

**APROXIMACIÓN ESTADÍSTICA A LA LINGÜÍSTICA
HISTÓRICA: RECONSTRUCCIÓN DEL PROTOGERMÁNICO A
PARTIR DEL MÉTODO DE MONTECARLO**

Francisco Javier Muñoz-Acebes
Universidad de Valladolid

Resumen: En el presente artículo investigaremos en torno a la posibilidad de uso de algoritmos matemáticos enfocados a la reconstrucción lingüística. Dado que consideramos que su uso puede aportar algunos beneficios a la mencionada reconstrucción. Con la posibilidad de manejar gran cantidad de datos y posibilidades podemos testar de un modo mucho más eficiente determinadas hipótesis en torno a los cambios que ocurren en la evolución lingüística. Nuestro enfoque se centrará en la posibilidad de reconstrucción del protogermánico a partir de los datos de los corpus conservados en gótico, alto alemán antiguo, inglés y nórdico antiguos, fundamentalmente. Del mismo modo, nos ceñiremos a la utilización del algoritmo matemático Montecarlo. Esto no indica que los materiales propuestos para la reconstrucción no puedan ser usados con otras lenguas o que otros algoritmos no puedan ser utilizados, pero ambas son cuestiones que quedan fuera de las pretensiones de la presente investigación.

Palabras clave: Reconstrucción lingüística, Protogermánico, Montecarlo.

Abstract: In this article, we will investigate the possibility of using mathematical algorithms for linguistic reconstruction, as we believe their use can contribute some benefits to the mentioned reconstruction. With the ability to handle a large amount of data and possibilities, we can more efficiently test certain hypotheses regarding changes that occur in linguistic evolution. Our focus will be on the possibility of reconstructing Proto-Germanic based on data from the preserved corpora in Gothic, Old High German, Old English, and Old Norse primarily. Likewise, we will adhere to the use of the Monte Carlo mathematical algorithm. This does not imply that the materials proposed for reconstruction cannot be used with other languages or that other algorithms cannot be utilized, but both are matters beyond the scope of this research.

Keywords: Linguistic Reconstruction, Protogermanic, Montecarlo.

1. INTRODUCCIÓN: LOS PROBLEMAS DE LA RECONSTRUCCIÓN DEL PROTOGERMÁNICO

El éxito del método comparativo en el siglo XIX mostró el proceso de las leyes que gobiernan el cambio lingüístico. Sin embargo, gran parte de los procesos evidencian cierto grado de improvisación, o incluso de intuición personal más que un modo

estricto de aplicación de principios mecánicos (Fox 1996: 57). A esto podemos añadir el problema de en qué grado una forma reconstruida se corresponde con una lengua que alguien habló en algún momento (Gilman 2012: 132). La reconstrucción puede darnos la falsa creencia de que la lengua es uniforme y de que no existían variedades dialectales. Las lenguas reconstruidas son por definición y método uniformes en su máxima expresión, pero la lengua que en realidad fue hablada ¿poseía ese grado de uniformidad? (Pulgram 1959: 424). La reconstrucción, evidentemente, no nos puede indicar el grado de dialectización de la lengua.

El método comparativo tradicional muestra especial fortaleza cuando nos encontramos con la repetición del mismo sonido en una serie de cognados. De este modo, si encontramos en los cognados del número *cuatro* y de *pájaro* la misma consonante inicial [f], tendremos la suficiente seguridad de que la protolengua también comenzaría por el mismo sonido:

<i>Gótico</i>	<i>Alto alemán antiguo</i>	<i>Nórdico antiguo</i>	<i>Inglés antiguo</i>
𐍆𐌺𐍅𐍂 (fidwōr)	fior	fjórir	fēower
𐍆𐌺𐍈𐌺𐍂 (fugls)	fogal	fogl, fugl	fugol

Tabla 1: Vocablos con [f] inicial en lenguas germánicas

Sin embargo, cuando encontramos diversos sonidos o incluso diversas formas léxicas para el mismo cognado en las lenguas tomadas como referencia, el método comparativo ha de tomar en cuenta los posibles cambios fónicos y su regularidad observando otros ejemplos, como en el caso del fonema [t]:

<i>Gótico</i>	<i>Alto alemán antiguo</i>	<i>Nórdico antiguo</i>	<i>Inglés antiguo</i>
𐍅𐌶𐌺𐌿𐍄𐌸𐌹𐌸 (dauhtar)	tohter	dóttir	dohtor
𐍆𐌺 (þu)	dū	þú	þū
𐍂𐌻𐌴𐍆 (blōþ)	bluot	blóð	blōð

Tabla 2: Vocablos con fonema [t] en lenguas germánicas

Uno de los problemas frecuentes lo encontramos cuando creemos estar trabajando con palabras que consideramos cognados y que en realidad son préstamos. Si observamos por ejemplo la palabra en alemán moderno *Eisen*, podemos constatar la relación con el inglés moderno *iron*, el danés *jern* o el neerlandés *ijzer*. Remontándonos a estratos más antiguos constatamos en alto alemán medio *isen*, *iser*, forma a su vez procedente del alto alemán antiguo *isarn*, que se corresponde con el inglés antiguo *isern*, el neerlandés medio *iser*. La forma habitualmente reconstruida en PGM **isarną* y que con gran probabilidad es un préstamo del proto céltico **isarnom* (Birkhan 1970: 128-129), atestiguado en gálico como nombre propio *Isarnus* (Matasović 2009: 172).

