Revista de Investigación Lingüistica, n.º 21 (2018)
ISSN: 1139-1146
Universidad de Murcia

De la glotocronología a la filogenética: estado de la cuestiónylos nuevos desarrollos en la metodología de clasificación lingüística

Javier Muñoz
Universidad de Valladolid
javi@fyl.uva.es

Resumen: En el presente artículo tratamos de mostrar una panorámica en la evolución de los estudios y la datación de lenguas antiguas y protolenguas. La glotocronología iniciada por M. Swadesh, pese a las polémicas y críticas suscitadas, ha evolucionado enormememente gracias a la incorporación de elementos estadísticos y etimológicos. Las nuevas metodologías utilizan grandes bases de datos y métodos relacionados con la filogenética. Proyectos como *Ielex*, ASJP o GDL muestran la importancia de la léxico-estadística actualmente y constituyen un elemento de gran valor para los investigadores así como un campo con enormes posibilidades dentro del aula.
Palabras clave: Glotocronología, Filogenética, Léxico-estadística, Historia de la Lengua.

Title: From Glotocronology to Phylogenetics: The evolution in linguistics classifications methodologies.
Abstract: In this work we try to make a panoramic view of the different methodologies applied to date ancient and proto-languages. The inicial effort of M. Swadesh, highly criticised, had evolved to new methods linked with database and phylogenetics. Our goal is to show how this new methodology can help the scholar and how to apply this in a linguistics or a history of the language clashroom. Projects like *Ielex*, ASJP or GDL show the importance of l exicostatistics and constitute a field with a great potential in teaching and research.
Keywords: Glottochronology, Phylogenetic, History of the Language, Lexicostatistics

1. INTRODUCCIÓN: LOS COMIENZOS DE LA GLOTOCRONOLOGÍA

El método comparativo ha sido la forma tradicional de aproximación para el análisis de lenguas arcaicas así como la vía para intentar reconstruir lenguas hoy perdidas. La base del procedimiento parte del análisis de cognados entre varias lenguas con el fin de delimitar un posible origen común e intentar una reconstrucción. Es el proceso seguido por los lingüistas desde el siglo XIX para intentar reconstruir el indoeuropeo o el protegermánico, por mencionar un par de ejemplos y esa metodología se ha mantenido estable en el ámbito de la lingüística diacrónica y ha sido considerada como válida sin excesivos cambios desde entonces.

En la década de los años 50 Morris Swadesh tratará de dar un paso adelante con el método comparativo e intentará partir de él para realizar una datación de las lenguas. Fuertemente influido por su maestro Edward Sapir y sobre todo por su teoría de la pervivencia de estructuras morfológicas básicas de lenguas emparentadas (1921), se embarcó en la tarea de buscar los términos básicos, resistentes a préstamos y que constituirán la base para sus análisis. Inicialmente constituirá un listado de 200 términos para reducirlo posteriormente a 100[1].

En él se incluyen elementos pertenecientes a partes del cuerpo, verbos, fenómenos naturales, etc. Los conceptos que podrían ser considerados como básicos en una lengua.

Este listado permite, siempre según Swadesh, establecer mediante la comparación el nivel de relación que existe entre dos lenguas. Para ello estableció una constante glotocronológica a partir del análisis de diversas lenguas cuya matriz era conocida. De este modo partió del hecho de que el cambio lingüístico es uniforme, algo que por ejemplo sostiene posteriormente Piotrowski con su uso de la función logística para delimitarlo (Marchuk 2003).

El ritmo constante del cambio es fijado por Swadesh en un promedio del 14% cada 1000 años, estableciendo una constante glotocronológica de retención de vo-

[1] El listado de Swadesh incluye los siguientes conceptos: *I (me), you, we, this, that, who, what, not, all, many, one, two, big, long, small, woman, man, person, fish, bird, dog, louse, tree, seed, leaf, root, bark, skin, flesh, blood, bone, grease, egg, horn, tail, feather, hair, head, ear, eye, nose, mouth, tongue, tooth, claw, foot, knee, hand, belly, neck, breasts, heart, liver, to eat, to drink, to bite, to see, to hear, to know, to sleep, to die, to kill, to swim, to fly, to walk, to lie, to come, to sit, to stand, to say, sun, moon, star, water, rain, stone, sand, earth, cloud, smoke, fire, ash, burn, path, mountain, red, green, yellow, white, black, night, hot, cold, full, new, good, round, dry, name, to give.*

cabulario del 0,86 (86%).

A partir de aquí la propuesta de Swadesh pasa por intentar delimitar la edad de una lengua a partir de la comparación de su lista de cognados.

