

LENGUAS Y GENES: APLICACIONES A LA PREHISTORIA DE LA LENGUA VASCA¹

Resumen: Partiendo de la crítica de varias aportaciones recientes a la reconstrucción de la situación lingüística y poblacional de Europa y en especial del País Vasco, entre las que destaca por su exhaustividad la de Villar, este trabajo se detiene en cuestiones de carácter metodológico. Se pasa revista a los diferentes modos de transmisión de la información genética y lingüística, con los consiguientes problemas de emparejamiento de ambos tipos de información. Se abordan los problemas inherentes al material toponímico antiguo en cuanto base para el establecimiento de clasificaciones lingüísticas, incidiendo sobre las características especiales que presentan los «cognados toponímicos». Por último, se muestran las debilidades y contradicciones en que caen muchas propuestas de reconstrucción genético-lingüísticas, aplicadas a la prehistoria de las lenguas indoeuropeas y de la lengua vasca en Europa.

Palabras clave: prehistoria lingüística de Europa; lengua vasca; lenguas indoeuropeas; genética de poblaciones; toponimia antigua.

Abstract: This paper deals with recent hypotheses concerning the linguistic and genetic Prehistory of Europe, especially the prehistory of the Basque language, and centres on methodological considerations. The author evaluates the different ways in which genetic and linguistic information is passed on, and the difficulties which arise when matching both types of information. Special attention is paid to ancient place names as a source of data for reconstruction in linguistic prehistory, with a discussion of the methodological bases for «toponymic cognates». The paper ends by identifying the weaknesses and contradictions inherent in some of these hypotheses for the genetic and linguistic Prehistory of Europe and the Basque Country.

Key words: Linguistic Prehistory of Europe; Basque language; Indo-European languages; Genetics; ancient place names.

1. He querido que esta contribución, cuyo alcance quedará limitado a reflexiones generales sin gran aparato documental, tocara de algún modo cuestiones de prehistoria, aunque sea de esta manera tan marginal y tangencial.

La prehistoria de la lengua vasca siempre ha sido un tema que ha suscitado el interés de estudiosos de múltiples disciplinas, aparte de los mismos filólogos. Junto a las aportaciones estrictamente arqueológicas de los prehistoriadores, los genetistas, en particular los especialistas en genética de poblaciones, también han intervenido últimamente con sus propuestas en la configuración de escenarios sobre la situación prehistórica de la lengua vasca y de otras de Europa.

Quizás las aportaciones más llamativas hayan sido aquellas que han intentado unir en una hipótesis armoniosa datos genéticos y lingüísticos con la pretensión de que resultaran un reflejo de procesos

¹ Este trabajo se ha realizado dentro del Proyecto HUM2006-13424-C04-03/FILO financiado por el Ministerio de Educación y Ciencia de España.

prehistóricos realmente ocurridos. Ha habido varias propuestas en este sentido, no todas construidas sobre el mismo tipo de datos ni con el mismo nivel de control ejercido sobre los propios datos. Las tres más notables que han afectado directamente a la lengua vasca son: a) la propuesta por Arnaiz Villena, en unión con Alonso, sobre el origen norteafricano, en concreto, beréber del pueblo vasco y su lengua, b) la hipótesis de Th. Vennemann sobre el origen vascónico de la toponimia llamada «europea antigua» de gran parte de Europa Occidental, cuya creación se debe a la expansión de la población vasca por Europa tras el último periodo glacial a partir de su refugio climático pirenaico y peninsular, y c) la de F. Villar, que adscribe la toponimia «europea antigua» al indoeuropeo primitivo, proponiendo una expansión muy antigua, con seguridad en época neolítica, sin descartar que hubiera habido indoeuropeización incluso anterior, de época mesolítica, que a su vez fuera expandida secundariamente por el mismo proceso de colonización de tierras vírgenes tras la glaciación, que he mencionado antes; en este escenario, la lengua vasca pierde el protagonismo que normalmente ha tenido en muchas propuestas lingüístico-genéticas anteriormente.

Ninguna de las tres propuestas posee una contrapartida arqueológica elaborada, al modo en que la hipótesis de C. Renfrew pretendía armonizar el escenario general del movimiento demográfico debido a la neolitización con culturas materiales perceptibles en el registro arqueológico y con la distribución de genes en la actualidad, en concreto, el llamado Componente Principal 1 de Cavalli-Sforza *et al.* (1994, 292, fig. 5.11.1)

2. La propuesta de Arnaiz & Alonso ha sido muy criticada por los dos lados de la ecuación. En lo que toca a su lado genético, está basada en una metodología, que no solo no tiene en cuenta los modernos estudios exhaustivos y secuenciales sobre el ADN mitocondrial y el cromosoma Y, sino que incluso se aparta de la mayoría de los estudios genéticos de su momento en la elección, tratamiento e interpretación de los marcadores genéticos. En concreto, según sus conclusiones, la población vasca muestra claros componentes de genes de origen norteafricano, haciendo del estrecho de Gibraltar el lugar de paso para el flujo de estos genes a la Península Ibérica. Este panorama ha sido expresamente negado por múltiple y variada investigación coetánea y posterior (véase Bosch, E. *et al.* 2001). Igualmente, las propuestas de equiparación entre el vasco y el beréber no tienen ningún apoyo lingüístico, y mucho menos las fantasiosas y totalmente acientíficas interpretaciones de los textos ibéricos y tartesios a partir de la lengua vasca, llevados a cabo por su colega Alonso (cf. De Hoz 1999).