Pero estos problemas no invalidan el método y, ciertamente, las múltiples dificultades no deben impedirnos ver las numerosas virtudes. Una de las claves es realizar un proceso sistemático a partir de las siguientes instrucciones (Durie, Ross 1996: 6-7):

1. Determinar, basándose en evidencia, que un conjunto de lenguas están genéticamente relacionadas, es decir, que constituyen una familia lingüística.
2. Recopilar conjuntos de cognados propuestos para la familia (tanto paradigmas morfológicos como elementos léxicos).
3. Determinar las correspondencias fonéticas a partir de los conjuntos de cognados, separando los conjuntos de cognados "irregulares".
4. Reconstruir el protolenguaje de la familia de la siguiente manera:
 - 4.1. Reconstruir la protofonología a partir de las correspondencias fonéticas desarrolladas en el paso 3, utilizando el conocimiento convencional sobre las direcciones de los cambios fonéticos.
 - 4.2. Reconstruir los protomorfemas (tanto paradigmas morfológicos como elementos léxicos) a partir de los conjuntos de cognados recopilados en el paso 2, utilizando

- la protofonología reconstruida en el paso 4.1.
5. Establecer innovaciones (fonológicas, léxicas, semánticas, morfológicas, morfosintácticas) compartidas por grupos de lenguas dentro de la familia en relación con el protolenguaje reconstruido.
 6. Tabular las innovaciones establecidas en el paso 5 para obtener una clasificación interna de la familia, es decir, un "árbol genealógico".
 7. Elaborar un diccionario etimológico que rastree los préstamos, los cambios semánticos, etc., para el léxico de la familia (o de una de las lenguas de la familia).

G. Jäger (2019: 153) las sintetiza en el siguiente diagrama de flujo:

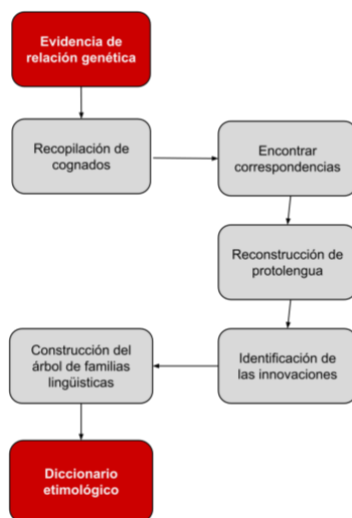


Figura 1: Diagrama del método comparativo

Del mismo modo, es importante establecer la relación entre los grupos lingüísticos y la época de los mismos con el objeto de poner en valor la importancia que debemos otorgar a cada lengua en nuestro intento de reconstrucción.

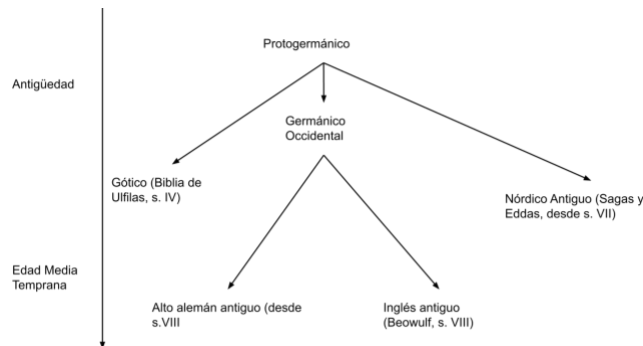


Figura 2: Árbol de familias lingüística y temporalización en el eje vertical

De modo más concreto y ciñéndonos al protogermánico, la reconstrucción del mismo presenta dificultades a causa de factores como la falta de evidencias por la ausencia de textos escritos o a la pérdida de léxico. Del mismo modo, la existencia de variedades dialectales en lenguas germánicas dificulta la identificación de una raíz común (la paradoja apuntada por Pulgram mencionada más arriba). De tal modo que la reconstrucción del protogermánico puede ser considerada como una abstracción que intenta crear una imagen completamente cerrada en la línea de lo que proponía la *Stammbaumtheorie* (Fulk 2018: 11), siendo en realidad una reconstrucción completamente especulativa.

A pesar de todo ello, muchas de las evidencias lingüísticas en el caso concreto del protogermánico nos permiten realizar una reconstrucción posible y confiable a nivel léxico-fonológico (Walkden, 2014: 58), aunque no exenta de controversia.

La base léxica además puede ser tomada como punto de inicio para una posible correspondencia sintáctica, pero teniendo en cuenta que una gramática no se transmite históricamente y poco podemos deducir de una proto-gramática a partir del estudio de las propiedades formales de las gramáticas de las

lenguas hijas (Lightfoot, 1980: 37)¹.

En ese complejo marco, pensamos que la estadística puede proporcionarnos datos que permiten ayudar al lingüista en la reconstrucción.

2. MÉTODOS ESTADÍSTICOS EN LINGÜÍSTICA HISTÓRICA

En los últimos años la lingüística histórica se está viendo beneficiada por el uso de herramientas computacionales y el análisis estadístico de datos. Los beneficios de este tipo de análisis estadístico se han podido constatar en múltiples ámbitos que abarcan los análisis estilométricos², la realización de filogenias o la reconstrucción de lenguas³.