Robert B. Lees propuso, en esta misma línea una formula glotocronológica que permitiría hallar la distancia temporal entre dos lenguas a partir de la división entre el tanto por ciento de cognados similares y la constante glotocronológica. Del siguiente modo:

$$t = \frac{lnc}{lnr}$$

Donde *t* sería la distancia temporal, *ln* el logaritmo natural (con base e); *c* el porcentaje de cognados comunes y *r* la constante glotocronológica reseñada más arriba.

Dicha fórmula podemos reproducirla fácilmente con una hoja de cálculo y valga como ejemplo la realizada con los alumnos de la asignatura de Historia de la Lengua del Grado de Lenguas Modernas y sus Literaturas: http://bit.ly/2jAFV1i

Podríamos, por tanto, siguiendo a Christopher Ehret delimitar la separación temporal a partir de la tasa de permanencia entre las dos lenguas (2000: 373-399):

Separación temporal en años	Porcentaje de similitudes
1000	74
2000	55
3000	40
4000	30
5000	22
6000	16
7000	12
8000	9
9000	7
10000	5

En resumen, los principios fundamentales de la glotocronología descrita por Swadesh pueden resumirse en varios principios básicos:

1. En el léxico de cualquier lengua puede localizarse un vocabulario que puede ser denominado como básico o estable. Un léxico que está menos sujeto a cambio que otros. Este vocabulario incluye términos como partes del cuerpo,

fenómenos meteorológicos, etc.

2. La tasa de retención de este vocabulario es constante a través del tiempo. Con ello significa que si en una lengua hay un cierto número de palabras del vocabulario básico, un porcentaje de este vocabulario permanecerá en la lengua a lo largo del tiempo.

3. Partiendo del porcentaje de cognados compartidos entre cualquier par de lenguas, podría computarse el tiempo transcurrido desde que ambas lenguas comenzaron a separarse

Sin embargo las críticas a Swadesh a partir de la exposición de su metodología fueron múltiples y en algún caso bastante justificadas. Baste, por ejemplo, mencionar la polémica iniciada por Eugenio Coseriu a partir del análisis glotocronológico de las lenguas romances (Coseriu 1965: 90ss).

Algunas de las dificultades del método radican en la propia lista de cognados, ya que en algunos casos podemos estar hablando de un préstamo posterior y no de un cognado real[2].

También, tal y como señalaba Coseriu, no existe un léxico fundamental universal (Coseriu 1962: 92), dado que su lista está pensada para el inglés[3].

Los cambios fónicos a lo largo de la historia podrían evitar un reconocimiento entre dos cognados y el desconocimiento de la evolución de las lenguas analizadas podría conducir a no reconocer la relación.

También la constante de retención o glotocronológica ha sido objeto de severas críticas, ya que podría variar con el tiempo y en función de la lengua o el significado del cognado.

Limitaciones que el propio Swadesh ya advirtió dado que los resultados nunca deberían ser considerados como dataciones absolutas de acontecimientos prehistóricos. Y de hecho, las dataciones glotocronológicas han de ser confrontadas con estudios históricos o arqueológicos, siempre y cuando sea posible. (Swadesh 1958: 551-559)

Las dificultades de aplicación son múltiples y el propio Swadesh intentará delimitar su metodología en trabajos posteriores (Swadesh 1971), o incluso compilan-

[2] Recordemos que dos palabras son cognados cuando se han desarrollado a partir de una misma proforma. Por ejemplo, Ziegel en alemán y teja en castellano, no podrían ser considerados como cognados, dado que Ziegel procede del préstamo latino tegula; a pesar de que en español la palabra teja tiene la misma procedencia que Ziegel, pero en el caso alemán se trata de un préstamo del latín al antiguo alto alemán.

[3] Coseriu señala el ejemplo de lie y de stand, que no tienen un significado simple en castellano o francés. Se refieren a las acepciones de esse o stare latinas, usadas en compuestos en lenguas románicas.

do varias versiones de su lista de conceptos.

Para percatarnos de las carencias del método glotocronológico podemos realizar un sencillo experimento, tal y como señala G. Jäger (2014: 4), partiendo de los datos de las lenguas indoeuropeas si comparamos los cognados del español y el hindi obtenemos un 22,5% de cognados comunes, mientras que si comparamos el español y el pastún obtenemos solamente un 14% de índice de cognados. Es evidente que el pastún y el hindi pertenecen, en ambos casos, a la rama indo-irania, y el español a la itálica, de modo que la distancia temporal cabría pensar que debería haber sido semejante.

