



ARTÍCULO

## Sobre la filogenia de las lenguas de México *On the philogeny of Mexican Amerindian languages*

Luis Fernando Lara  
El Colegio de México  
Miembro de El Colegio Nacional  
lara@colmex.mx

Original recibido: 2021/02/15

Dictamen enviado al autor: 2021/08/23

Aceptado: 2021/08/23

### Abstract

This article is a critical review of Joshua Greenberg's theory on the phylogenesis of Amerindian languages (especially Mexican languages) as related to Luigi Cavalli-Sforza's theory on the evolutionary genetics of Earth's peoples. It exposes briefly the history and methodology of romance and Germanic linguistics, as considered the source of Amerindian phylogenetic proposals, and reviews the efforts and consequences of Morris Swadesh's lexicostatistics and glottochronology.

**Keywords:** History; methods; Spanish phylogenetic; Amerindian languages

### Resumen

Este artículo es una revisión crítica de la teoría de Joshua Greenberg sobre la filogenia de las lenguas amerindias (en especial las lenguas mexicanas)

en relación con la teoría de Luigi Cavalli-Sforza acerca de la genética evolucionista de la humanidad. El artículo expone brevemente la historia y la metodología de la lingüística romance y la germánica, consideradas como el origen de las propuestas de la filogenia amerindia; además este texto revisa los esfuerzos y los resultados de la lexicoestadística y la glotocronología de Morris Swadesh.

**Palabras clave:** historia; métodos; filogenia del español; lenguas amerindias

El estudio de las lenguas romance desde principios del siglo XIX destaca en comparación con el resto de las familias lingüísticas de la Tierra debido a que contamos con multitud de datos, tanto del latín y del griego, como incluso del celta, el árabe y las lenguas germánicas que dejaron sus huellas en ellas, así como de cada una de las quince lenguas con reconocimiento nacional y sus varias decenas de dialectos. Enumero las quince: rumano, friulano, ladino dolomítico, romanche o retorromano, italiano, sardo, corso, francés, occitano o provenzal, catalán, aragonés, español, asturiano, gallego y portugués. Agrego el sefardí o ladino, lengua de los judíos de origen español y, aunque extinguida, la dalmata, que se habló hasta mediados del siglo XIX en lo que hoy es Croacia (ver Figura 1). El amplio y profundo conocimiento que tenemos de estas lenguas nos ha dado una experiencia de dos siglos, que nos permite tanto documentar las regularidades de su evolución como la influencia determinante de las invasiones en la Antigüedad, de las culturas y también de los accidentes geográficos, como los ríos, los valles y las montañas que tienen su papel en la



Figura 1. Árbol genealógico de las lenguas romances

diversificación de los dialectos y de las lenguas. Tal conocimiento no tiene comparación con el que se tiene de las lenguas germánicas o las eslavas, puesto que casi no hay documentos antiguos de esas lenguas. No se diga de las amerindias, cuyo pasado, salvo el más cercano de algunas de ellas como el náhuatl, el otomí o hñahñú, el purépecha o tarasco, o el maya yucateco.

Cuenta como iniciador de la lingüística románica, dedicada precisamente al estudio comparativo de las lenguas del tronco latino, el alemán Friedrich Diez (1794–1876), contemporáneo de Goethe, de los hermanos Grimm o de Augusto Schlegel, iniciadores, a su vez, de la lingüística germánica.<sup>1</sup> Los métodos utilizados por Diez fueron los mismos que se habían comenzado a aplicar desde que William Jones en 1786 afirmó el parentesco entre el sánscrito, el griego y el latín, basado en la comparación de *formas de la expresión*<sup>2</sup> en este caso, *palabras*, a partir de la similitud de las formas y la correspondencia entre sus contenidos o significados. Así por ejemplo *padre* en español e italiano, *père* en francés remiten a *pater* en latín; *vater* en alemán, *father* en inglés remiten a un gótico *fadar*; en griego *pater*, en sánscrito *pitá*. Según el romanista suizo recién fallecido Peter Wunderli (2001), el comparativismo de Diez se distingue del de los germanistas en cuanto que no era reconstructivista —es decir, no pretendía reconstruir hipotéticamente vocablos de etapas anteriores de evolución o

---

<sup>1</sup> En 1818 publicó romances españoles antiguos y estudios sobre trovadores provenzales. Entre 1830 y 42 publicó su *Grammatik der romanischen Sprachen* (3 vols.) y entre 1853 y 54 su *Etymologisches Wörterbuch der romanischen Sprachen*.

<sup>2</sup> Utilizo la terminología de Louis Hjelmslev (1943).

incluso de lenguas ancestrales—, sino evolucionista y clasificadorio, puesto que disponía de datos de la lengua madre, el latín, y buscaba explicar los procesos evolutivos sufridos por cada lengua romance a partir de aquélla, así como considerando la influencia de otras lenguas sobre ellas.<sup>3</sup> El comparatismo de los germanistas o los eslavistas y los indoeuropeístas, por el contrario, solo podía ser reconstructivo, dedicado a extrapolar hipotéticas formas originarias a partir de esas lenguas con una historia documental reducida. No me parece solo anecdótico que el comparatismo reconstructivista sea el método de los estudiosos de las lenguas amerindias, precisamente porque tampoco de ellas hay documentos antiguos y también por el hecho de que los lingüistas americanos tienen su origen en la lingüística germánica y en la necesaria lingüística descriptiva, taxonómica, no precisamente histórica.

Fue en el ámbito de los germanistas en donde se desarrolló el interés por encontrar una lengua madre, de la cual derivarían las romance, las germánicas, el sánscrito, etc. August Schleicher (1821–1868),<sup>4</sup> en quien se reconoce una fuerte influencia de Darwin, fue el primero en proponer que habría habido una lengua de la que derivarían aquellas: el indoeuropeo o indogermánico (ver Figura 2). Ya Diez señalaba como mal método la mera comparación entre palabras y proponía como más segura la comparación entre fonemas (aunque

---

<sup>3</sup> Un resumen sustancioso de la historia del comparativismo en las lenguas romance se encuentra en el artículo “Linguistique comparée” de Joseph Herman (1989: 704-718).

<sup>4</sup> En Schmidt (1873).

todavía en su época la noción de *fonema* era confusa<sup>5</sup>). Schleicher igualmente optó por llevar a cabo la comparación entre fonemas, para establecer ciertas reglas de su evolución, una idea posteriormente desarrollada por una pléyade de jóvenes indoeuropeístas y germanistas, que adoptaron el nombre provocador de *Junggrammatiker* (“los jóvenes gramáticos”), a los que el lingüista Graziadio Isaia Ascoli optó por llamar *Neogramáticos*. Destacan en ese grupo Karl Brugmann, August Leskien y Hermann Osthoff.