En este contexto, el uso de algoritmos estadísticos en el estudio de la lingüística histórica puede ser enormemente positivo ya que permitiría al investigador desarrollar nuevas perspectivas o testar diferentes posibilidades.

Desde la década de los 60 del siglo pasado, varios investigadores han aplicado métodos computacionales (Jäger 2019: 170-171) en aspectos del método comparativo (Ringe 1992; Baxter y Manaster Ramer 2000; Kessler 2001), para evidenciar relaciones de familias lingüísticas, para la recopilación y detección de cognados (Kay 1964, Rama y List, 2012, Kondrak, 2002 y Covington, 1996), para la reconstrucción (Lowe y Mazaudon 1994), en la recopilación de cognados y etimologías (Oakes 2000) y para la reconstrucción de familias lingüísticas (Ringe et al. 2002; Rama, 2018; Muñoz, 2019 y 2022; Ni, 2020) e incluso para determinar la patria originaria indoeuropea (Bouckaert et al. 2012).

¹ Los estudios de Walkden (2014) a partir de corpus digitales etiquetados de inglés antiguo, sajón antiguo, alto alemán antiguo, islandés antiguo y gótico han ofrecido solución a aspectos concretos de la sintaxis del protogermánico, como la posición del verbo, los elementos “wh” o la elisión de objetos y sujetos.

² Prueba de ello es la aparición de múltiples paquetes en R que permiten este tipo de análisis, como *stylo*, *quanteda*, *textmineR*, *stylometry*, *TextAnalysis*, por mencionar algunos.

³ Con paquetes como *stringDist* o *Ape*.

En el ámbito específico de la reconstrucción de lenguas muertas, es indispensable mencionar Bouchard-Côté (Bouchard-Côté et al. 2013: 4224-4229) y su propuesta de reconstrucción de la protolengua austronesia a partir de un algoritmo Montecarlo, sin embargo, las herramientas algorítmicas no han sido publicadas. Del mismo modo, Jäger (2019: 151-182) realiza la reconstrucción del proto-romance a través de un método de inferencia bayesiana (Jäger, 2019: 170-171)⁴.

El análisis cuantitativo ha mostrado grandes posibilidades a nivel fonemático (Bouchard-Côté et al, 2007), combinando la fortaleza del método de reconstrucción tradicional con la del método probabilístico. El uso de determinados algoritmos estadísticos puede darnos respuestas en casos en los que la lingüística histórica tradicional puede ofrecer dudas.

El enfoque que queremos presentar propone un acercamiento al problema de la reconstrucción del protogermánico a partir del enfoque estadístico, para lo cual nos valdremos de RStudio.

R se ha convertido en los últimos años en una herramienta usada con frecuencia dentro del ámbito educativo universitario por su simplicidad y versatilidad frente a otros lenguajes de programación como Matlab o Bugs (Robert, Casella, 2010: vii) y ha supuesto un enorme cambio en el ámbito de las humanidades digitales dado que su uso no requiere unos grandes conocimientos informáticos o estadísticos. Algo que habitualmente no está incluido en los programas de formación de estudiantes o investigadores de humanidades. De este modo, muchos paquetes proporcionan un entorno gráfico que posibilita la elección de opciones sin necesidad de tener conocimientos del código de programación o de los procesos estadísticos involucrados.

El uso de métodos estadísticos en la reconstrucción

⁴ En este caso con completa publicación de los datos para facilitar la replicabilidad.

lingüística es una herramienta muy efectiva sobre todo con lenguas reconstruidas o de Corpus Mínimos⁵. Tradicionalmente la lingüística histórica se basa en la regla de la mayoría que vimos al inicio de este trabajo para reconstruir protoformas (Walkden 2014: 47-48). Por la cual, si la mayoría de un grupo de cognados presenta un rasgo común, es muy probable que dicho rasgo esté presente en un estadio lingüístico anterior; igualmente, el principio de desarrollo natural: que implica que algunos tipos de cambio lingüístico son más frecuentes que otros (así, el cambio de oclusivas a fricativas, la desaparición de vocales finales, el ensordecimiento de consonantes finales o la sonorización de consonantes intervocálicas, por mencionar algún ejemplo).

El cambio sustancial respecto a la disciplina tradicional lo proporciona el volumen de datos que pueden ser analizados con herramientas computacionales. En ese sentido, la estadística proporciona la posibilidad de realizar inferencias a partir de gran cantidad de datos. Básicamente permite realizar el procedimiento comparativo tradicional usado desde Schleicher pero con la posibilidad de incrementar los datos y de testar posibilidades.

De este modo el uso de métodos estadísticos en la reconstrucción lingüística es un elemento fundamental para entender el desarrollo histórico de las lenguas. Dado que permite llenar alguno de los frecuentes vacíos con los que se suele encontrar la reconstrucción.