2. LA EVOLUCIÓN DEL MÉTODO GLOTOCRONOLÓGICO: EL MÉTODO STAROSTIN

Las dificultades del método glotocronológico, fácilmente constatables, provocaron que múltiples lingüistas criticasen el procedimiento y que algunos constatasen la escasa validez del mismo. Es el caso de J. Tischler (1973) que intentó determinar la edad del indoeuropeo utilizando las listas de 100 y 200 términos. Sus conclusiones con las dataciones no dejan lugar a dudas:

> [Die Glottochronologie] ist sicher falsch und kann höchstens die durch andere Methoden wie Archäologie, vergleichende Ethnographie und linguistische Paläontologie erzielten Ergebnisse ergänzen (Tischler 1973: 139).

Este abandono y descrédito han sido la tónica habitual hasta fechas relativamente recientes.

Colin Renfrew en 1987 inicia una cambio que abogó por un renacimiento de los métodos léxico-estadísticos y alentó a una nueva generación de lingüistas que utilizaran los recursos técnicos a su alcance para dar un nuevo impulso a la metodología iniciada por Swadesh.

En ese sentido, el método fue abandonado prematuramente y quizá de un modo incorrecto (S. Starostin 1989, 3). Cualquier lingüista que ha trabajado con el procedimiento glotocronológico comprueba fácilmente que dialectos estrechamente relacionados tienen una tasa de cognados cercana al 90%, con la lista de 100 términos de Swadesh; lenguas estrechamente relacionadas como las de los grupos románico o germánico (con lenguas que comenzaron a divergir hace entre un milenio y dos) comparten un 70% o incluso un 80% de cognados; y familias lingüísticas como la indoeuropea que se separaron hace cinco o seis mil años tienen una ratio del 25% al 30%. Obviamente no hay una precision exacta, pero determinados elemen-

tos del método resultan indudablemente útiles.

Isidore Dyen, Joseph Kruskal y Paul Black (Dyen 1992) recogieron la iniciativa e intentaron validar el método léxico-estadístico con la familia indoeuropea (Dyen 1992, 2), así como separarse de la metodología usada por Tischler. La base de su estudio será un corpus de la lista Swadesh realizado con 31 lenguas y sus conclusiones serán tajantes: "The resulting lexicostatistical classification of Indoeuropean languages approximates the generally accepted classification". (Dyen 1992: 77)

En la misma línea, el lingüista ruso Sergei Starostin (1999) propuso una serie de modificaciones en el cálculo de la datación temporal atendiendo a la consideración de los préstamos léxicos, puesto que su inclusión en los porcentajes de cognados alteraba notoriamente los resultados. Su formulación es la siguiente:

$$t = \sqrt{\frac{lnN(t)}{-\lambda N_0}}$$

Su fórmula introduce un cambio en la constante de Swadesh, ya que para Starostin ésta ha de ser calculada individualmente. ln refleja la desaceleración en el proceso de reemplazo, de modo que los términos susceptibles de cambiar son los primeros en ser sustituidos y por tanto van permaneciendo los cognados más estables. Por el contrario, la raíz cuadrada representara la tendencia inversa[4].

Starostin trató de calcular el nivel de estabilidad de los listados de Swadesh partiendo del criterio general del número de raíces existentes en una familia lingüística para denotar el concepto. El resultado fue la denominada lista Jaxontov, una recopilación con los 35 conceptos más estables. La cuestión es que las lenguas emparentadas deberían mostrar un mayor porcentaje de coincidencias en esta lista y por el contrario, los otros 65 conceptos mostrarían el contacto lingüístico más que el parentesco. (Starostin S., 2007 [5])

La idea de que existe una constante de retención de vocabulario uniforme es rechazada y se adopta la hipótesis de que existe una correlación entre la constante y el tiempo de separación. De hecho se puede asumir que cuanto mayor es el valor de t existe mayor probabilidad de que una palabra del vocabulario básico desaparezca. (Starostin S. 1989: 10)

Además, ahora se incorpora λ que es un coeficiente de aceleración diferente

[4] Un interesante análisis siguiendo la metodología de Starostin para la divergencia entre el alemán y el francés lo realizan Petra Novotná y Václav Blazek en "Glottochronology and its application to the balto-slavic languages", Baltistica XLII (2) 2007, pág. 193.

[5] Citado en inglés por G. Starostin.