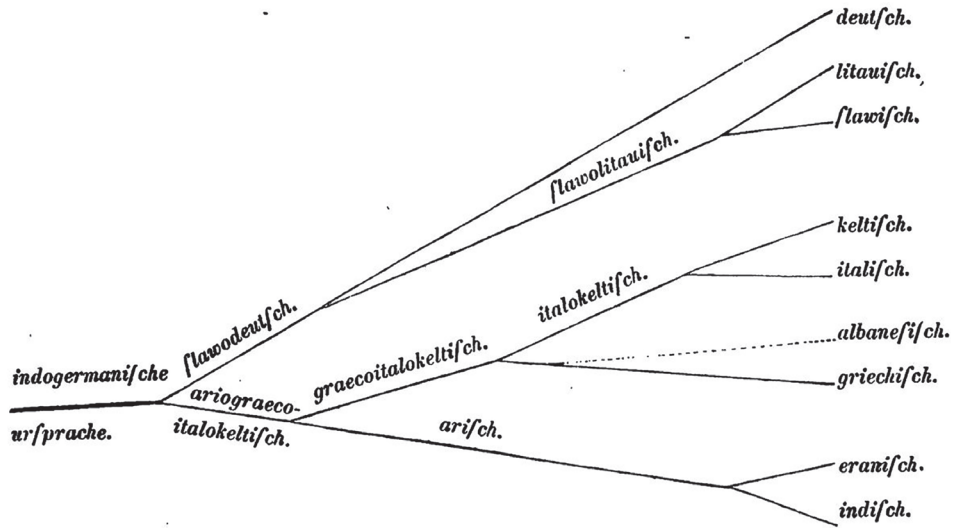


Figura 2. Árbol del indoeuropeo según Schleicher

<sup>5</sup> Al respecto, André Martinet (1975) ofrece una concisa explicación acerca de la historia de esta noción.

Los neogramáticos establecieron una serie de fenómenos fonético–fonológicos que se producían sistemáticamente en la reconstrucción evolutiva del indoeuropeo. Por supuesto, en un proceso de doble vía: comparaban las lenguas modernas de la hipotética familia indoeuropea para extrapolar un fonema o un conjunto estructurado de fonemas sobre la base de los datos bien documentados de esas lenguas y luego reconstruían el supuesto indoeuropeo. Tales investigaciones dieron lugar a las llamadas “leyes fonéticas” (*Lautgesetze*)<sup>6</sup> en el último cuarto del siglo XIX. Las leyes fonéticas establecían regularidades en la evolución de ciertos sonidos, regularidades que no están sujetas a la actividad consciente de los hablantes y no parecen influidas por la cultura. Una expresión inicial de una ley fonética es la llamada ley de Grimm, que establece que las consonantes oclusivas sordas del sánscrito y del griego (y, por lo tanto, atribuidas al indoeuropeo) evolucionaron en gótico (antecedente del alemán) a fricativas sordas cuando aparecen en posición inicial de palabra, por lo que /p/ > /f/, /t/ > /þ/ y /k/ > /χ/, por ejemplo el sánscrito *pitár* y el griego *πατήρ* demuestran la evolución al gótico en *fadar* (en inglés *father*, en alemán *vater*); en cambio en latín *pater* conservó las oclusivas iniciales (en español e italiano *padre*, en francés *père*). En posición intervocálica esas mismas consonantes evolucionaron a consonantes sonoras /b/, /d/, /g/, debido a la acción sobre la palabra del acento de intensidad en indoeuropeo: Así en sánscrito *bhrātar* > gótico

---

<sup>6</sup> Véase Brugmann, Karl & Berthold Delbrück, *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen*, Zweite Auflage, fünf Bände. 1897–1916.

*brōPar* (inglés moderno *brother*, en alemán *bruder* ‘hermano’); en cambio en latín *frater* conservó la sorda intervocálica /t/.<sup>7</sup> Esos procesos evolutivos regulares no se produjeron de la misma manera en las lenguas romance, como hemos visto en los ejemplos anteriores; es muy claro, por ejemplo, el caso del grupo consonántico -kt- del latín: en español se produjo un nuevo fonema /ch/: *noctem* dio en español *noche*, pero en francés *nuit* y en italiano *notte*; el latín *lacte* dio en español *leche*, *lait* en francés, *latte* en italiano. Quizá muchas evoluciones sean casos de mutación que se produce en la organización de los sistemas fonológicos de las lenguas, en los que sus características fonéticas alteran los subsistemas, como el caso probado por la ley de Grimm en las relaciones entre p/b, t/d y k/g, ligados también al acento de intensidad y a la carga funcional de las oposiciones entre fonemas,<sup>8</sup> pero también es necesario considerar los otros componentes de los sistemas lingüísticos: sus morfologías, sus procesos de formación de palabras, los de gramaticalización y los de lexicalización. Es decir, hay que partir de descripciones completas de las lenguas que se comparan y no solamente, ni de palabras aisladas, como criticaba Diez, ni de sistemas fonológicos exclusivamente, ni mucho menos de fonemas aislados. En otros casos esas evoluciones estaban (y siguen estando) condicionadas por diferentes factores debidos a la

---

<sup>7</sup> Más tarde italiano *fratello*, mientras en francés *frère* en donde la /t/ intervocálica desapareció; en español, *hermano* < *germanus* que en latín quería decir ‘hermano auténtico o carnal’, pero después se introdujeron, como cultismos, *fraterno*, *fraternal*, etc.

<sup>8</sup> Acerca de la noción de *carga funcional* véase mi artículo “Una condición de la evolución de las lenguas: la medida del rendimiento funcional” en prensa, El Colegio Nacional.



influencia de las lenguas anteriores a la llegada del latín, es decir, la influencia del sustrato, o al desarrollo de las culturas. Un caso ejemplar es la evolución de *f*- inicial latina, que dio lugar en castellano a una aspiración y luego a su desaparición: lat. *facere* > *hacer*, *filium* > *hijo*; *formosu* > *hermoso*; *forno* > *horno* por influencia del vasco, cuyo sistema fonológico no tiene /f/.

El papel de los neogramáticos en el desarrollo de la lingüística comparativa y evolutiva fue determinante, pues no solo introdujo la observación de la evolución fonológica y abrió el camino a la noción de sistematicidad, correspondiente al hecho de la sistematicidad de las lenguas (posteriormente desarrollada por Ferdinand de Saussure), sino también llevó a revisar la cronología de la evolución de las lenguas indoeuropeas. Por un lado, como señala Joseph Herman (1989), el estudio del sistema fonológico, al tratarse de un sistema de pocos elementos y claramente estructurado, facilita la elaboración de hipótesis acerca de estados de lengua más antiguos; por el otro, elude en principio la influencia de la cultura, la interferencia de otras lenguas y la actividad consciente de los hablantes, que, como se verá más adelante, constituye un fuerte obstáculo para la reconstrucción fidedigna de las protolenguas, las familias de lenguas y, si acaso existió, la lengua primigenia.<sup>9</sup> Los neogramáticos postulaban,

---

<sup>9</sup> Hay que aclarar lo que en este contexto entiendo por “lengua primigenia”: históricamente, se considera, sobre la base de los mitos de la creación que crearon el primer horizonte de estudio de las lenguas, que la lengua primigenia fue la supuesta lengua de Adán y Eva. Al respecto, véase Umberto Eco (1994). Pero en este contexto me refiero a la idea que se encuentra, por ejemplo, en la obra de Cavalli-Sforza (2001), de que, si hubo unos

primero, la regularidad de los cambios fonológicos; segundo, que si, entre elementos comparables (fonológica, morfológica y sintácticamente) se encuentran diferencias, son indicación de evoluciones divergentes a partir de una protolengua. Como el primer postulado se contradiría con el segundo, pues los cambios debieran ser regulares y sistemáticos en una lengua, lo que revela el segundo es la intervención de influencias externas sobre la evolución (algo semejante a lo que se observa en las divergencias genéticas entre poblaciones). Precisamente para contrarrestar el carácter absoluto de las leyes fonéticas apareció en la lingüística de los últimos años del XIX la llamada “teoría de las ondas” (*Wellentheorie*) –es interesante compararla con las hipótesis de Cavalli-Sforza (2001) acerca de la propagación de la agricultura– propuesta por Hugo Schuchardt y Johannes Schmidt,<sup>10</sup> que proponía que las leyes fonéticas se veían alteradas por la interferencia entre diversas variedades lingüísticas a partir de ciertos centros de difusión de los cambios. Igualmente con ese propósito surgió la necesidad de estudiar los dialectos en el campo, en sus regiones, en sus peculiaridades, sin someterlos a la lente deformadora de la gramática normativa, tema de la dialectología iniciada, entre otros, por Georg Wenker, seguidor, paradójicamente, de los postulados de los neogramáticos y muy desarrollada en el ámbito romance desde que

---

*homin sapiens* originarios en África, de donde deriva todo el género humano, puede haber habido una primera lengua de esos hombres, una “lengua primigenia”.