2.1. El método de Montecarlo

El método de Montecarlo es un enfoque estadístico que utiliza simulación para obtener resultados. Fue desarrollado en los años 40 por los físicos John von Neumann y Stanislaw Ulam durante el proyecto Manhattan. Inicialmente se utilizó para

⁵ Con el término “Corpus Mínimos” nos referimos a lenguas de corpus de las que solo contamos con algunas palabras o nombres propios, un vocabulario, por tanto, muy limitado y una morfología generalmente desconocida. En alemán este tipo de corpus limitado a palabras sueltas o a inscripciones se denomina *Trümmersprachen* (Untermann, 1980: 7-8), literalmente “lenguas en escombros/restos”.

modelar sistemas físicos complejos, como los reactores nucleares, pero rápidamente se expandió a otras áreas, como las finanzas, la ingeniería o la biología. El nombre hace referencia a los casinos de Montecarlo y fue inspirado por un familiar de Ulam que siempre pedía dinero prestado porque “just had to go to Monte Carlo” (Metropolis 1987: 3).

Consiste en realizar un gran número de simulaciones para un problema, y utilizar los resultados para obtener una respuesta aproximada al problema original. Cada simulación se basa en un conjunto de números aleatorios, y se hace un seguimiento de los resultados obtenidos para obtener una distribución estadística de las respuestas posibles.

Por ejemplo, si queremos conocer el área de un círculo podemos hacerlo generando puntos aleatorios dentro de un círculo y a partir de ellos estimar el área⁶.

Referente a la reconstrucción de lenguas muertas o desaparecidas, el método de Montecarlo podemos usarlo para generar hipótesis sobre la evolución de una lengua. Por ejemplo, se pueden simular diferentes escenarios de cambio fonético y comparar los resultados con los datos disponibles para determinar la hipótesis más probable. También se puede utilizar para evaluar la robustez de las reconstrucciones lingüísticas existentes.

2.2. Código de programación

R nos ofrece múltiples posibilidades de integración de simulaciones Montecarlo (Robert y Casella, 2010). En nuestra propuesta partimos de un programa⁷ en el que definimos las

⁶ Hemos realizado un programa en R que realiza justamente eso: utiliza un método probabilístico para estimar el área de un círculo generando puntos aleatorios y contando cuántos de esos puntos caen dentro del círculo. Cuanto mayor sea el número de puntos generados, más precisa será la estimación del área del círculo. (<https://doi.org/10.5281/zenodo.8013716>).

⁷ El programa está disponible en el repositorio zenodo en el siguiente enlace: <https://doi.org/10.5281/zenodo.8034005>.

palabras de las lenguas hijas, especificando una lista de palabras en diferentes lenguas germánicas antiguas. Estas palabras serán la base de la reconstrucción en la lengua madre.

A continuación, definimos el número de iteraciones para el muestreo Montecarlo, partimos de la premisa de que, cuantas más iteraciones, más precisa será la reconstrucción final, pero también incrementando ostensiblemente los tiempos de proceso.

El siguiente punto es la definición de las leyes fonéticas y sus posibilidades de cambio. Cada ley se establece como una función que realiza un cambio fonético específico en una palabra. Podemos introducir cambio de un fonema por otro, eliminación de algún fonema (simplemente dejando en blanco el fonema destino) o la sustitución de un fonema por dos o más. También se asigna una probabilidad entre 0 y 1 de ese cambio lingüístico. Este parámetro determina la frecuencia con la que se aplicará la ley en el proceso de reconstrucción.

```
ley1 = list(ley = function(x) gsub("i", "u", x),
probabilidad = 0.4), # Cambio de la "i" por "u"
ley2 = list(ley = function(x) gsub("p", "b", x),
probabilidad = 0.2), # Reemplazo de "p" por "b"
ley3 = list(ley = function(x) gsub("f", "b", x),
probabilidad = 0.3), # Reemplazo de "f" por "b"
ley4 = list(ley = function(x) gsub("ou", "au",
x), probabilidad = 0.4), # Reemplazo de "ou" por
"au"
ley5 = list(ley = function(x) gsub("t", "d", x),
probabilidad = 0.4) # Reemplazo de "t" por "d"
```

Figura 3: Fragmento del código con las leyes fonéticas introducidas para la reconstrucción de la palabra *haubuda (cabeza).

Cada regla contiene tres elementos configurables: el primer elemento es el fonema (o letra) original, el segundo elemento es el fonema (o letra) al que se cambia y el tercer elemento es una probabilidad entre 0 y 1 que indica la

probabilidad de que se aplique esa regla en particular.

Elegir los porcentajes adecuados para las reglas de cambio fonético puede ser un desafío, pues depende de la lengua específica y de la palabra objetivo que se desea reconstruir. Sin embargo, hay algunos principios generales que pueden ser de ayuda.

Un enfoque puede basar los porcentajes en la frecuencia de cambio de cada letra en la evolución del idioma, basándonos en los patrones de cambio observados o en reglas ya conocidas. Por ejemplo, si se observa cómo cambiaron las palabras similares a la palabra objetivo en las lenguas germánicas, podría obtenerse una idea de qué reglas son las más comunes y asignar un porcentaje más alto a estas reglas. Por ejemplo, otorgando un mayor peso a los fonemas presentes en estratos lingüísticos más antiguos. En este punto la labor y el conocimiento del lingüista es el que ha de determinar los datos a introducir.

En cualquier caso, el programa permite aplicar las reglas de cambio fónico que consideremos para cada caso, o incluso prescindir de ellas totalmente y desarrollar un muestreo exclusivamente estadístico como veremos más adelante.