La propuesta de Th. Vennemann (1994) está basada primordialmente sobre material toponímico de Europa, tanto el atestiguado en fuentes antiguas como el existente actualmente, que constituye la base del catálogo de topónimos que desde H. Krahe se conoce como toponimia europea antigua, «alteuropäische» o paleo-europea. Mientras que Krahe pensaba que esta capa toponímica pertenecía a una fase específica del indoeuropeo, anterior a su desmembración in situ en las lenguas históricas conocidas, Vennemann en atención a ciertos criterios formales y tipológicos cree que esta toponimia se entiende mejor si es adscrita a la antecesora de la lengua vasca. Para ello, establece unos análisis de morfemas, un inventario de fonemas y una identificación del último elemento vocálico *-a* de la serie con el artículo definido vasco en *-a*, que en su opinión se adecuan bien a la estructura morfológica y tipológica del vasco o proto-vasco. Por el lado histórico, el proceso que explicaría esta distribución europea de los topónimos sería la expansión de población vascófona desde su refugio climático peninsular hacia el norte de Europa a medida que la cota de hielo va retrocediendo tras el último periodo glacial. Sería, por tanto, una colonización de tierras vírgenes y estaríamos ante un proceso de adjudicación primigenia de nombres a los ríos y accidentes geográficos. El reflejo genético de esta expansión demográfica se encontraría, según algunos genetistas como A. Torroni *et al.* (1998), en la

distribución en gradiente Sur-Norte de algunos Haplogrupos, especialmente el Hg V del ADNmt, que surgió en la península ibérica hace unos 13.000-12.200 años, y otros como el H3 y H1 del ADNmt.

No es el momento para volver a revisar en detalle ni los datos ni los argumentos utilizados por Vennemann en defensa de su hipótesis; ya ha sido hecho previamente tanto por mí como por J. Lakarra en trabajos anteriores². Observábamos en nuestra crítica que el establecimiento del inventario fonológico y de la distribución de los fonemas no se adecuaba a lo documentado ni a lo reconstruido por la reconstrucción estándar y más aceptada, que el análisis de los morfemas era muy particular y no se compadecía con el más común y habitual, que la interpretación del elemento final *-a* como artículo no se sostenía desde los datos más básicos de la protohistoria vasca.

Su hipótesis mostraba también dos características: a) no se preocupaba por explicar la toponimia del propio País Vasco, ni la antigua ni la moderna; b) las etimologías protovascas resultaban de las características fonotácticas presentes en los topónimos paleo-europeos, sin que resultaran confirmadas independientemente por los métodos de reconstrucción internos de la lengua vasca. Como colofón, resultaba que esta nueva visión del «vascónico» no aportaba nada a una mejor comprensión de la prehistoria de la propia lengua vasca, limitándose a dar una explicación alternativa, en ocasiones de peor calidad, a términos ya antes explicados³.

3. Recientemente F. Villar (Villar & Prósper 2005) ha dedicado un largo trabajo al estudio comparado de material genético y lingüístico, centrado exclusivamente en la toponimia antigua, en orden al esclarecimiento de la Prehistoria lingüística de Europa y de la posición de la lengua vasca en el marco general. Al igual que Arnaiz concede un gran peso al aporte de la genética de poblaciones en la discusión, aunque le separan de él puntos importantes: critica el emparejamiento mecánico de lenguas y genes; no utiliza material procedente de ninguna lengua viva, sino material toponímico antiguo; se basa en datos y metodologías, tanto genéticos como especialmente lingüísticos, de mayor aceptación entre los respectivos especialistas. Al igual que Vennemann tiene como banco de pruebas principal el conjunto de topónimos paleo-europeos, aunque disiente radicalmente en la adscripción lingüística del conjunto, que considera indoeuropeo.

Resumiré a continuación las conclusiones que obtengo de su trabajo, que a veces no quedan más que apuntadas o sugeridas, sobre todo en lo que concierne a la relación entre datos genéticos y lingüísticos.

- a. en lo que concierne a la distribución de genes, la población de Europa posee de modo mayoritario y también bastante uniforme, con diferencias menores a nivel regional, un conjunto de genes que remonta a época paleolítica (80% del total) y otro 20 % que se debe al aporte de gentes relacionadas con la neolitización, estos últimos distribuidos en gradiente decreciente Este-Oeste.
- b. los genes de la población vasca no son diferentes de los del resto de los europeos, y lo son menos que los de otras poblaciones como los Saami⁴.

² Gorrochategui & Lakarra 2001; Lakarra 1996 [1999]

³ He comprobado que en su reciente obra (Vennemann 2004) no hace mención del *Vasconic* como sustrato lingüístico de las islas Británicas —que explicaría un fenómeno como la inexistencia de poseedores externos tanto en celta insular antiguo como en inglés moderno—, sino solo del semítico.