<sup>10</sup> Primero enunciada por Hugo Schuchardt hacia 1870 (pero publicada hasta 1885) para oponerse a la teoría de los neogramáticos, y elaborada por Johannes Schmidt (1872).

Jean Pierre Rousselot y Jules Gilliéron fundaron la *Revue des patois gallo-romans* en 1887 (Malmberg 1966). La teoría de las ondas más que contradecir las leyes fonéticas contribuye a reconocer que muchos cambios lingüísticos se producen por contactos e interferencias entre poblaciones humanas; la dialectología demuestra, por su parte, la relatividad de esas leyes a causa de la constante variación que se produce en las comunidades lingüísticas, debido a sus necesidades de adaptación al ámbito en que viven y a las diferencias culturales correspondientes.

La existencia del indoeuropeo no deja de ser hipotética, puesto que no hay documentos que la prueben, así como también la de las diversas lenguas intermedias que se necesita postular para poder llegar en la cadena evolutiva a las lenguas contemporáneas. En el ámbito germánico y eslavo se ha tenido que proponer un “germánico común” y un “eslavo común” como lenguas procedentes del indoeuropeo, de donde habrán evolucionado sus correspondientes modernas. En el ámbito romance la situación es relativamente diferente, puesto que no solo hay una amplia documentación del llamado “latín clásico”, que encontramos, por ejemplo, en las obras de Virgilio, de Cicerón, de Séneca, de Marco Aurelio, sino que se encuentran documentos de las variedades populares latinas, el llamado “latín vulgar”, mediante los cuales se puede reconstruir en muchos casos cómo habrán sido esas variedades populares y, en particular, cómo intervinieron en ellas las lenguas de los pueblos, ya sea aborígenes de cierto territorio, como

el vasco o el ibérico en la Península Ibérica,<sup>11</sup> ya colonizadores, como los celtas, los visigodos o los árabes.<sup>12</sup>

La lingüística románica es la que nos ofrece suficientes datos y pruebas de la manera en que evolucionaron las lenguas romance. Incluso, debido a la procedencia latina de estas lenguas, contamos con suficientes datos de carácter cronológico, que nos permiten datar muchos cambios; es decir, hacer lingüística histórica real. A la vez, la lingüística románica ofrece un catálogo de métodos y observaciones críticas que no se pueden soslayar cuando se trata de reconstruir familias o troncos lingüísticos de otras regiones de la Tierra.

Tal es el caso de la filogenia de las lenguas amerindias. El iniciador más destacado de esta rama de la lingüística en el continente americano fue el lingüista de Yale Edward Sapir, alumno de Franz Boas y educado en la escuela neogramática, quien propuso la existencia de un grupo humano procedente de Asia, que se avecindó en el extremo noroeste del continente, del cual habrán derivado, entre otras lenguas, el atabasca, el tlingit, el navajo y el apache. Sapir (1915) llamó *na-dene* a ese grupo<sup>13</sup> después de hacer estudios comparativos a partir de

---

<sup>11</sup> Cavalli-Sforza (2001) afirma que el vasco, que supone se hablaba en el extremo noreste de la Península hace 12 mil años, corresponde a un período preneolítico, quizá relacionado con los hombres de CroMagnon, por lo que puede haber habido en Europa varias lenguas pre-indoeuropeas en el paleolítico.

<sup>12</sup> Como se puede leer, en cuanto al español, en todas las historias de la lengua; véase Lara (2013).

<sup>13</sup> En su artículo explica el nombre: “The name that I have chosen for the stock, *Na-dene*, may be justified by reference to no. 51 of the comparative vocabulary. *Dene*, in various dialectic forms, is a wide-spread Athabaskan term for “person, people”; the element *\*-ne*

descripciones modernas de las lenguas atabasca, haida y tlingit, posiblemente elaboradas por él mismo. Sobre su obra inicial se han venido construyendo hipótesis acerca de los orígenes de las lenguas de nuestro continente; primero de las del norte y, a partir de los libros de Lyle Campbell y Marianne Methun, *The languages of native America* (1979) y de Joseph Greenberg, *Language in the Americas* (1986), las del sur y del centro.

Mientras que Sapir tomó en consideración tanto elementos fonológicos como morfológicos y parcialmente sintácticos para su propuesta de una lengua madre na-dené de aquellas lenguas modernas del noroeste de América, pero que se extienden hasta los apaches y los navajos del norte de México, a partir de Mauricio Swadesh<sup>14</sup> las investigaciones filogenéticas se han basado en varias listas de “reactivos léxicos”, la más socorrida de las cuales es la de cien

---

(\*-n, \*-ŋ) wich forms part of it is an old stem for “person, people” which, as suffix or prefix, is frequently used in Athabaskan in that sense. It is cognate with H[aida] *na* “to dwell; house” and T[inglit] *na* “people”. The compound term *Na-dene* thus designates by means of native stems the speakers of the three languages concerned, besides continuing the use of the old term *Dene* for the Athabaskan branch of the stock.

<sup>14</sup> Mauricio o Morris Swadesh, de padres nacidos en Besarabia, fue alumno de Sapir. Durante el gobierno de Lázaro Cárdenas contribuyó al despegue de las investigaciones modernas de las lenguas de México. Dirigió el Consejo de Lenguas Indígenas, el Departamento de Asuntos Indígenas y el Proyecto Tarasco en Michoacán en el año de 1939. Tras su participación en la guerra volvió a México a causa de la persecución macarthista; creó la formación en lingüística en el Instituto Politécnico Nacional y después en la Escuela Nacional de Antropología e Historia; en la UNAM fue el iniciador de las investigaciones filogenéticas de las lenguas con ayuda de la computación electrónica.

reactivos propuestos por él.<sup>15</sup> Digo “reactivos léxicos” porque supongo que primeramente los habrá elaborado Swadesh precisamente como instrumentos de interrogación a sus informantes cuando comienza el estudio descriptivo de una lengua. En realidad, todavía hoy muchos lingüistas descriptivistas siguen utilizándola para llevar a cabo sus investigaciones de campo, sin pretender reconstruir genealogías. Esos reactivos buscan *denominaciones* de objetos, de personas y de algunos procesos verbales que “deben” formar parte de cualquier lengua del mundo, es decir, no deben estar, por lo general, sujetos a cambio a lo largo del tiempo ni a influencias de otras lenguas.<sup>16</sup> No he logrado encontrar ningún documento que me indique en qué se basó Swadesh para elaborar su lista; es posible que proviniera de su propia experiencia, buscando que contuviera exclusivamente vocabulario que no fuera producto de la cultura o del contacto con otras

---

<sup>15</sup> Tal lista fue objeto de varias versiones, de una mayor, de 251 reactivos, a la última, de cien. Hoy se han construido varias listas más por diferentes investigadores, pero con el mismo propósito. Holman *et al.* (2008) investigaron la estabilidad relativa de las palabras en la lista de Swadesh de 100 vocablos, comparando sus porcentajes de retención en familias de lenguas bien establecidas, con lo que llegaron a proponer una lista de solo cuarenta reactivos.