Posteriormente se realiza el muestreo Montecarlo en el que se aplican las leyes fonéticas y se reconstruyen las palabras en la lengua madre en cada iteración. Las palabras hijas se someten a las leyes fonéticas de manera aleatoria, lo que genera diferentes variaciones en cada iteración.

Finalmente se muestra en pantalla la palabra reconstruida que aparece con mayor frecuencia después de todas las iteraciones del muestreo Montecarlo. La palabra obtenida sería la posible reconstrucción.

Los datos generados con el programa se han subido al repositorio Zenodo para facilitar su consulta⁸.

2.3. *Aplicación*

⁸ <https://doi.org/10.5281/zenodo.8034042>.

Para testar las posibilidades de la reconstrucción mediante algoritmos estadísticos vamos a analizar un pequeño corpus de palabras en lenguas germánicas antiguas: alto alemán antiguo, nórdico antiguo, inglés antiguo, gótico y sajón antiguo, lenguas de las que se han conservado textos y están razonablemente bien documentadas.

<i>Vocablo</i>	<i>Alto alemán antiguo</i>	<i>Nórdico antiguo</i>	<i>Inglés antiguo</i>	<i>Gótico</i>	<i>Sajón antiguo</i>
alto	hōh	hár	hēah	hauhs	hōh
amigo	friunt	frændi	frēond	frijōnds	friund
cabeza	houbit, haupt	hēafod	haufuð	haubiþ	héafod
hablar	sprehhan	spræka	spræcan, sprecan	sprekan	sprekan
padre	fater	faðir	fæder	fadar	fæder, feder
raíz	wurz	urt	wyrt	waurts	wurt
rey	kuning	konungr	cýning	kuni	kuning
semana	wehha	vika	wice, wicu	wikō	wika

Tabla 3: Corpus de prueba

La elección de los vocablos responde a varias razones. En primer lugar, probar la reconstrucción de fonemas variados y en algunos casos probar la reconstrucción estadística con conceptos que suscitan dudas o resultan problemáticos de cara a la reconstrucción.

2.4. Resultados

Las gráficas nos muestran la frecuencia de resultados

generados y tomamos las frecuencias más altas como resultados más probables a partir de 100.000 iteraciones.

<i>Vocablo protogermánico</i> ⁹	<i>Reconstrucción Montecarlo</i>	<i>Reconstrucción Montecarlo sin leyes fonéticas</i>
*fader (p.119)	*fader	*fader
*fri, frjôn (p.128)	*friund	*friund
*haubuda/hauba (p.45)	*haubud	*haufið
*hauha/hauga (p.54)	*hauh	*hauh
*kuninga (p.27)	*kuning	*kuning
*sprekan (p.256)	*sprekan/sprecan	*sprekan
*vikô(n) (p.205)	*wika	*wika
*vurti (p.200)	*vurt	*wurt

Tabla 4: Vocablos reconstruidos

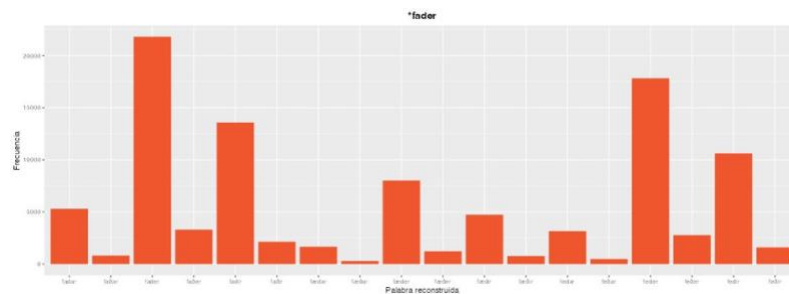


Figura 4: Reconstrucción de *fader¹⁰

⁹ Seguimos el léxico reconstruido tradicionalmente por Fick (1909). Los números de página siguen la edición electrónica de Crist y Studer (2006).

¹⁰ Todos los gráficos de frecuencias (Figura 4 a 11) han sido realizados con la librería esquisse (Meyer y Perrier, 2023).