⁴ Véase en este sentido también, Simoni *et al.*, 2000, para quienes, aparte de la clara particularidad de los Saami, la distribución del ADNmt y nuclear no coincide en el resto de Europa; se dan mayores diferencias en la zona mediterránea que en la central y septentrional, de modo que esta distribución no cuadra bien, en su opinión, con una expansión a partir de refugios glaciales en el mesolítico.

- c. los genes de los vascos han merecido una atención preferente debido a la singularidad lingüística, pero en el fondo no presentan una particularidad mayor que la de zonas vecinas, p. ej. la de los cántabros.
- d. la expansión démica producida en el periodo de deshielo que siguió al Dryas reciente surgió de varios refugios climáticos: el del Suroeste de Europa no se puede equiparar a población antecesora de la vasca.
- e. muchos de los topónimos antiguos pueden clasificarse en dos patrones distribucionales: el patrón A, de etimología indoeuropea, tiene su foco de irradiación en el Oriente Próximo y llega hasta la Península Ibérica, donde presenta sus mayores porcentajes en Andalucía. El patrón B, cuyos topónimos mayoritariamente pueden explicarse también como indoeuropeos, tienen su foco de irradiación en la Península Ibérica, extendiéndose por Europa, alcanzando la zona báltica.
- f. los topónimos del patrón A deben su razón de ser a la expansión démica del neolítico, porque coinciden con los mapas distribuciones de los genes atribuibles a este proceso (en ello coincide parcialmente con Renfrew); los del patrón B, sugiere, se deberán a la expansión postglacial desde Iberia hacia el Norte: ello trae como consecuencia, no expresada explícitamente, la presencia de indoeuropeos en la Península Ibérica en pleno periodo mesolítico, hace 13 mil o 12 mil años.
- g. la toponimia antigua del País Vasco (entiéndase en el sentido restringido de Comunidad Autónoma del País Vasco o Euskadi) y Navarra (española) no muestra casi ningún rasgo euskérico, sino más bien abundantes índices de indoeuropeísmo arcaico, anterior a la presencia céltica histórica; muchos de los topónimos se adscriben a uno u otro patrón distribucional. Consiguientemente, la lengua vasca no fue una lengua peninsular ni autóctona en el País Vasco; probablemente se introdujo en su territorio histórico en época tardoantigua desde el Norte de los Pirineos, desde Aquitania.

4. Hay en este trabajo de Villar, junto al tratamiento de numerosos datos de índole genética y toponímica, reflexiones de naturaleza metodológica de carácter general que son también de interés. Consiguientemente, en lo que sigue incidiré especialmente en este tipo de cuestiones, sin desdeñar el comentario de detalle, pero limitándolo a unos cuantos casos.

4.1. La primera cuestión que no hay que olvidar es que la transmisión de los genes y de las lenguas sigue pautas diferentes, de suerte que el emparejamiento de ambas realidades es un fenómeno más bien extraordinario que habitual; a pesar de ello, los genetistas suelen tender inexorablemente al emparejamiento de ambos ámbitos, a veces de un modo excesivamente simplista. Tienden a unir las particularidades regionales que hallan en el marco general con particularidades étnicas y lingüísticas actuales o recientes⁵. En este sentido el pueblo vasco, poseedor de una lengua genéticamente aislada en el concierto de las lenguas, ofrece señaladas oportunidades para convertirse en el anclaje histórico de ciertas particularidades genéticas regionales.

⁵ «As Zegura points out, the geneticist's usual null hypothesis is that linguistic and human genetic development will be parallel, whereas the linguists' null hypothesis is usually that they will not match» (Campbell [en prensa]). Cavalli-Sforza *et al.* 1994, 99: «The one-to-one

correspondence between genetic clusters and linguistic families is remarkably high, but it is not perfect»; véase fig. 2.6.2 donde se emparejan árboles genéticos y familias de lenguas.

Pero, los lingüistas sabemos que la suplantación de lenguas, es decir el abandono de una lengua por otra en el seno de una comunidad, es un fenómeno relativamente frecuente, tal como nos muestra la gran cantidad de lenguas muertas de las que tenemos noticia en el corto espacio de la historia humana conocida. Para poner un ejemplo cercano, sabemos que en los dos últimos milenios en un lugar como Zaragoza (la antigua *Salduia*) se ha cambiado de lengua cuatro veces (ibérico > latín > árabe > aragonés > castellano)⁶. Por otro lado, el conocimiento actual de las comunidades de hablantes nos enseña que la situación monolingüe total es un fenómeno minoritario, siendo mucho más normal los estados de contacto de lengua en cualquiera de sus múltiples variantes: desde situaciones en que una de las lenguas puede considerarse de la comunidad mientras que la otra sea externa (y con funciones variadas), hasta situaciones en que anden en juego de manera normal y estable hasta diez lenguas, como en los casos de exogamia lingüística del área del Vaupés (Colombia-Brasil). En estos casos, ¿cómo es posible equiparar linajes genéticos y linajes lingüísticos?