[6] El software puede descargarse de la página central del proyecto, así como las diversas bases de datos realizadas hasta el momento (http://starling.rinet.ru/new100/downloads.htm [21-06-2016])

para cada lengua, pero al margen de esta consideración, el que hemos denominado "método Starostin" presta atención al elemento etimológico: "the combination of lexicostatistics and etymostatistics allow us to obtain more precise datings and classifications, both for normal families and macro-families". (Starostin 1989: 27)

Está será la base de el proyecto *Tower of Babel de The Global Lexicostatistical Database* (GDL) que trataremos más adelante. La aplicación de los algoritmos es más compleja que la de Swadesh, pero es factible realizarla a través del software del GDL donde están implementadas diversas funciones, desde la tradicional glotocronológica a las más experimentales prouestas por S. Starostin[6].

Al mismo tiempo existe una evolución muy interesante en la datación de protolenguas, no sólo a partir de la evolución de las fórmulas matemáticas que intentan explicar el cambio lingüístico, sino a partir del tratamiento electrónico de las bases de datos y la aplicación de complejos algoritmos matemáticos.

3. LA LÉXICO-ESTADÍSTICA: EL INICIO DE LAS BASES DE DATOS ELECTRÓNICAS

La primera gran recopilación de cognados a partir de las listas de Swadesh será la realizada por el lingüista norteamericano Isidore Dyen, mencionado anteriormente. Se trata de la Comparative Indoeuropean Database. La base de datos inicialmente contenía 200 conceptos de la lista de Swadesh en 95 lenguas diferentes que en la actualidad han llegado ya a 163 lenguas y dialectos.

Esta base de datos es el punto de partida del grupo *Evolutionary Processes in Language and Culture* que ha puesto en marcha el proyecto *Ielex (Indo-European Lexical Cognacy Database).*

El tratamiento de los cognados es semejante al análisis de una mutación biológica, y de este modo los cognados pueden ser tratados como caracteres biológicos.

En ese sentido, este grupo asentado en el instituto Max Planck de Nimega y bajo la dirección de Michael Dunn ha aportado un importante elemento para el análisis de cognados, puesto que ha convertido los listados en lenguaje binario, en un lenguaje apto para ser analizado e interpretado por herramientas informáticas. Y dado que estamos hablando de datos semejantes a los biológicos, el uso de herramientas de análisis bioinformático será especialmente relevante.

Para realizar el paso a lenguaje binario se parte de dos posibles estadios: 0/1. Siendo 0 la ausencia de cognado de la lengua analizada respecto a la protolengua, y 1 la presencia de un cognado vinculado a la lengua matriz.

El profesor Dunn puso en marcha un programa informático[7] con el fin de clasificar los distintos cognados. Las secuencias numéricas obtenidas reflejarían la presencia o ausencia de una coincidencia en cada uno de los cognados del listado.

Una base de datos similar en su concepción será la realizada en el marco del proyecto ASJP (Automated Similarity Judgment Program) bajo la dirección del profesor Wichmann del instituto Max Planck de Leipzig. La base de datos creada consta de 40 términos de la lista Swadesh, seleccionados como los más estables y menos propicios a cambios. El número de lenguas se ha incrementado enormemente con respecto al *Ielex*, ya que la base de datos de ASJP consta actualmente de 6895 lenguas de todos los continents, inluyendo lenguas, dialectos, lenguas muertas e incluso lenguajes artificiales.

El listado de ASJP incorpora los siguientes conceptos:

Körperteile (partes del cuerpo)	*Tiere und Pflanzen (animales y plantas)*	*Menschen (el género humano)*
Auge *Ohr* *Nase Zunge* *Zahn* *Hand* *Knie* *Blut* *Knochen* *Brust (der Frau)* *Leber* *Haut*	*Laus* *Hund* *Fisch* *Horn (von Tieren)* *Baum* *Blatt*	*Mensch* *Name*
Natur (naturaleza)	*Verben und Adjektive (Verbos y adjetivos)*	*Ordnungszahlen und Pronomen (ordinales y pronombres)*
Sonne *Stern* *Wasser* *Feuer* *Stein* *Pfad* *Berg* *Nacht*	*Trinken* *Sterben* *Sehen* *Hören* *Kommen* *Neu* *Voll*	*Eins* *Zwei* *Ich* *Du*

[7] El software está disponible en https://bitbucket.org/evoling/lexdb

Pero con respecto a la binarización propuesta por el profesor Dunn, el ASJP incorpora un nuevo concepto: la Distancia Levenshtein (DL). La denominación se debe a Vladimir Levenshtein, matemático ruso que en la década de los años 60 del siglo XX trabajó en torno a la medición de distancias entre dos cadenas de caracteres.