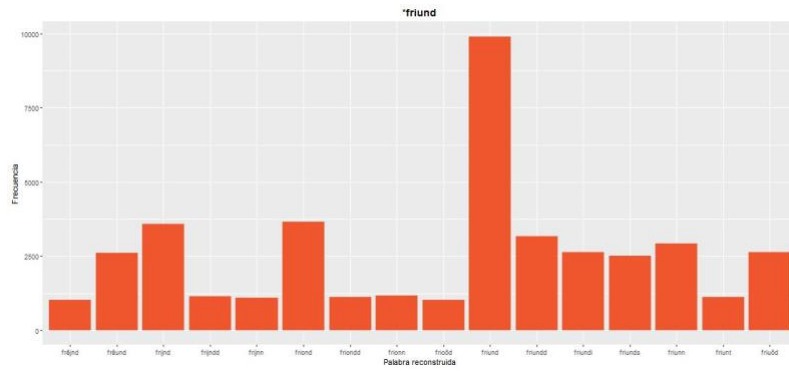


Figura 5: Reconstrucción de *friund

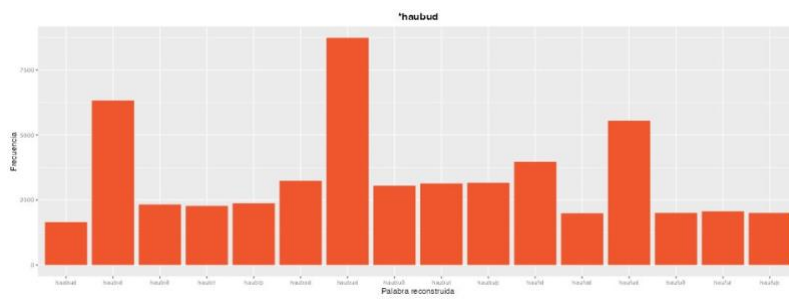


Figura 6: Reconstrucción de *haubud

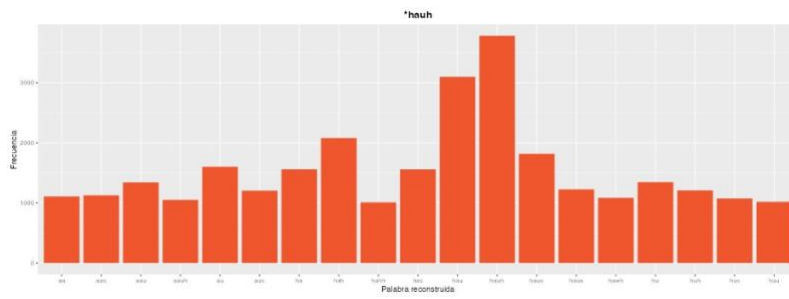


Figura 7: Reconstrucción de *hauh

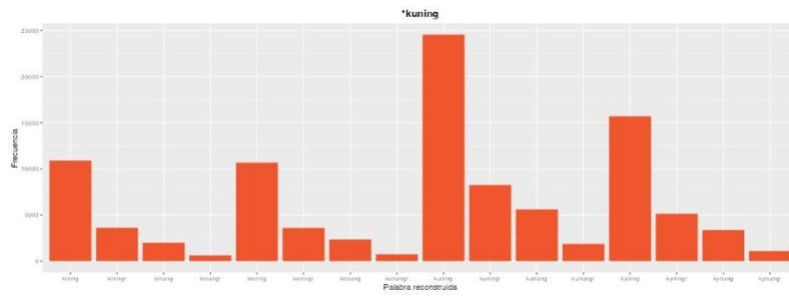


Figura 8: Reconstrucción de *kuning

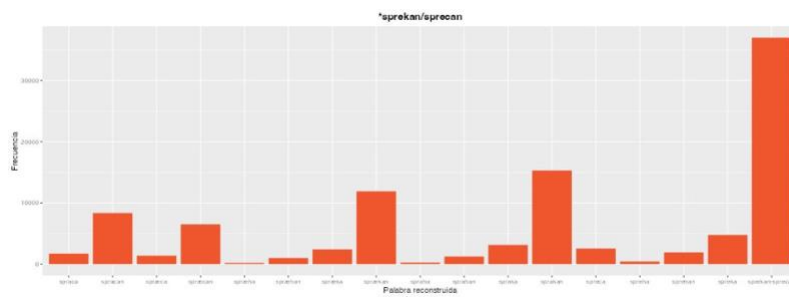


Figura 9: Reconstrucción de *sprekan/sprecan

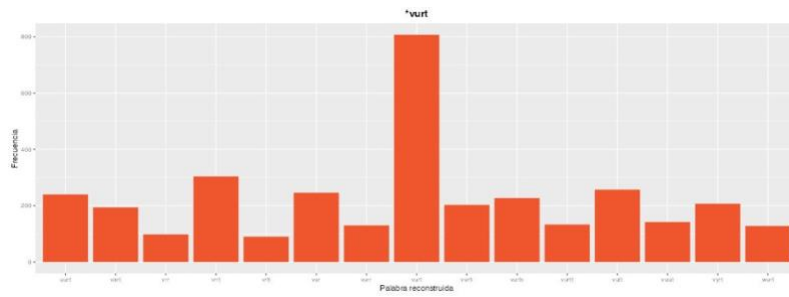


Figura 10: Reconstrucción de *vurt

El vocablo permanece prácticamente idéntico a su fuente 2000 años después. Este tipo de préstamos pueden proporcionar datos muy interesantes en torno a las etapas tempranas del germánico (Fortson, 2010: 338).

3. CONCLUSIONES

Es el momento de valorar los resultados obtenidos a partir de los ejemplos propuestos. Cabe preguntarse si los vocablos obtenidos no estarían completamente condicionados por las leyes fonéticas que hemos incorporado al código y para ello hemos intentado validar la reconstrucción usando también un código de reconstrucción que no implementaba dichas leyes¹² y realizando una reconstrucción exclusivamente estadística (ver tabla 4).

Lo primero que hay que señalar es que un enfoque exclusivamente estadístico no resulta viable para su uso en lingüística histórica, la labor de un lingüista es indispensable a la hora de incorporar datos o de fijar las posibles reglas de cambio lingüístico.