<sup>16</sup> Lo que suscita el problema del *tertium comparationis*, como lo explicó Klaus Heger (1974). El problema consiste en el uso de los significados de los signos de una lengua real, como si fueran *conceptos* de carácter universal. Tal idea es tan antigua como los seres humanos pero supone el sobreseimiento de la complejidad y la densidad del significado que se construye en cada lengua. Véase al respecto Lara (2001). Cuando esto sucede, siempre hay el peligro de forzar los signos de las otras lenguas a parecerse —o, peor, a ser “equivalentes”— a los de la lengua utilizada como base de la comparación, lo que deforma sus significados.

lenguas, es decir, que no contuviera préstamos o que no se hubiera creado culturalmente, lo que lo volvería estable a lo largo del tiempo. Dicho de otra manera, que ese léxico correspondiera a unos “universales empíricos y sustanciales”.<sup>17</sup> A ese vocabulario lo llamó “vocabulario básico”. Hoy, quienes hemos construido subconjuntos del léxico de una lengua que llamamos “básicos” o “fundamentales”, lo hacemos a partir de recuentos estadísticos basados en grandes corpus de *texto* –no de palabras aisladas–, por lo que no hay sesgos en la selección, pero ciertamente tales vocabularios tampoco eluden el hecho de que hay préstamos de otras lenguas o que son manifestaciones evidentes de la cultura de cada comunidad lingüística.<sup>18</sup> La “lista de Swadesh” o las diversas variaciones que se han hecho de ella siguen siendo la piedra de toque para la reconstrucción comparativa de las lenguas de América y su clasificación. Un hecho, como se verá en adelante, sorprendente.

Swadesh llamó al método que preconizaba *lexicoestadística*. La aplicación de este método para establecer cronologías de los cambios

---

<sup>17</sup> Como se puede colegir a partir de la nota anterior, esos universales empíricos contradicen la teoría semántica basada en la teoría del signo de Ferdinand de Saussure, que reconoce la especificidad inmanente a cada lengua de sus signos. La disciplina lexicológica llamada *onomasiología* permite la construcción de universales formales solo como método de investigación. En el mismo caso se encuentra el esfuerzo chomskyano por construir una “gramática universal” *formal*, independientemente de su aspiración a que tenga un fundamento genético.

<sup>18</sup> Al respecto, véase las publicaciones acerca del “español mexicano fundamental” de Lara, García & Ham (1980) y Lara (1990), en donde se explican los fundamentos estadísticos de esos cálculos.

sufridos por las lenguas se llama *glotocronología*. Evidentemente, la inexistencia de datos sobre estados antiguos de las lenguas en general, pero sobre todo de las amerindias, o sobre sus lenguas ancestrales, no permite situarlos en la historia, en acontecimientos reales acaecidos en el tiempo, por lo que estrictamente hablando no se debiera hablar de *lingüística histórica*, sino de *diacronía lingüística* o *lingüística diacrónica*, en la cual la reconstrucción hipotética de los signos de las lenguas en cuestión solo permite proponer ancestros, pero no determinar su existencia histórica (si uno revisa los resultados que obtienen los genealogistas de las lenguas, verá que la mayor parte de los signos propuestos como orígenes de los registrados son hipotéticos, marcados con asterisco, siguiendo la convención que estableció Schleicher). Swadesh observaba que el cambio lingüístico opera lentamente y sigue ciertas tendencias evolutivas que producen divergencias entre las lenguas a partir de un tronco común, lo que le permitiría calcular el tiempo que han tardado esos cambios, e incluso llega a hablar de un “reloj lingüístico”, como en biología el “reloj molecular”, aunque no explica en qué consiste, lo que también señala Cavalli-Sforza (2001). Los criterios propuestos por Swadesh para elaborar la genealogía son cuatro: en primer lugar, parte de la idea de que, si entre dos o más lenguas hay similitudes, debe haber habido una época en que se trataba de una sola lengua. En efecto, la evolución del latín a las lenguas romance lo prueba, aunque sea imposible calcular contra reloj o contra calendario en qué momento o cuánto tardaron en producirse las divergencias. En cuanto al paso del latín a las lenguas romance todo hace pensar que su origen se debe situar en la variación



dialectal que se produjo muy pronto en los territorios dominados por Roma, de poblaciones romanizadas ibéricas, celtas, galas, griegas en el sur de Italia, y poco más tarde germánicas, una variación contemporánea con el latín clásico, como lo demuestran algunas citas del latín popular en el *Satiricón* de Petronio a finales del siglo I d.C.<sup>19</sup> Los primeros documentos del “francés” o del “castellano” –para aquella época no tiene sentido denominarlos así; lo hago para abreviar explicaciones– se sitúan hacia el siglo noveno, pero ese milenio transcurrido entre el latín y las lenguas romance no se puede extrapolar para las demás lenguas de la Tierra, pues las condiciones, esas sí, históricas, de la evolución de la cultura y del dominio romano es más que probable que no se hayan dado de manera semejante en otras regiones. En segundo lugar, pensaba Swadesh, puesto que hay grados de similitud, también habrá grados de contacto tras la separación y se puede hacer inferencias acerca de su tasa de divergencia. Esas similitudes se manifiestan mediante signos cognados entre sí, como *padre* y *père* o *vater* y *father*. Observa que esos vocablos cognados dan información acerca de rasgo culturales antiguos y de características del entorno físico cuando todavía no se separaban; cuando no hay cognados,

---

<sup>19</sup> La llamada “lingüística espacial” o también “neolingüística”, propuesta por el lingüista italiano Matteo Bartoli a mediados del siglo xx se proponía establecer no solo épocas de formación de las lenguas romance, sino también áreas en que se producían ciertos cambios. Su obra *Saggi di lingüistica spaziale* (1945) está hoy injustamente olvidada, pues podría contribuir a las discusiones contemporáneas acerca de los “espacios comunicativos” propuestos por el alemán Wulf Österreicher (2002) y las “zonas” como Mesoamérica, que proponen Lyle Campbell, Kaufman & Smith-Stark (1986).

pero se encuentran correspondencias en relación con el reactivo o el “concepto”, se tratará de palabras resultantes de contactos culturales prehistóricos. Cavalli-Sforza comenta (2001: 138) que la glotocronología no tiene el rigor de la genética y la suposición de que hay una tasa constante de cambio de las palabras no se sostiene, pues es pensable que cada palabra tenga su propia historia, como lo comprobamos con creces quienes nos dedicamos a la investigación lexicológica.<sup>20</sup>

La lista de Swadesh y sus sucesoras toman en consideración denominaciones de partes del cuerpo, de algunos fenómenos meteorológicos y accidentes geográficos, de algunos objetos de la naturaleza que son alimenticios, ciertas acciones que llevamos a cabo los seres humanos *porque somos humanos*, como caminar o correr, nombres de colores y partículas mostrativas. Harald Haarman (1990) sometió a escrutinio varias listas, cuyos resultados demuestran la imposibilidad de basarse en una lista de esa clase a propósito del indoeuropeo, no se diga del resto de las lenguas de la Tierra. Tomemos por ejemplo los nombres de los colores en la lista modificada por Holman et al. (2008): Los reactivos son *yellow* y *green*. El inglés y el alemán *yellow* y *gelb* provienen de un protogermánico *\*gelwaz* ‘radiante’. El español

---

<sup>20</sup> En la réplica de Kenneth M. Weiss y Ellen Woolford, antropólogos, al artículo de Greenberg et al. (1986: 491), observan: “An application of glottochronology to English and the English-based pidgin of Papua New Guinea, Tok pisin, gives a separation date of something like 2,000 B.P., but Tok pisin has not existed a tenth of that time”. Algo semejante se podría decir en cuanto a las lenguas criollas caribeñas de Haití o de Curazao, Bonaire y Aruba, que se formaron a partir del siglo xvii, sobre la base del francés y el portugués con lenguas de la costa occidental de África.

