cuya diversidad lingüística me llamó la atención. Durante la era de Mandela, Sudáfrica pasó de tener dos lenguas oficiales (el inglés y el afrikaans) a 11: inglés, afrikaans, zulú, khoisán, sotho del norte, sotho del sur, tswana, shangaan, swati, venda y ndebele.

Me detendré en la lengua o lenguas khoisanas,³¹ por su interés lingüístico y sus características únicas. Estas lenguas provienen de los khoi (hotentotes) y los san (antes llamados bosquimanos), originarios probablemente de las regiones de Tanzania y Etiopía, donde aparecieron los primeros *Homo sapiens*. Es muy probable, por tanto, que las lenguas khoisanas sean las más antiguas de la Tierra.³² Estas lenguas poseen unos sonidos peculiares y únicos: los chasquidos o clics. Estos clics tienen representaciones escritas y equivalen aproximadamente a lo siguiente:

<i>Símbolo</i>	<i>Nombre en español</i>	<i>Sonido aproximado</i>
0	Clic labial	Un beso muy fuerte
//	Clic lateral	Chasquido para arrear un caballo
!	Clic alveolar	Ruido al descorchar una botella
/	Clic dental	Sonido de disgusto o fastidio

FUENTE: Enrique Bernárdez, *¿Qué son las lenguas?*, Alianza Editorial, Madrid, 2001, p. 141.

³¹ La lengua khoisán se divide en varios dialectos parecidos; se puede, pues, hablar de "lenguas khoisanas".

³² Luigi Luca Cavalli-Sforza, *op. cit.*, indica que las lenguas khoisanas son las más antiguas de las familias lingüísticas africanas, que, como es bien sabido, son las más antiguas de la Tierra.