En la Tabla 4 con los vocablos reconstruidos, podemos constatar las diferencias y similitudes entre un tipo y otro de reconstrucción. Del mismo modo, se ha incluido la propuesta de reconstrucción de los vocablos analizados realizada a principios del siglo XX por August Fick. Existen múltiples coincidencias entre la reconstrucción estadística y la tradicional, aunque también notables diferencias, sobre todo con los fonemas [v] y [w] o el vocalismo en el vocablo *frîjôn. En el primer caso, no podemos conocer la posible pronunciación del sonido en protogermánico y las diferencias fónicas entre uno y otro sonido podrían no ser relevantes. Pensemos en la palabra *wulfaz (“lobo”) ¿su pronunciación sería sonora labiodental y fricativa [v] o más bien labiovelar aproximativa [w]? Parece claro que un proceso de delabialización de la labiovelar [w] se produjo en

¹² <https://doi.org/10.5281/zenodo.8027252>

algún momento de la evolución lingüística sin que se conozca cuándo y por qué (Ringe, 2006: 111).

El caso de *frijōn presenta otro tipo de circunstancias que también es necesario señalar. D. Ringe y A. Taylor (2014: 62) reconstruyen el vocablo “amigo” en germánico occidental como *friund, con un cambio *ijō a *iu¹³ (Luick, 1914-40: 118). El sustantivo *frijōndz sería un estrato lingüístico anterior reflejo de los participios presentes activos del PIE. Esos participios fueron “remodelados” en las lenguas germánicas antiguas dejando una clase residual de sustantivos en *-nd. (Ringe, 2006: 199).

A pesar de las evidentes coincidencias, nuestro breve estudio no pretende reformular un diccionario de protogermánico, si no simplemente testar las posibilidades de un método de reconstrucción estadístico en lenguas de corpus mínimos o que han de ser completamente reconstruidas.

El número de vocablos que hemos usado es reducido y ha de tomarse como una prueba de concepto del uso de algoritmos estadísticos en la reconstrucción del protogermánico y, ciertamente, nuestra aportación habrá de ser ampliada y testada con una base de datos mayor y con el análisis de otros corpus con el fin de ampliar los datos, cuestiones que sin duda quedan fuera de las pretensiones de este artículo. En ese sentido, alguno de los puntos que han de ser mejorados y completados son los siguientes:

1. Ampliación del conjunto de datos.
2. Tratamiento de la reconstrucción a nivel fonético.
3. Mejorar la solución para fenómenos que implican cambios consonánticos de orden como la metátesis o fenómenos de haplogía.
4. Inclusión en la reconstrucción otros posibles modelos estadísticos.

¹³ Esto lo podríamos haber definido como ley fonética en el código, pero al no estar suficientemente constatado el cambio, nos hemos decantado por no incluirlo; en línea de lo aseverado por Ringe que ve inadmisibles la realización de una regla general a partir de la historia de esta palabra (Ringe 2014: 62).

Lo positivo de nuestra propuesta es la posibilidad de condicionar los resultados a través de la incorporación de reglas fonéticas conocidas. De este modo, el código permite introducir datos con flexibilidad y puede ser útil para la reconstrucción lingüística o incluso para visualizar los efectos del cambio lingüístico.

Se abre también un posterior desarrollo que incorpore las distancias lingüísticas entre los cognados de las lenguas usadas con el fin de precisar el peso de los fonemas de cada lengua, en consonancia con lo planteado por G. Jäger en la reconstrucción del protorromance (2019: 170-171).

En definitiva, hemos pretendido demostrar que el uso de enfoques estadísticos y herramientas computacionales - específicamente R- en la reconstrucción lingüística ha ampliado las posibilidades de análisis y ha mejorado nuestra comprensión del desarrollo histórico de las lenguas. Estos enfoques permiten abordar los desafíos y limitaciones de la reconstrucción lingüística tradicional y proporcionan nuevas perspectivas para la investigación en este campo.

4. BIBLIOGRAFIA

- Baxter, William. H. y Manaster Ramer, Alexis (2000): Beyond lumping and splitting. Probabilistic issues in historical linguistics, en Colin Renfrew et al. (eds.), *Time depth in historical linguistics*, vol. 1, Cambridge, McDonald Institute for Archaeological Research, pp. 167-188.
- Birkhan, Helmut (1970): *Germanen und Kelten bis zum Ausgang der Römerzeit. Der Aussagewert von Wörtern und Sachen für die frühesten keltisch-germanischen Kulturbeziehungen*, Österr. Akademie der Wissenschaften, Phil.-hist. Kl., Sitzungsberichte, 272. Bd. Viena, Böhlau.
- Bouchard-Côté, Alexandre; Liang, Percy S.; Griffiths, Thomas; Klein, Dan (2007): «A probabilistic approach to language change», *Advances in Neural Information Processing Systems 20 (NeurIPS) 20*, pp. 169-176.
- Bouchard-Côté, Alexandre; Hall, David; Griffiths, Thomas L.; Klein, Dan (2013): «Automated reconstruction of ancient languages using probabilistic models of sound change», *Proceedings of the National Academy of Sciences*, 110, pp. 4224-4229. <https://doi.org/10.1073/pnas.120467811>.
- Covington, Michael A. (1996): «An algorithm to align words for historical comparison», *Computational Linguistics* 22(4), pp. 481-496.
- Fortson, Benjamin (2010): *Indo-European Language and Culture. An Introduction*, Oxford,