La DL entre dos cadenas sería la cifra mínima de operaciones de edición que existe entre los dos elementos, teniendo en cuenta que es posible la eliminación, la inserción o la sustitución. De este modo, y por mostrar un ejemplo tomemos la palabra agua en antiguo alto alemán (*wazzar*) y gótico (*wato*), la distancia estaría definida por la sustitución de t por z, la eliminación de z, la sustitución de o por a y la eliminación de r: resultando en una DL de 4.

A partir de aquí y según el procedimiento descrito por G. Jäger (2014: 9) ha de obtenerse la distancia Levenshtein normalizada, dividiendo la distancia obtenida por la longitud total de la cadena. De este modo y analizando pares de lenguas se van obteniendo las correspondientes distancias promediadas que permiten una alineación inicial de las lenguas más cercanas y también de las más distantes[8].

Otra de las iniciativas más interesantes actuales es la del proyecto GDL liderado por G. Starostin, antes mencionado, que continúa la labor de su padre S. Starostin en la compilación de listas de palabras en múltiples lenguas y dialectos. El proyecto está enmarcado en el programa *Tower of Babel / Evolution of Human Language* y su objetivo es agrupar y permiter el acceso de la más completa colección de vocabulario básico (listas Swadesh) de todas las lenguas tanto al público general como a especialistas. La novedad con respecto a otras bases de datos estructura es la estructuración de las aportaciones en 3 niveles; el primero constituido por una relativamente pequeña base de datos que contiene lenguas cercanas, con las que no existe controversia y cuya edad no excede los 3000 años (por ejemplo con lenguas germánicas); el segundo nivel los constituyen las protolenguas, y el ámbito temporal se amplía hasta los 6000 años (indoeuropeo); el último nivel es el más controvertido y enmarca el rango temporal de los 6000 a los 8000 años. En la terminología del GLD se usa el término grupo para el nivel 1, familia para el 2 y *macrofamilia* para el 3. La información en el grupo se complementa con etimología, uso, etc[9]. El

[8] El procedimiento es algo más complejo y es descrito pormenorizadamente por G. Jäger (2014: 9-14) y pasa por la realización de un minucioso análisis de las distancias entre las distintas lenguas y la calibración de los resultados mediante el valor-p (una función de estadística de frecuencia), así como mediante el algoritmo Needleman-Wunsch, que permite realizar alineamientos entre dos secuencias. Por medio de este algoritmo se consiguen los mejores resultados (Jäger: 2014, 15).

[9] La base de datos puede consultarse en http://lexstat.tk/databases/ (Starostin G. 2011-2016)

resultado es una completa herramienta que permite analizar a varios niveles las relaciones lingüísticas entre lenguas de la misma o distinta familia. El proyecto está en continuo proceso de ampliación y recientemente han sido incluídas (22-5-2016) el antiguo italiano, basado en el corpus de Dante Alighieri y el antiguo francés, basado a su vez en las obras de Chrétien de Troyes.

4. LA BIOINFORMÁTICA Y LA RECONSTRUCCIÓN DE PROTOLENGUAS

Como hemos señalado, existen, por tanto dos procedimientos bien diferenciados en la generación de datos clasificables informáticamente. Por un lado tenemos los datos descriptivos basados en un análisis de caracteres (*Ielex*) y por el otro el análisis de las distancias entre los cognados que es la base para los listados de ASJP. Se trata respectivamente de criterios de similitud y de distancia.

Ambos tipos de datos son similares a los utilizados en la metodología de análisis filogenético y de este modo podemos utilizar las herramientas usadas en esta disciplina para realizar un acercamiento a la evolución lingüística.

La similitud entre un análisis de ADN y una comparación lingüística en la que las mutaciones de caracteres tienen una perfecta correspondencia con las mutaciones de las especies. De este modo los algoritmos de análisis de los programas de filogenética se convierten en una herramienta sumamente útil para la reconstrucción lingüística (Jäger: 2016).

Existen diversos métodos usados habitualmente en filogenética: la cladística que usa el principio de Máxima Parsimonia (MP), la Máxima Verosimilitud - Maximum Likelihood (ML) y la Inferencia Bayesiana (IB) (Peña 2011: 265).

El fin de un análisis filogenético es hallar una filogenia, esto es un árbol que muestre la evolución de un grupo de estudio. Se basa en las relaciones de proximidad de las distintas ramas y a partir de ahí se intenta reconstruir la evolución biológica. En caso concreto que nos ocupa, de una *familia* lingüística o un grupo específico de lenguas. El análisis informático permite que mediante algoritmos se evalúen todas las posibilidades, todos los árboles, y mediante las correspondientes metodologías de análisis se obtenga el árbol más probable.

[10] Paup* es uno de los programas más usado para el cálculo de inferencia filogenética. Realizado por la Florida State University puede ser descargado en la siguiente dirección http://paup.sc.fsu.edu/downl.html.