Villar es perfectamente consciente de este problema y critica en varias ocasiones la debilidad de este planteamiento: la ocasión más señalada es su rechazo de las conclusiones a las que llegan McEvoy *et al.* 2004 en su estudio comparativo de las poblaciones de Irlanda y el País Vasco, cuando asignan a toda la población de las islas y de la fachada atlántica continental, incluido el País Vasco, un linaje celta (p. 395-6). Igualmente rechazables son los resultados a que han llegado otros genetistas, como S. Oppenheimer, que basándose en las mismas relaciones genéticas propone, inversamente, una ancestral población vasca de las Islas Británicas, muy poco modificada posteriormente por aportes celtas (éstos también vía meridional) y germánicos⁷.

4.2. En muchísimas ocasiones los genetistas, para sus emparejamientos lingüísticos, se basan en clasificaciones muy problemáticas de lenguas y, por lo general, poco aceptadas por los especialistas. En este sentido, está muy extendida la utilización de propuestas enormemente reduccionistas, como las de J. Greenberg o M. Ruhlen, para sus comparaciones. Como hay una pulsión innata al emparejamiento de datos y mapas genéticos con los lingüísticos, y como por otro lado los datos genéticos adquieren la mayoría de las veces cierto sentido solo si se los observa a escala continental, resulta de ello que las clasificaciones lingüísticas reduccionistas, que tienden también a clasificaciones continentales, proporcionan una buena contrapartida. Pero, ninguno de los especialistas en las diversas lenguas de América acepta la clasificación de Greenberg, ya que es imposible reducir por los métodos

⁶ Convengo en que la lengua hablada en *Salduia* fuera el ibérico, a partir de la documentación onomástica del Bronce de Ascoli y otros testimonios epigráficos cercanos. Aquí no se tienen en cuenta los estados de bilingüismo que han debido existir en la mayor parte del periodo señalado.

⁷ Opiniones publicadas en su libro, *The Origins of the British: A Genetic Detective Story*, 2006, obtenidas a partir de su divulgación por Internet. Uno de los valedores de la interdisciplinariedad entre genética, lenguas y arqueología como C. Renfrew es consciente de la dificultad de las equiparaciones simplistas: «The foregoing discussions should have made clear not only that there are areas where the three disciplines of archaeology, historical linguistics and molecular genetics intersect, but that

there are also very considerable methodological difficulties in relating and reconciling them. For that reason suggestions of equations and equivalencies should be regarded with caution. ... It is, of course, perfectly clear that there is no direct correlation between specific languages and specific genes. As we have seen, language replacement can occur in a population with only very minor genetic effects precisely because a language can be learnt and need not be inherited. Such may often be the case when language replacement occurs through elite dominance (although we might expect the incoming elite to leave some molecular genetic trace, if only a minor one). And such may also be the case when contact-induced language change occurs (Renfrew, 2002, 69)

comparativos estándares las numerosas lenguas y familias lingüísticas de toda América a tres troncos, de los cuales el amerindio es con mucho el más rico⁸.

También los lingüistas, muchas veces de manera inconsciente, se ven atrapados en estas equiparaciones genéticas realizadas sin el suficiente control:

«Esa convicción [sobre la conexión entre genes y lenguas] tiene, sin embargo, un punto de partida lógico bien fundado, cuya constatación empírica es en parte posible. En efecto, si examinamos un mapa genético del mundo, tanto del ADNmt como del cromosoma Y, veremos que en líneas muy generales y a escala mundial hay una cierta correspondencia entre lenguas y genes. Por ejemplo, entre los hablantes de lenguas amerindias predomina el haplogrupo Q del cromosoma Y y los Hg A, B, C del ADNmt; o C+ K y N respectivamente entre los hablantes de las lenguas aborígenes australianas» (Villar, p. 380-1).

Acabamos de decir que no existe una familia lingüística amerindia: el predominio del Hg Q del cromosoma Y en América se deberá explicar mediante recursos estrictamente genéticos, pero no puede emparejarse con una supuesta familia de lenguas concreta. En Australia ocurre lo mismo: aunque la familia Pama Nyunga es muy extendida⁹, existen muchas otras familias menores, así como lenguas aisladas que no pueden reducirse a un australiano aborígen común. En ambos casos son fenómenos continentales. Parece que lo razonable en estos casos es aplicar también la máxima de que el emparejamiento genes-lenguas no es la hipótesis cero. Dejo apuntada ahora la idea, que trataré en un punto más abajo, de que el emparejamiento sería asumible como hipótesis cero, sin embargo, solamente en el caso de poblamiento fundacional.