*amarillo* proviene del lat. *amarus* ‘amargo’, que dio *amarellus*; en cambio it. *giallo* y fr. *jaune* provienen del lat. *galbus* o *galbinus* ‘verde-amarillo’. En galés, que tenía *glas* ‘azul, verde, gris’ se incorporó más tarde *gwyrdd* del lat. *viridis*, para especificar lo verde en comparación con el matiz de azul o de gris de su vocablo *glas*. De *viridis* derivaron esp. e it. *verde*, fr. *vert*, mientras que ingl. *green* y al. *grün* derivan de un proto germánico \**groni*. Tanto *viridis* como *groni* significaban el color de la hierba, lo fresco, lo joven. Estos pocos ejemplos demuestran el carácter profundamente cultural de los nombres de los colores. El artículo de Haarmann es devastador.<sup>21</sup>

El lingüista Joseph Greenberg fue colega de Cavalli-Sforza en Stanford. Después de haber publicado en 1963 una clasificación y filogenia de las lenguas de África, en 1986 publicó sus hipótesis acerca de las lenguas del continente americano.<sup>22</sup>

Greenberg sostiene que debe haber habido al menos tres migraciones hacia el continente americano, todas cruzando el estrecho de Bering: la primera habría sido hace 30 mil años, a finales del pleistoceno, extendida por la mayor parte de América, lo que lo lleva

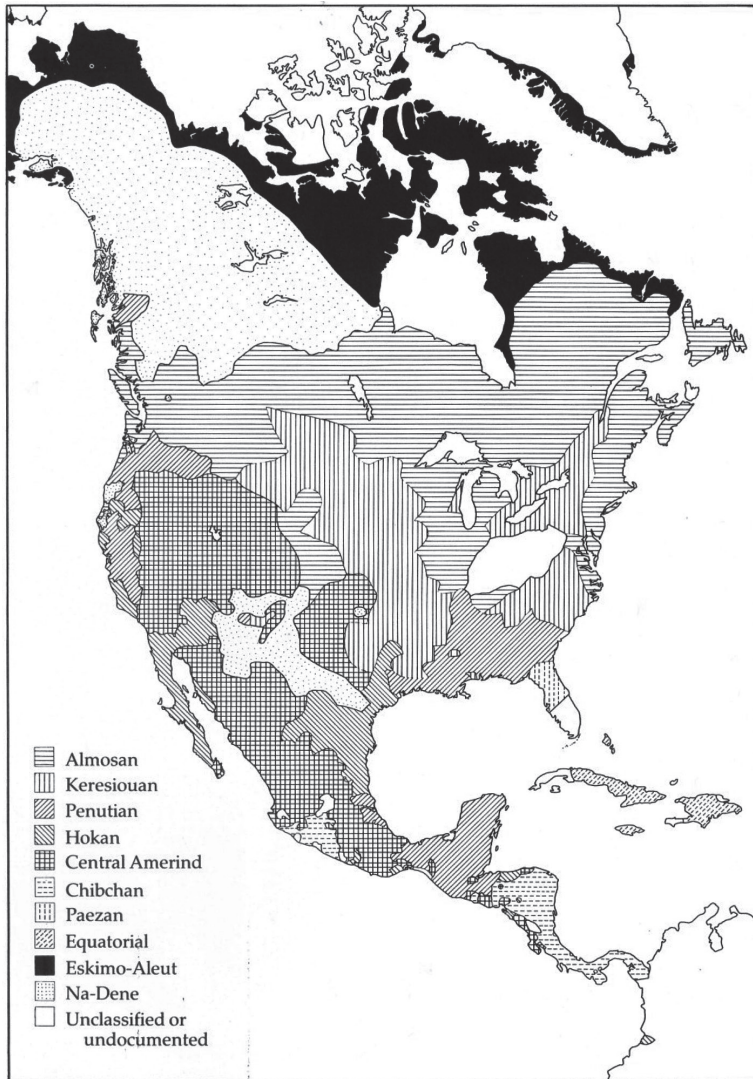
---

<sup>21</sup> Basado en la lista de Swadesh, Greenberg considera que los numerales en cada lengua son suficientemente estables como para tomarlos en cuenta como “universales”. Campbell, Kaufman & Smith-Stark (1986) hacen notar que una de las características propias del área cultural mesoamericana es la numeración vigesimal –como la céltica, por otra parte–, que la separa tanto de las lenguas al norte y el sur de la frontera cultural mesoamericana como de lenguas del mismo tronco, como las utoaztecas, por ejemplo el cora.

<sup>22</sup> Buenos recuentos de las investigaciones que se han hecho al respecto desde finales del siglo XIX hasta Greenberg y Campbell aparece en el libro de Campbell & Mithun (1979), así como en el artículo de Greenberg, Turner & Zegura (1986).

a proponer un gran tronco o macrofamilia amerindio –*amerind* en inglés, en vez del común *amerindian*, seguro para resaltar su valor terminológico y, en tal caso, en español se puede formar *amerindiano*, diferente de *amerindio*–, del cual formarían parte todas las de México. La segunda sería la na-dené, llegada hace unos siete mil años, y la tercera la esquimo-aleutiana, extendida a lo largo del círculo polar ártico hace entre tres a seis mil años. Las tres migraciones: amerindiana, na-dené y esquimo-aleutiana corresponden, según Greenberg, a los datos que ofrece la investigación arqueológica. En cuanto a sus reconstrucciones e hipótesis, dice haberse basado en la recopilación de una gran base de datos. El antropólogo Christy Turner, en el artículo que firma con Greenberg y Zegura, dedicado a comparar datos acerca de las dentaduras de los esqueletos hallados y de muestras de habitantes contemporáneos, sobre todo del norte del continente, afirma que todos los amerindios se parecen más entre sí que con los europeos, y que la variación dental que se encuentra entre ellos es mayor en el norte que en el sur. Sostiene, en consecuencia, que las tres migraciones postuladas por Greenberg se comprueban en su investigación y que los ancestros amerindios pueden haber provenido del norte de China, de pueblos relacionados con la cultura Clovis, una localidad en Nuevo México, de cazadores paleolíticos, hace cerca de 12 mil años (Figura 3).

¡Sin embargo, los sudamericanos tendrían ya 15 mil años! Es sorprendente que los dos colaboradores de Greenberg hayan llegado a la conclusión de que sus clasificaciones lingüísticas son más seguras



*The Amerind family, North and Central America*

**Figura 3. Mapa de las lenguas amerindias de Norteamérica según Greenberg**

que las que provienen de la antropología física y la genética<sup>23</sup> (quizá una explicable cortesía). La Figura 3 muestra la hipótesis general de Greenberg a propósito de las lenguas *amerindianas*. La mayor parte de las lenguas amerindias mexicanas queda incluida como *amerindiano central*, al cual correspondería la familia yutonáhuatl, pero también la otomangue; sorprende –como Greenberg mismo lo afirma– que clasifique al tarasco (purépecha) y al cuitlateco como pertenecientes a otra macrofamilia, principalmente sudamericana, el chibcha-paezano;<sup>24</sup> al hokano pertenecerían el seri, el subtiaba o tlapaneco y el tequistlateco o chontal de Oaxaca. Al penutiano de California Sapir lo había relacionado con el mixe-zoque y el huave; Greenberg agrega la familia maya y el totonaco-tepehuano. Si bien, como se ve en el mapa, la macrofamilia amerindiana central podría tener cierta continuidad, tiene uno que sorprenderse de que una lengua californiana tenga que ver con las mayas, distantes entre sí cerca de cuatro mil kilómetros y, agregaría, con enormes diferencias culturales.