[11] El árbol de consenso se construye a partir de otros árboles, siendo una especie de resumen de un conjunto de árboles.

En el método de MP, el árbol que se genera parte del principio de que el más simple sería el que mejor explica las relaciones. Es decir parte de la premisa de que las mutaciones son poco probables. El árbol filogenético resultante implica un mínimo de cambios evolutivos (Peña 2011: 266). De este modo, el árbol que se genera parte del principio de simplicidad. En el gráfico que a continuación reproducimos mostramos un árbol de lenguas indoeuropeas generado a partir de un algoritmo de Máxima Parsimonia con Paup* [10]. El resultado global fueron 16 árboles y a partir de ellos se elaboró un único árbol de consenso[11] (Figura 1).

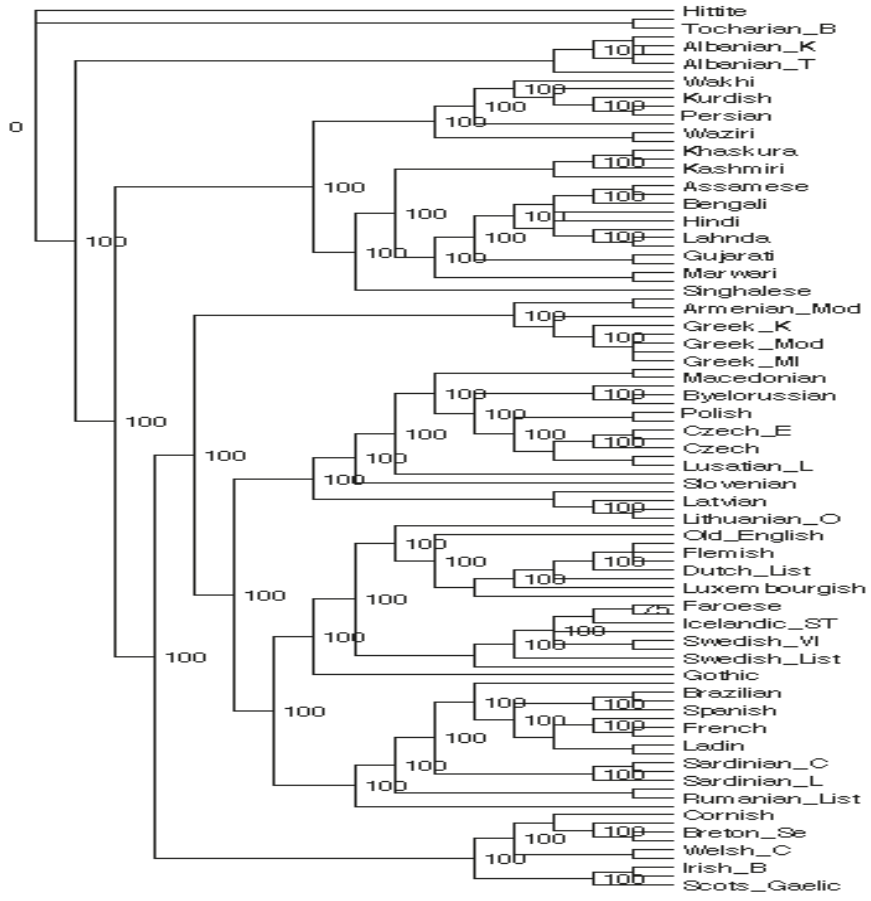

*Figura 1: Árbol de Máxima Parsimonia de las lenguas infoeuropeas generado por Paup**

El resultado con este algoritmo delimita de un modo bastante claro las distin-

tas familias indoeuropeas, como podemos constatar, por ejemplo, con las familias germánica, eslava o románica.

Otro tipo de análisis es el de máxima verosimilitud ML. Esta metodología presupone que las mutaciones no son tan raras. El árbol que se genera es el que tiene mayor probabilidad de haber generado todos los datos examinados.