4.3. Junto a la búsqueda de la familia lingüística apropiada por parte del genetista, existe también análogamente la búsqueda del mapa genético apropiado o preferido por parte del lingüista. El trabajo de Villar nos ofrece un ejemplo de este proceder. En una primera fase del trabajo, Villar identifica una serie de topónimos que, en virtud de su clara etimología indoeuropea y su área de distribución territorial, han tenido su origen en el foco de mayor densidad de atestigüaciones: en el Próximo Oriente. Se trata de los topónimos de patrón distribucional A, que presentan una gran densidad en Anatolia con una difusión decreciente hacia Occidente, donde hay ciertos picos en Italia y sur de la Península Ibérica. Este mapa recuerda rápidamente a los mapas genéticos en gradiente Este-Oeste, como es el del Componente Principal I de Cavalli-Sforza y el de distribución de determinados haplogrupos del cromosoma Y, como el E y J (cita a McEvoy *et al.*, 2004). La cuestión es, sin embargo, que un mapa sintético de todos ellos coloca el epicentro de la difusión demica neolítica, responsable de la expansión de estos haplogrupos, en Palestina, lugar, como todo el mundo sabe, de lenguas no indoeuropeas, sino semíticas. En cambio, la distribución de una variedad concreta, la del Hg J-M67,

⁸ He aquí la conclusión a la que llegan los especialistas en lenguas amerindias sobre la clasificación de Greenberg y su utilización por los genetistas (Bolnick *et al.* 2004):

«In future studies comparing genetic and linguistic variation in the Americas, we recommend use of the consensus linguistic classification, as given in Campbell (1997), Goddard (1996), and Mithun (1999), rather than Greenberg's tripartite classification (Greenberg *et al.* 1986; Greenberg 1987). In addition, since there is no

legitimate reason to believe that «Amerind» is a unified group (linguistic or otherwise), it has been essentially abandoned in linguistics and should not be used in genetic analyses.»

⁹ Dixon incluso considera la familia Pama Nyunga como resultado de procesos de convergencia a partir de lenguas en origen diferentes antes que una familia lingüística al uso, surgida por especiación a partir de un ancestro común.

cuyo mapa ofrecen Semino *et al.* 2004, 1026 «se ciñe de manera más estricta al patrón distribucional de nuestros topónimos de tipo A en Europa. Y al igual que nuestros topónimos, su presencia se detecta también en Asia, alcanzando la India, donde penetraron junto con la agricultura neolítica. La hipótesis de mayor parsimonia para explicar las coincidencias filogeográficas de estos dos fenómenos de naturaleza tan dispar es considerar que ambos (genes y topónimos) fueron introducidos en Europa desde Anatolia y Oriente Próximo con la difusión démica del neolítico» (p. 140-1).

El problema está en que tan neolíticos y tan europeos son las otras variedades del Hg J (como alguna del Hg E, como E-M78), de modo que estamos eligiendo una única variedad solamente en virtud de la distribución territorial que presentan las lenguas indoeuropeas anatólicas en épocas históricas: estamos haciendo una equiparación de mapa genético y mapa lingüístico (prefiriendo Anatolia a Palestina), aunque en este caso se trate de lenguas muertas y no actuales, haciendo caso omiso de lo que nos enseña la lingüística sobre la improbabilidad de los emparejamientos (supra § 4.1)¹⁰.

4.4. En la extensa cita del punto anterior Villar considera, sin embargo, que la «hipótesis de mayor parsimonia» es la historia paralela de genes y topónimos, en contradicción con lo que él mismo ha afirmado en varios lugares. Entonces, ¿acaso está permitido el emparejamiento en circunstancias especiales, como por ejemplo, la colonización inicial y fundacional de un territorio? Ésta parece ser la idea de Villar. Pudiera ser, pero nada nos asegura que los colonizadores neolíticos de Europa fueran hablantes de una sola lengua o de una familia muy cercana de lenguas; en vez de ser la hipótesis más normal, sería incluso la más difícil de admitir para un lingüista que sabe que las comunidades humanas pueden presentar múltiples situaciones sociolingüísticas (ver § 4.1).

Si la neolitización de Europa se debe a una difusión démica procedente de un foco único en Oriente Próximo (sin aporte de otros focos secundarios independientes, cuestión muy debatida), existen varios problemas: a) la zona originaria es un cruce geográfico entre África, Asia y Europa, muy abierta a todo tipo de influencias culturales y de población; históricamente es patria de lenguas de familias diversas, y no hay por qué pensar que no lo fuera en el inicio del neolítico; b) Europa no estaba vacía; esto nos lo dicen ahora los mismos genetistas. De modo que hay que pensar en procesos normales de sustitución de lenguas y no en procesos «extraordinarios» de carácter fundacional; c) el proceso no es como la colonización de una isla despoblada por una exigua población uniforme, sino la de un continente abierto con población existente y con aporte seguramente multilingüístico. En conclusión, hay que abogar por el principio del unitarismo científico, según el cual debemos suponer para épocas prehistóricas el mismo comportamiento en los procesos de evolución, contacto y difusión de lenguas que observamos en los periodos documentados históricamente.