- Wiley/Blackwell.
- Fox, Anthony (1995): *Linguistic Reconstruction: An Introduction to Theory and Method*, Oxford, OUP.
- Fick, August Friedrich (1909), *Wortschatz der Germanischen Spracheinheit*, en *Wörterbuch der Indogermanischen Sprachen*, Göttingen, Vandenhoeck and Ruprecht. Versión electrónica a cargo de Sean Crist y Dieter Studer (2006): en línea http://www.germanic-lexicon-project.org/pdf/pgmc_torp/pgmc_torp_20061216.pdf [21-7-2023]
- Fulk, Robert Dennis (2018): *A comparative Grammar of the Early Germanic Languages*, Amsterdam, John Benjamins. <https://doi.org/10.1075/sigl.3>
- Gilman, Sophia. (2012): «Comparative Method Algorithm», en *Cambridge Occasional Papers in linguistics*, 6(5), pp. 131-175.
- Jäger, Gerhart. (2019): «Computational Historical Linguistics», en *Theoretical Linguistics* 45: 151-182. <https://doi.org/10.1515/tl-2019-0011>.
- Kay, Martin (1964): *The logic of cognate recognition in historical linguistics*. Santa Monica, CA: Rand Corporation.
- Keretchashvili, George (2018-2023): *Chrestomathy of Gothic and Anglo-Saxon*. En línea: <https://germanic.ge/en/ang/> [2-6-2023]
- Kessler, Brett (2001): *The Significance of Word Lists. Statistical Tests for Investigating Historical Connections Between Languages*, Chicago, Center for the Study of Language and Information, CSLI Publications.
- Kondrak, Grzegorz (2002): *Algorithms for language reconstruction*, University of Toronto, PhD thesis.
- Krahe, Hans (20197): *Germanische Sprachwissenschaft / Einleitung und Lautlehre*, Berlín, De Gruyter.
- Lightfoot, David (1980): «On reconstructing a proto-syntax», en Paolo Ramat (ed.) *Linguistic reconstruction and Indo-European syntax: proceedings of the colloquium of the 'Indogermanische Gesellschaft'*, 27-45. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.19.05lig>
- Lowe, John y Mazaudon, Martine (1994): «The reconstruction engine: A computer implementation of the comparative method», *Computational Linguistics* 20(3). 381–417.
- Luick, K. (1914-40): *Historische Grammatik der englischen Sprachen*, Leipzig: Tauchnitz.
- Matasović, Ranko (2009): *Etymological Dictionary of Proto-Celtic* (Leiden Indo-European Etymological Dictionary Series; 9), Leiden, Brill.
- Metropolis, Nicholas (1987): «Beginning of the Monte Carlo method», *Los Alamos Science Special Issue*, pp. 125–130. En línea: <https://fas.org/sgp/othergov/doc/lanl/pubs/00326866.pdf> [20-05-2023]
- Meyer, Fanny, Perrier, Victor: (2023): *esquisse: Explore and Visualize Your Data Interactively*. <https://dreamrs.github.io/esquisse/>, <https://github.com/dreamRs/esquisse>.
- Muñoz-Acebes, J. (2019): «De la glotocronología a la filogenética: Estado de la cuestión y los nuevos desarrollos en la metodología de clasificación lingüística», *Revista de Investigación Lingüística*, 21, 170–184. <https://doi.org/10.6018/rii.21.367611>
- Muñoz-Acebes, J. (2022): «Cladística e Historia de la Lengua Alemana: introducción al uso de PAUP*», *Revista De Humanidades Digitales*, 7, 40–56.

- <https://doi.org/10.5944/rhd.vol.7.2022.33655>
- Ni, Tianyi (2020): «Markov Chain Monte-Carlo Phylogenetic Inference Construction in Computational Historical Linguistics», *arXiv preprint arXiv: 2002.09637*. <https://doi.org/10.48550/arXiv.2002.09637>
- Oakes, M. P. (2000): «Computer estimation of vocabulary in a protolanguage from word lists in four daughter languages», *Journal of Quantitative Linguistics* 7(3), pp. 233–243.
- Pulgram, Ernst (1959): «Proto-indoeuropean reality and reconstruction», en *Language*, 35(3), pp. 421-426.
- Ringe, Don (1992): «On calculating the factor of chance in language comparison», *Transactions of the American Philosophical Society* 82(1), pp. 1–110.
- Ringe, Don, Warnow, Tandy y Taylor, Ann (2002): «Indo-European and computational cladistics», *Transactions of the Philological Society*, 100(1), pp. 59–129.
- Ringe, Don (2006): *From Proto-Indo-European to Proto-Germanic*, Oxford, OUP.
- Ringe, Donald y Taylor, Ann (2014): *The Development of Old English* (A Linguistic History of English; 2), Oxford: Oxford University Press.
- Robert, Christian y Casella, George, (2010), *Introducing Monte Carlo Methods with R*, New York/Dordrecht/Heidelberg/London, Springer, DOI: 10.1007/978-1-4419-1576-4
- Durie, Mark y Ross, Malcolm (eds.) (1996): *The comparative method reviewed: regularity and irregularity in language change*. New York & Oxford: Oxford University Press.
- Walkden, G. (2009): *The comparative method in syntactic reconstruction*, Cambridge, University of Cambridge, Doctoral dissertation. En línea: http://walkden.space/Walkden_2009_CompMeth.pdf [15-6-2023]