En el ejemplo que mostramos podemos observar un análisis de Máxima Verosimilitud referido a diversas lenguas germánicas. (Figura 2)

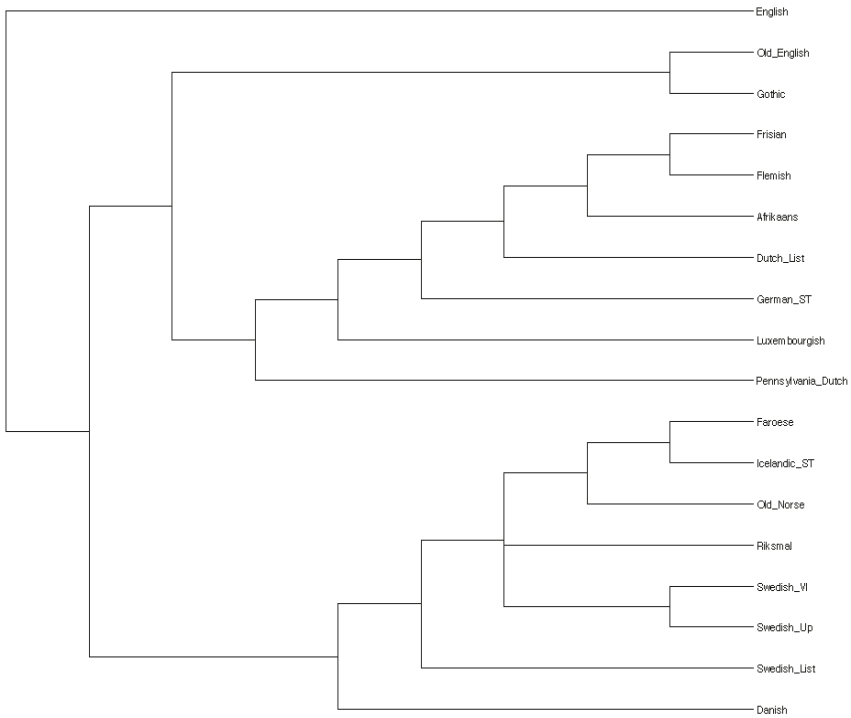

Figura 2: Árbol de Máxima Verosimilitud generado por PAUP para la familia de lenguas germánicas*

Es evidente la diferencia con el árbol tradicional, pero las relaciones entre el Afrikaans o el Neerlandés están perfectamente marca¬das, o también cómo ha sido delimitada la rama de las len¬guas escandinavas. La separación completa del inglés

o la relación en¬tre el antiguo inglés o el gótico, son obviamente los elementos más controvertidos del árbol generado.

La inferencia bayesiana es otra de las metodologías de acercamiento que hemos mencionado. La base es el teorema de Bayes, mediante el cual se calcula la probabilidad de que nuestro árbol sea correcto condicionado por los datos que tenemos. Los métodos de IB tienen una relación bastante estrecha con los de ML, ya que la hipótesis preferida es aquella que tiene mayor probabilidad posterior, y esto se calcula en función de la verosimilitud. La potencia de cálculo necesaria es inferior a los algoritmos de ML y suele ser más rápida en la generación de resultados.

5. CONCLUSIÓN

Ciertamente el análisis a partir de herramientas filogenéticas no está exento de problemas o dificultades, tal y como hemos tenido ocasión de ejemplificar, y en gran parte de los casos es necesaria la intervención de un especialista que revise o modifique los resultados obtenidos y los evalúe. Esta es nuestra posición, sostenida también por G. Starostin con su defensa de la combinación de métodos léxico-estadísticos y etimo-estadísticos (Starostin 1989: 27). De todos modos, no tenemos que olvidar que estamos probando unos modelos de análisis a partir de unos datos ya conocidos y quizá los resultados más valorables los podamos obtener con el análisis de protolenguas o incluso en la reconstrucción lingüística. Es el caso, por ejemplo, de A. Bouchard-Côté que ha desarrollado un modelo de reconstrucción lingüística a partir de métodos estadísticos, en concreto a partir del método Monte Carlo (Bouchard 2009). O también el proyecto liderado por R. Bouckaert, M. Dunn, Q. Atquinson o S. Greenhill entre otros, quienes partiendo de procedimientos bayesianos han desarrollado un método de inferencia filogeográfica para apoyar la tesis anatolia en el origen del pueblo indoeuropeo[12] (Bouckaert 2012: 957-960). En este punto, con ausencia de datos lingüísticos o históricos, es cuando la filogenética puede ayudarnos a entender, a relacionar o a inferir una evolución lingüística, geográfica o incluso cronológica.

En cualquier caso, la evolución de la glotocronología iniciada por M. Swadesh ofrece en los últimos tiempos múltiples posibilidades para el estudio de sincrónico y diacrónico de la lengua y su aplicación tanto en el ámbito investigador como en

[12] El grupo sostenido por la Universidad de Auckland posee una página informativa en la siguiente dirección: language.cs.auckland.ac.nz, con el título Mapping the origin of Indeuropean, en la que existe abundante material gráfico para complementar el artículo mencionado de *Science*.

el docente. Se trata indudablemente de una interesante renovación de los estudios tradicionales a partir de algoritmos matemáticos que pueden ayudar al profesor de lenguas extranjeras a tener una sólida fundamentación teórica que le sirva para sustentar el enfoque metodológico en el que se basan sus clases de lengua. Teniendo siempre presente que el aspecto léxico es una parte relativamente fluctuante en la evolución lingüística, con las limitaciones que esto conlleva.