Existe otro proceso histórico que presenta cierta analogía con el de la neolitización, cuyo carácter fundacional «ex novo» se me antoja incluso más nítido: me refiero al de la recolonización de Europa desde los refugios glaciales meridionales a partir del 10.000 antes de ahora. Como sabemos, esta es

¹⁰ Independientemente de la crítica expuesta, que es de naturaleza general y metodológica, puede hacerse otra concreta relativa a los datos concernidos: en el mapa de distribución de la variante Hg J-M67 el centro se coloca en Anatolia septentrional; si comprobamos la tabla donde Semino *et al.* 2004, 1029, ofrecen los porcentajes de esta variedad entre la población de la muestra, los resultados son: 13,3% entre los georgianos, 9,6% entre italianos del Norte, 8,9% entre cala-

breses, 6,30% entre balcarianos del Cáucaso meridional, 4,9% entre judíos askenasim, 4,5% entre iraquíes, 4,3% entre griegos, 3,6% entre catalanes y 3,2% entre andaluces; no hay constancia entre vascos (ni peninsulares ni continentales), ni entre bearneses. En la zona del Próximo Oriente, ninguna población de la muestra es indoeuropea ni está ubicada en zona supuestamente indoeuropea antigua.

la hipótesis preferida de Vennemann para la vasconización de Europa. Veamos cómo la critica Villar, de modo admirable:

«La cuarta y última hipótesis (que los repobladores paleolíticos hablaran paleoeuskera) es simplemente gratuita. Suponer que en todos los lugares de partida de aquella repoblación se hablara una misma lengua (sea ésa euskérica o de cualquier otra filiación) no es más que una conjetura arbitraria. De la misma manera en que cada punto de partida pudo aportar diferentes variedades genéticas a la repoblación es posible que aportara también distintas lenguas o variedades lingüísticas» (p. 413-4)

Estoy totalmente de acuerdo con esta apreciación. Simplemente entiendo que su validez no queda limitada a la repoblación mesolítica, sino que sirve también, con más razón si cabe, para la difusión neolítica.

5. Toponimia vs. Lenguas. Uno de los supuestos metodológicos de Villar consiste en que la toponimia, por tratarse de un material especialmente refractario al cambio o sustitución, mucho más que las lenguas en su conjunto, podría ser reflejo más fiel de un poblamiento antiguo y de este modo tener más legitimidad para el emparejamiento con los mapas genéticos¹¹.

5.1. Hay que señalar, antes que nada, que una toponimia, para que tenga valor lingüístico, debe ser adjudicada a una lengua concreta, catalogada desde un punto de vista clasificatorio, de modo que funcione como el sucedáneo de esa lengua extinta, como su representante en el momento presente. Esta clasificación lingüística tiene sus problemas, menores y de poca envergadura en los casos en que la toponimia sea transparente para los hablantes o estudiosos de alguna lengua bien conocida, mucho mayores en casos de lenguas extintas, atestiguación toponímica escasa y mal transmitida y deficiente información histórica complementaria.

No cabe duda de que la toponimia proporciona información de gran valor para conocer los avatares (pre-)históricos de un territorio dado. Pero su alcance queda limitado a la cronología de la propia toponimia. Si en un territorio dado, donde se habla la lengua A, se documenta toponimia perteneciente a la lengua B, podremos sustituir el mapa obtenido de la lengua A por el mapa proporcionado por la toponimia de B, pero estaremos como al principio en lo que atañe al emparejamiento mecánico entre mapas genéticos y lingüísticos, solo que habremos retrotraído la equiparación unos cuantos siglos (o algún milenio en los casos favorables). Tiene gran interés para cuestiones históricas y protohistóricas que caen dentro del alcance cronológico de la sustitución lingüística, pero mucho menor para dilucidar problemas de gran antigüedad ocurridos hace 10 milenios o más.

La toponimia, por tanto, no ofrece información de índole diferente de la que ofrece una lengua viva, solamente actúa en representación de una lengua determinada que ha desaparecido en los últimos periodos de la historia de un territorio dado.

¹¹ Ya hemos visto su emparejamiento de los topónimos del patrón A con el mapa genético del Hg J-M67 del cromosoma Y. Lo formula expresamente en los preliminares de su capítulo sobre la toponimia vasca: «la solución del tan debatido y crucificado problema de la indoeuropeización de Europa, uno de cuyos flecos es la cronología de la presencia del euskera en la zona no puede basarse correlacionando genes con lenguas actualmente habladas

sino *correlacionando genes y conjuntos toponímicos* pertenecientes a lenguas distintas de las actuales, en el caso de que los haya. Pero en Europa, cuya historia está plagada de episodios de élites dominantes, *es previsible que la correlación entre conjuntos toponímicos y genes sea superior a la de genes y lenguas actualmente habladas*» (p. 427) [subrayado original del autor]

5.2 Ahora bien, lo que la toponimia gana en penetración prehistórica pierde en seguridad lingüística¹².

A medida que nos alejamos en el tiempo, nuestro conocimiento de las lenguas antiguas decrece. Aun así, en algunos casos concretos somos afortunados: no solo contamos con una información suficiente de las lenguas habladas en un territorio, sino que tenemos una idea correcta de sus estructuras, aunque por lo general pobre. Si la lengua en concreto pertenece a una familia lingüística conocida, como es la indoeuropea, contaremos con una ayuda complementaria de gran valor. Este es el panorama aproximado de Europa, especialmente de la zona mediterránea occidental, entre el 500 a.C y el 500 d.C. Los lingüistas tienen instrumentos para identificar correctamente muchos topónimos célticos, en sus variedades gala, celtibérica, britónica, etc., así como ibéricos, etruscos, itálicos, griegos, godos. Pero existen centenares de ellos, para los que no hay más que una vaga sugerencia de explicación, siendo muy frecuente también la existencia de vías alternativas no concluyentes, cuando no la total oscuridad.