La obra de Greenberg ha sido muy controvertida; Campbell & Mithun (1979), quienes más ha estudiado las lenguas amerindias, y en particular las del territorio mexicano con los mismos fines clasificatorios, distinguen doscientos grupos de lenguas amerindias. En su crítica

---

<sup>23</sup> Los participantes en la réplica al artículo citado no solo cuestionan las hipótesis de Greenberg, sino también las de Turner.

<sup>24</sup> Afirma que el reactivo *belly* ‘abdomen’ en purépecha da *-va-* ‘inside, dentro’. Otras formas en las otras lenguas de la familia son *bi* ‘dentro’, *pu* o *bu(wo)*. Para *burn* ‘quemar o quemado’ da en cuitlateco *čibe* y en tarasco *čuhpi*. Habría que conocer los razonamientos que lo llevan a estas atribuciones.

al artículo de Greenberg, Turner & Zegura (1984: 488) consideran que “neither their linguistic classification nor its dental/genetic correlation is supported, the conclusions about migrations are unwarranted, and the whole speculative adventure should be abandoned”. A pesar de ello Cavalli-Sforza se basa en ella y en la de Merrit Ruhlen (1994) –alumno del primero–, para ofrecer el dendrograma ilustrado en Figura 4, que combina los resultados de la investigación genética hasta 1988, extendida a todos los pueblos de la Tierra, con la hipótesis de Greenberg/Ruhlen; evidentemente, el interés de Cavalli-Sforza va dirigido a sostener la necesaria existencia de una lengua primigenia, una lengua correspondiente al primer *homo sapiens sapiens*, una lengua de “Eva”; las hipótesis de Greenberg no llegan a tal propuesta, aunque crea un superfilum “nostrático”<sup>25</sup> que relaciona los pueblos afroasiáticos, indoeuropeos, dravídicos, uralo-yukaghir, sinotibetanos, altaicos, amerindios, esquimo-aleutianos y chukchi-kamchatkos.

El artículo de Campbell, Kaufman & Smith-Stark (1986) es una buena demostración de la clase de método que hay que seguir para problematizar una filogenia como la postulada por Greenberg y Ruhlen. Sobre la base de un conocimiento suficiente –supongo– de las lenguas mesoamericanas, así como de las que se hallan al norte y al sur de ellas, postulan la existencia rasgos lingüísticos comunes a estas lenguas, no como fonemas o formas del contenido similares,

---

<sup>25</sup> Este nombre fue creación del lingüista danés Holger Pedersen y publicado en un artículo de 1903 dedicado al turco. Pedersen señala: “As a comprehensive designation for the families of languages which are related to Indo-European, we may employ the expression *Nostratian languages* (from Latin *nostrās* “our countryman”).



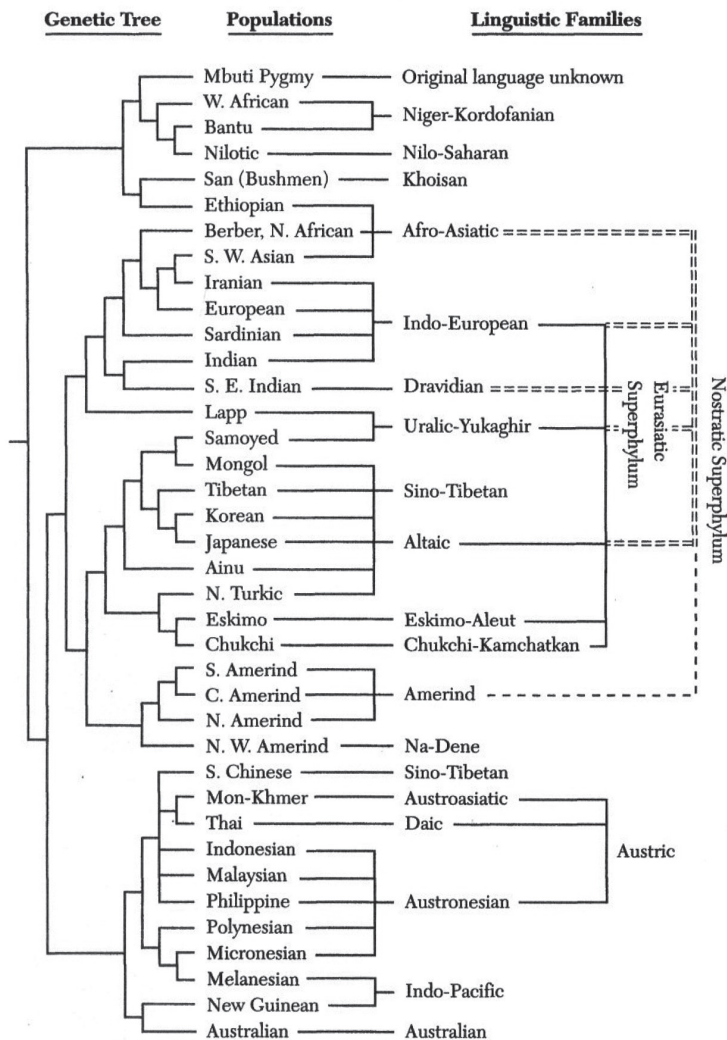


Figure 12. The comparison of genetic and linguistic trees (Cavalli-Sforza et al. 1988, pp. 6002-6).

Figura 4. Comparación de Cavalli-Sforza/Greenberg



es decir, no como “identidades”, sino como fenómenos morfológicos y sintácticos compartidos. Entre ellos la ausencia de marcas de plural en los nombres, los locativos derivados de los nombres de partes del cuerpo como en mixteco *čihi* ‘estómago’ > ‘abajo’, *ini* ‘corazón’ > ‘dentro’, *šini* ‘espalda’ > ‘detrás’;<sup>26</sup> la existencia de “nombres relacionales”, que consisten en la unión de un nombre con una partícula relacional (una preposición en español) como en chol *k-ik’ot* ‘conmigo’, (*k-* ‘mío’), *aw-ik’ot* ‘contigo’ (*aw-* ‘tuyo’), *y-ik’ot* ‘consigo’ (*y-* ‘suyo’);<sup>27</sup> los sistemas numéricos vigesimales; el orden de palabras en la oración, que nunca es sujeto-objeto-verbo; la existencia de morfemas absolutivos, como en náhuatl *tlaxacalli* ‘tortilla’ o *xitomatl* ‘jitomate’, etc. Concluyen estos autores en que el área lingüística mesoamericana se define fundamentalmente mediante: (a) la llamada “posesión nominal”, que consiste en la necesidad de que el sustantivo se componga de una raíz o de un lexema y un pronombre “posesivo”: por ejemplo en quiché *u-?’i:2 le: acih* ‘su perro-el hombre’, es decir, no se puede aislar ‘perro’ de la indicación de su poseedor. (b) la existencia, ya explicada, de “nombres relacionales”, (c) los sistemas numéricos vigesimales, y los rasgos negativos: (d) el orden de palabras en la oración que no puede terminar en verbo, y (e) la falta de partículas anafóricas para referir desde la oración subordinada al sujeto de la principal (switch-reference).