Con metodologías de aprendizaje cada vez más centradas en el léxico y la comunicación, las aportaciones de las bases de datos realizadas por *Ielex*, ASJP o GDL proporcionan una fuente de información altamente valorable en la enseñanza contrastiva de lenguas. En ese sentido, la glotocronología constituye, a nuestro modo de ver, un campo de estudio que puede ofrecer un enfoque metodológico enormemente novedoso y sugestivo.

6. REFERENCIAS BIBLIOGRÁFICAS

BOUCHARD-CÔTÉ, A., et. al. (2009), «Improved Reconstruction of Protolanguage Word Forms». *Proceedings of the North American Chapter of the Association for Computational Linguistics* (NAACL09). 7, 65-73.

BOUCKAERT, R., LEMEY, P., DUNN, M., GREENHILL, S. J., ALEKSEYENKO, A. V., DRUMMOND, A. J., GRAY, R. D., SUCHARD, M. A., & ATKINSON, Q. D. (2012) «Mapping the origins and expansion of the Indo-European language family». *Science*, 337, 957–960.

COSERIU, E. (1965), «Critique de la glottochronologie appliquée aux langues romanes.» *Linguistique et Philologie Romanes. Xe Congrès International de Linguistique et Philologie Romanes*. París: Klincksieck.

DYEN, I., KRUSKAL, J. B., BLACK, P. (1992), *An Indoeuropean Classification: A Lexicostatistical Experiment,* The American Philosophical Society, Philadelphia.

EHRET, CH. (2000), «Testing the Expectations of Glottochronology against the Correlations of Language and Archeology in Africa.» en Colin Renfrew, April McMahon and Larry Trask, (eds.) *Time Depth in Historical Linguistics*, Cambridge: MacDonald Institute for Archaeological Research, 373-399.

JÄGER, G. (2014), «Lexikostatistik 2.0», en A. Plewnia & A. Witt, eds., *Sprachverfall? Dynamik - Wandel - Variation. Jahrbuch 2013 des Instituts für Deutsche Sprache*, Berlin, de Gruyter, 197-216.

MARCHUK, Y. N. (2003), «The Burdens and Blessings of Blazing the Trail». *Journal of Quantitative Linguistics.* Bd. 10, 2 , 81–85

JÄGER, G. (2016), Investigating the potential of ancestral state reconstruction al-

gorithms in historical linguistics, C. Bentz, G. Jäger and I. Yanovich, eds.
PEÑA, C. (2011), «Métodos de inferencia filogenética», Rev. Per. biol. 18(2), 265-267.
SAPIR, E. (1921), *Language. An introduction to the study of speech,* New York, Brace.
STAROSTIN, S. (2007), Opredelenije ustojčivosti bazisnoj leksiki [Defining the Stability of Basic Lexicon] // S. Starostin. Trudy po jazykoznaniju [Works in Linguistics]. Moscow, Jazyki slav'anskix kul'tur, (2007) 825–839
STAROSTIN, G. (1989), "Sravnitel'no-istoričeskoe jazykoznanie i leksikostatistika", en "Lingvističeskaja rekonstrukcija i drevnejšaja istorija Vostoka", Moscú (trad. al inglés por I. Peiros y N. Evans Comparative-historical linguistics and lexicostatistics. Historical Linguistics and Lexicostatistics), 3-50
STAROSTIN, G. (2010), Preliminary lexicostatistics as a basis for language classification: A new approach*, Journal of Language Relationship,* No. 3: 79–116
STAROSTIN, G. (Ed.) (2011-2016), *The Global Lexicostatistical Database*. Moscow: Russian State University for the Humanities, & Santa Fe: Santa Fe Institute. Disponible online en http://lexstat.tk/databases/, acceso [28-5-2016].
SWADESH, M. (1955), «Towards Greater Accuracy in Lexicostatistic Dating.» *International Journal of American Linguistics*. 21: 121- 137.
SWADESH, M. (1958), «Sobre la clasificación del otomí-pame», Actas del 33 Congreso Internacional de Americanistas 2: 551-559.
SWADESH, M. (1971), *The Origin and Diversification of Language*. Aldine, Chicago.

Fecha de recepción: 7 de noviembre de 2017
Fecha de aceptación: 28 de febrero de 2018