Muchos de estos topónimos poco claros se documentan en el territorio de los topónimos claros y coherentes en su distribución territorial con las lenguas históricas conocidas: constituyen, por tanto, los testimonios del estrato lingüístico anterior. El problema está en la identificación de ese estrato, tanto en su vertiente lingüística como geográfica. Para el primer cometido las dificultades son muy grandes, ya que carecemos por completo de información sobre las lenguas de ese estrato; solo nos queda la lingüística comparada para darnos luz sobre la clasificación familiar de los topónimos: podemos llegar así a aceptar la indoeuropeidad de ciertos nombres de lugar, pero no podemos adscribirlos a una lengua concreta o conocida. Si contáramos con series nutridas de topónimos, variantes sobre las mismas raíces, y formas con ciertas particularidades fonéticas o morfológicas, delimitadas además coherentemente a un territorio concreto, podríamos estar legitimados a pensar que nos hallamos ante una «área toponomástica», que sería el pálido reflejo de la existencia de una lengua anterior. Estaríamos, con todo, en un horizonte cronológico limitado aún, como sumo, a mediados o fines del II milenio a. C. Queda mucho trecho todavía para poder identificar los supuestos mapas de estas «áreas toponomásticas» con hechos acaecidos hace 12.000-10.000 años.

5.3. Toponimia antigua vs. toponimia moderna. Relacionado con lo anterior está la cuestión de la utilización de un tipo u otro de material toponímico. Es obvio que la toponimia antigua nos proporciona una ventaja para saber sobre la situación lingüística de la antigüedad. La toponimia moderna tiene la ventaja, por regla general, de una nutrida representación documental y amplia distribución en el territorio, con lo que se pueden perfilar mejor los detalles lingüísticos y geográficos. Pero se le achaca, con razón, que ha podido sufrir las contaminaciones inherentes al paso del tiempo: desfiguraciones, etimología popular, pérdida natural, extensiones secundarias, etc. Por eso la toponimia antigua es más fidedigna para saber cómo eran los topónimos *contemporáneos* a la redacción de las fuentes antiguas¹³. Pero si los queremos utilizar como fósiles de un estrato lingüístico anterior

¹² Esta es una consecuencia que se observa en toda disciplina histórica. También en la genética, el conocimiento del ADN antiguo, obtenido de los restos orgánicos de los hallazgos arqueológicos, es de un grandísimo interés para compararlo con el obtenido del análisis de la población actual, pero los problemas técnicos en su obtención, tratamiento, falta de contaminación y fragmentariedad son también de gran calado.

¹³ No hay que olvidar, sin embargo, un hecho fundamental que arroja una gran sombra sobre la toponimia antigua: el peligro de corrupción del nombre con motivo de los errores de copia en la transmisión textual. Véase Gorrochategui, 2001, 105-7; 2006a, 113-8.

se nos convierten inmediata y automáticamente en «modernos» con respecto a ese objetivo, con las mismas limitaciones que acabamos de indicar para la toponimia moderna actual. Solo que se encuentran en una situación de gran desventaja frente a la toponimia actual: en este caso podemos detectar los problemas de etimología popular, de extensión secundaria, variación dialectal u otros que le hayan ocurrido al topónimo, porque nuestro conocimiento de las lenguas y de la historia externa es suficientemente rico. En el caso de la toponimia antigua, debemos pensar, en aplicación otra vez del principio del unitarismo científico, que todos esos peligros han acechado de modo similar sobre los nombres de lugar antiguos, aunque no dispongamos de los medios para detectarlos.

5.4. Toponimia vs. otra Onomástica. Al hilo de lo comentado en el apartado anterior, se ha contrapuesto también la diferente capacidad explicativa para dilucidar la prehistoria que presentan la toponimia, por un lado, y las onomásticas personales por otro. Estoy de acuerdo en la mayor virtualidad de la toponimia para dilucidar fases o estratos anteriores, siempre que esa toponimia pueda ser adscrita a una lengua concreta y definida o, al menos, a un «área toponomástica» como ha sido descrita en el punto anterior. Ello es así, porque en principio la toponimia tiene una mayor vocación de pervivencia que la antroponimia y la teonimia. Ahora bien, una mayor vocación no quiere decir una preeminencia absoluta y mecánica. Estimo que los dos tipos de información deben conjugarse en una interpretación histórica coherente con otros testimonios conocidos. Así, p.ej., la capital de los Convenae, pueblo aquitano localizado en el curso pirenaico del Garona y alrededores, era *Lugdunum*, nombre de nítida prosapia gala, mientras que la onomástica personal más firmemente documentada en la zona es la aquitana de ascendencia éuskara. En este caso, gracias a nuestro conocimiento de la historia, podemos corregir el dictado que nos aboca a la preeminencia de la toponimia y pensar que se trata de una nominación hecha a instancia de los romanos, que quisieron tener en Aquitania un lugar que emulara a la *Lugdunum* gala, sede del culto imperial en la Provincia.