<sup>26</sup> Yo diría, en este caso, que tal fenómeno de significación es propio del esquema antropocéntrico humano, que no permite llegar a una forma primigenia.

<sup>27</sup> El mismo fenómeno se dio en latín *mecum*, *tecum*, *secum* que dieron en español *conmigo*, *contigo*, *consigo*.

Estos rasgos compartidos, que los lleva a sostener que Mesoamérica constituye un área lingüística, correspondiente a la Mesoamérica propuesta por la antropología, no llevan necesariamente a inducir una filogenia mesoamericana, puesto que se trata de *fenómenos de los sistemas lingüísticos*, de *tipología de las lenguas*, cuya relación debe ser materia de profundización teórica, y todavía es improbable atribuirles un origen filogenético; más bien parecen corresponder, ante todo, a las múltiples posibilidades de articulación de cada sistema lingüístico, ofrecidas por la capacidad o facultad humana del lenguaje, esa sí, única. Lo que también demuestra este estudio es la posibilidad de influencia de unas lenguas sobre otras y, consecuentemente, de adopción no solo de vocablos –los préstamos– sino sobre todo de procedimientos de significación; creo que la extensión del sistema numeral vigesimal entre todos los mesoamericanos lo demuestra.<sup>28</sup>

La posición de Campbell acerca de la filogenia de las lenguas amerindias es probablemente la que mejor se puede sostener. Por su parte ofrece la siguiente clasificación de las lenguas en territorio mexicano: otomangue, que comprende el llamado “mixtecano” (mixteco y cuicateco); “popolocano” (mazateco, popoloca, chocho, ixcateco, chiapaneco-mangue –chiapaneco ya desaparecido, mangue en Centroamérica–); “otopameano” (otomí, mazahua, matlaltzinca, ocuilteco);

---

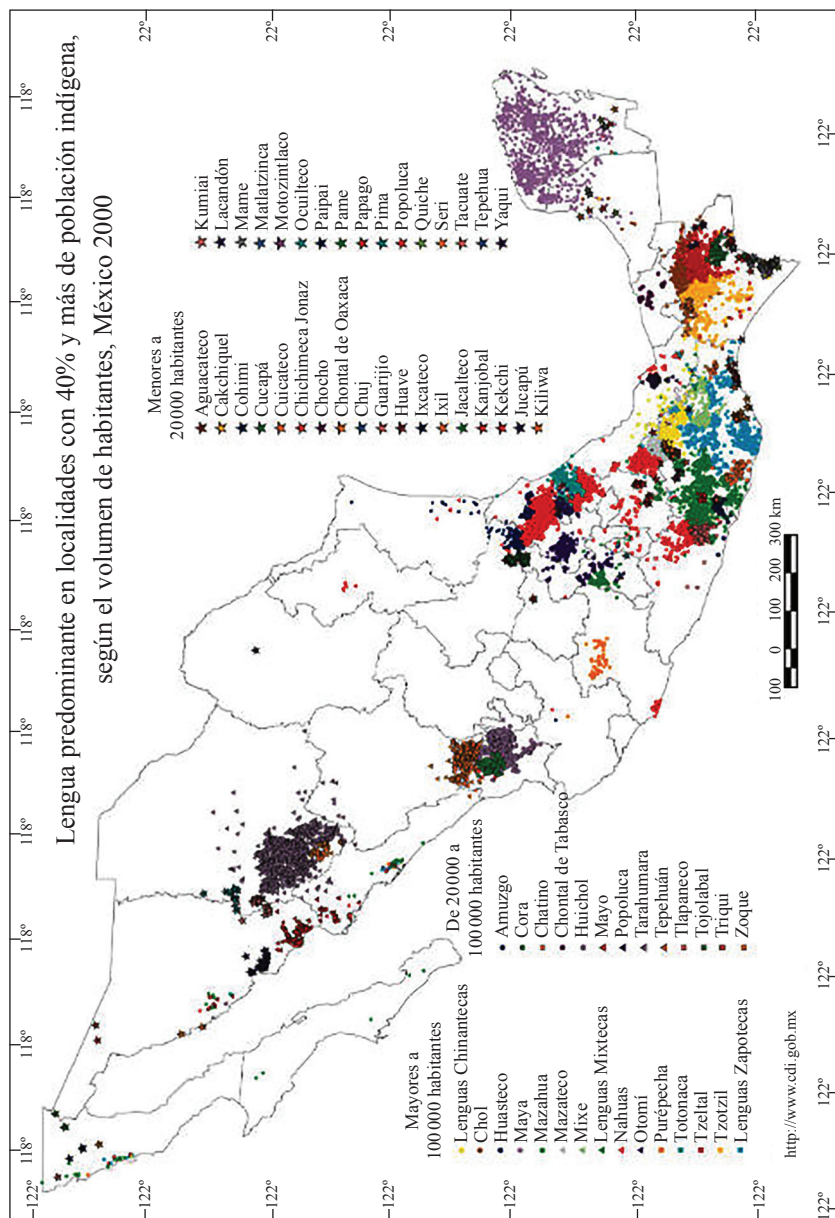
<sup>28</sup> Señalan Campbell & Mithun (1979: 488): “There is no deterministic connection between language and gene pools or culture. A single language can be spoken by a genetically and/or culturally diverse community; a culturally and /or genetically homogeneous population can speak more than one language. That is, language shift and multilingualism are facts of linguistics (and cultural) life; genes neither cause nor cater to them.”

“pameano” (pame norte y sur); chichimeco jonaz; “zapotecano” (zapoteco –de seis a 56 lenguas diferentes–, chatino); chinanteco; amuzgo; “hokano” (teuistlatecano o chontal de Oax); tlapaneco-subtiaba; maya; huave y “totonacano”.<sup>29</sup> Leopoldo Valiñas (2019), nos ofrece una clasificación de las lenguas amerindias mexicanas actuales en la edición especial No. 85 de *Arqueología mexicana*.

A la llegada de los conquistadores españoles el número de lenguas amerindias en territorio mexicano debe haber sido mayor que el actual. Lamentablemente, de la mayor parte de ellas solo se cuenta con sus nombres, sin que se pueda saber si tales nombres realmente designaban lenguas o designaban pueblos, que quizá compartían la lengua con otros más. Uno puede sospecharlo si revisa la *Geografía de las lenguas y carta etnográfica de México*, de Manuel Orozco y Berra, publicada en 1864. Es más que notable la discrepancia actual entre las 282 lenguas amerindias de México que lista el Instituto Lingüístico de Verano en su portal *Ethnologue* y las 68 que reconocen el Instituto Nacional de Estadística, Geografía e Informática y el Instituto Nacional de las Lenguas Indígenas (ver Figura 5). Lo que revela esa discrepancia es que no hay un criterio metodológico bien definido y compartido por todos para definir la relación entre dialectos y lenguas. Posibles causas para ello son, por un lado, los criterios lingüísticos sistemáticos seguidos por los investigadores; por ejemplo, sabemos que en español hay al menos dos sistemas fonológicos: uno castellano, que distingue *ese*

---

<sup>29</sup> Es claro que la manera de denominar las agrupaciones de lenguas por los lingüistas estadounidenses se debe a la necesidad de distinguir las agrupaciones propuestas respecto de los nombres de las lenguas comprendidas.



**Figura 5. Mapa de lenguas indígenas según el INALI**

y *zeta* y otro “atlántico” que comprende Andalucía, parte de Extremadura, las Islas Canarias y todo el continente hispanoamericano, que solo tiene *ese*. También cualquier persona nota muchas diferencias en la formación de palabras, en algunas construcciones sintácticas y, sobre todo, en vocabulario entre todos los dialectos actuales del español. Sin embargo, todos los hispanohablantes reconocemos una sola lengua. Lo mismo puede suceder en el campo amerindio, solo que allí las diferencias se conciben como *señales de identidad* de las diversas comunidades, de manera semejante a como los bordados de los huipiles distinguen unos pueblos de otros que, sin embargo, se entienden en “la misma lengua”. Es probable que *Ethnologue* reúna ambos aspectos: el de una sistematización aislante de una lengua y el de autorreconocimiento diferenciador de las comunidades para distinguir tantas lenguas diferentes... y colaborar a su fragmentación.

La disponibilidad actual de sistemas de cálculo y de modelos computarizados de simulación de procesos, por ejemplo en biología, ha dado lugar a varios intentos de elaboración filogenética de las lenguas. No he revisado los más citados pero creo que basta con considerar el artículo de Barbançon, Evans, Nakleh, Ringe & Warnow (2013) para poderse formar una idea acerca de ellos: se trata de un estudio que revela la idea subyacente de que no hay mucha diferencia entre el establecimiento de filogenias genéticas y lingüísticas, lo cual significa un gran desconocimiento de la naturaleza de las lenguas; además no pone en cuestión los conceptos de la lexicoestadística y la glotocronología de Swadesh y tampoco los datos lingüísticos que ofrecen los seguidores de esos conceptos y métodos, sino que se ocupa solo *instrumentalmente*

de los varios sistemas de cómputo que pueden utilizarse para establecer filogenias lingüísticas. Campbell & Mithun afirmaban en 1979 que la investigación actual acerca de la filogenia de las lenguas amerindias “has more often consisted of the perpetuation of the hypotheses of influential scholars without regard to the rigor of their methods or the weight of their evidence” (1979: 57). Yo agrego que, al parecer, pues no es esa mi especialidad, hace falta una profunda revisión de conceptos y métodos en lingüística que permita orientar la investigación filogenética por derroteros más dignos de confianza. A la vez, creo que los biólogos y computólogos que han venido publicando en revistas de ciencias naturales, como *Science* o *Nature*, contribuciones sobre filogenia de las lenguas inspirados por lo que ofrece la investigación genética de las poblaciones debieran acercarse con cuidado a las especulaciones lingüísticas que todavía circulan. La naturaleza de los fenómenos genéticos es muy diferente de la de los fenómenos lingüísticos; estos son siempre hechos de cultura. El indoeuropeísta inglés Benjamin W. Fortson IV (2010:49) afirma: “in spite of much wishful thinking along these lines, genes cannot be matched with languages”.

## REFERENCIAS

Barbançon, François; Evans, Steven N.; Nakhleh, Luay; Ringe, Don & Warnow, Tandy 2013. An experimental study comparing linguistic phylogenetic reconstruction methods. *Diachronica* 30(2). 143–170.

- Bartoli, Matteo. 1945. *Saggi di linguistica spaziale*. Torino: V. Bona.
- Campbell, Lyle & Mithun, Marianne. 1979. *The languages of native America: Historical and comparative assessment*. Austin: University of Texas Press.
- Campbell, Lyle; Kaufman, Terrence & Smith-Stark, Thomas. 1986. Mesoamerica as a linguistic area. *Language*. 62(3). 531–70.
- Cavalli-Sforza, Luigi Luca. 2001. *Genes, peoples and languages*. Los Angeles: University of California Press.
- Eberhard, David M.; Simmons, Gary F. & Fenning, Charles D. (eds.). 2019. *Ethnologue. Languages of the World*, 22a. edición, Texas: Summer Institute of Linguistics. <https://www.ethnologue.com/country/MX>.
- Eco, Umberto. 1994. *La búsqueda de la lengua perfecta*. Barcelona: Crítica.
- Fortson, Benjamin W. 2010. *Indo-European language and culture: An introduction*. Oxford: Wiley-Blackwell.
- Greenberg, Joseph H. 1987. *Language in the Americas*. California: Stanford University Press.
- Greenberg, Joseph H.M.; Turner II, Christy G.; Zegura, Stephen L.; Campbell, Lyle; Fox, James A.; Laughlin, W. S.; Szathmary, Emöke J. E.; Weiss, Kenneth M. & Woolford, Ellen. 1986. The settlement of the Americas: A comparison of the linguistic, dental, and genetic evidence [and comments and reply]. *Current anthropology*. 27(5).

- Haarmann, Harald. 1990. "Basic" vocabulary and language contacts: The disillusion of glottochronology. *Indogermanische Forschungen* 95. 1–37. <https://doi.org/10.1515/9783110243369.1>
- Heger, Klaus. 1974. Las bases metodológicas de la onomasiología y de la clasificación por conceptos. En *Teoría semántica II*. 1–32. Madrid: Alcalá.
- Herman, Joseph. 1989. Linguistique comparée, *Lexikon der romanistischen Linguistik* I(2). 704–718.
- Hjelmslev, Louis. 1943. *Omkring Sprogteoriens Grundlaeggelse*. Kobenhavn: Akademisk Forlag.
- Holman, Eric W.; Wichmann, Søren; Brown, Cecil H.; Velupillai, Viveka; Müller, André & Bakker, Dik. 2008. Explorations in automated language classification. *Folia Linguistica* 42(2). 331–354.
- Lara, Luis Fernando, García Hidalgo, Isabel & Ham, Roberto. 1980. *Investigaciones lingüísticas en lexicografía*. México: El Colegio de México.
- Lara, Luis Fernando. 1990. *Dimensiones de la lexicografía*. México: El Colegio de México.
- Lara, Luis Fernando. 2001. *Ensayos de teoría semántica: lengua natural y lenguajes científicos*. México: El Colegio de México.
- Lara, Luis Fernando. 2013. *Historia mínima de la lengua española*. México: El Colegio de México/ El Colegio Nacional.
- Malmberg, Bertil. 1966. *Les nouvelles tendances de la linguistique*. París: Presses Universitaires de France.
- Martinet, André (dir.). 1975. *La lingüística: guía alfabética*. Anagrama. Madrid.



- Orozco y Berra, Manuel. 1864. *Geografía de las lenguas y carta etnográfica de México*. México: J.M.Andrade y F. Escalante.
- Österreicher, Wulf. 2002. El español, lengua pluricéntrica: perspectivas y límites de una autoafirmación lingüística nacional en Hispanoamérica. El caso mexicano,. *Lexis* 26(2). 275–304.
- Ruhlen, Merrit. 1994. *The origin of language: Tracing the evolution of the mother tongue*. Nueva York: J. Wiley.
- Sapir, Edward. 1915. The Na-Dene languages: A preliminary report. *American Anthropologist* 17(3). 534–558.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau.
- Schuchardt, Hugo. 1885, *Über die Lautgesetze. Gegen die Junggrammatiker*. Berlín: Oppenheim.
- Swadesh, Morris. 1967. Lexicostatistic classification en Swadesh, M., McQuown. *Handbook of middle American Indians, Vol. 5 Linguistics*, 79–115. Austin: University of Texas.
- Valiñas, Leopoldo. 2019. Lenguas aisladas, *Arqueología mexicana* 85. 18
- Wunderli, Peter. 2001. Die Romanische Philologie von Diez bis zu den Junggrammatikern en Holtus, Günter et al. (eds). *Lexikon der romanistischen Linguistik*, 121–175. Tübingen.