Se dice que la antroponimia está sujeta a moda, mientras que la toponimia lo está al terruño. Siendo esto verdad en líneas generales, tampoco hay que llevar las cosas a un extremo irreal: hay topónimos viajeros, de la misma manera que hay antroponimia pertinaz. Villar gusta de citar ejemplos de toponimia árabe en la Península, como ejemplo de mantenimiento de nombres de lugar después de que la lengua árabe haya desaparecido de la Península hace más de 500 años. Y es buen ejemplo, gráfico y pertinente. Pero, no se me negará que tanta o mayor vocación de pervivencia han mostrado entre nosotros los antropónimos godos; nombres como Fernando, Álvaro, Rodrigo, etc. existen a centenares y han existido, con más o menos altibajos, a lo largo de toda la historia de nuestros romances, más de 1.200 años después de la desaparición de los godos en la península. Su tasa de eficacia en la transmisión es infinitamente mayor que la de los topónimos godos, como Frómista o Toro.

Pensar que la toponimia, en virtud de su mayor vocación de pervivencia, no está sujeta a la mortandad natural por el paso del tiempo y a los fenómenos de sustitución por cambio de lengua, de modo que pueda salvarse de la erosión total y llegar hasta el estrato fundacional, es exagerar esa capacidad de permanencia y admitir de manera gratuita que no ha habido el suficiente número de sustituciones lingüísticas como para eliminar toda presencia razonablemente comprobable de toponimia fundacional.

5.5. La lingüística comparada. La lingüística comparada es un instrumento eficaz, quizá el más eficaz de los que disponemos, para arrojar luz sobre la prehistoria no solo de las lenguas, sino incluso de los pueblos. Una de las nociones centrales de la disciplina es la de *cognado*, es decir, términos de

lenguas diferentes relacionados entre sí por razón de su mismo origen, por sus relaciones de homología. El establecimiento de cognados entre el diferente material lingüístico con que se enfrenta el especialista es la labor más delicada y trascendental de su quehacer, en el que tiene que extremar todos los cuidados para asegurar la bondad de su propuesta, ya que de ese acto pionero y fundamental depende todo el edificio interpretativo que construirá a continuación. En el establecimiento de cognados hay varios niveles de actuación, unos iniciales y provisionales, en los que se pretende que las correspondencias fonéticas sean estrictas y las correspondencias semánticas idénticas; una vez asegurado el parentesco lingüístico, se pueden admitir como cognados términos que presenten una perfecta correspondencia formal, aunque carezcan de equivalencia semántica¹⁴. Cuanto más seguro sea el parentesco más podremos arriesgarnos en los límites de la comparación.

De lo anterior se deduce que el cognado debe cumplir con las dos caras del signo lingüístico *a la vez*, con las correspondencias fonéticas regulares y con la equivalencia semántica. Si falla una de las dos caras, la semántica, el nivel de control sobre la bondad de su naturaleza como cognado decrece en picado. Existen infinidad de ejemplos sobre parecidos externos entre palabras de lenguas diferentes. Incluso la tasa de «cognados» con equivalencia formal y semántica debidos al puro azar es relativamente elevada como muestran diversos estudios¹⁵. Bien, el problema con la toponimia y con toda onomástica en general es que carecen de «sentido», a excepción de unos pocos nombres motivados y, por tanto, transparentes en la lengua de creación: «Villanueva» entre nombres de lugar y «Mozo» entre los de persona, por ejemplo. Por esta razón, si unos topónimos no pueden ser adscritos a una lengua conocida con seguridad, porque identificamos en sus constituyentes elementos nítidos de esa lengua, se convierten, al igual que los gases nobles, en un material tan atractivo como inerte. En esto los topónimos y los antropónimos adolecen de los mismos defectos.

Una posibilidad para darles cierta fuerza consiste en suplir de alguna manera esa carencia por medio de algún elemento externo e independiente al propio nombre, que limite las enormes probabilidades estadísticas de emparejamiento que se dan cuando no concurre la semántica. Uno de estos factores, en mi opinión imprescindible, es la geografía. Puede haber otros, todos complementarios, como por ejemplo la coherencia en los inventarios fonológicos, los procesos de formación de palabras, correlación fehaciente entre nombre y referente, etc. Lo mejor es la existencia no contradictoria de todos ellos en un territorio definible con cierta precisión.

En conclusión, un comparatista tiene como objetivo establecer relaciones genéticas entre lenguas o entre lenguas y conjuntos onomásticos, que no son sino reflejos de lenguas perdidas. Pero su cometido solo lo logra si consigue superar todos los obstáculos puestos a la hipótesis del origen común, si elimina todas las explicaciones alternativas que se ciernen sobre el candidato a cognado: en especial, el del préstamo y el azar.

6. Villar propone en este trabajo una hipótesis de largo alcance sobre la expansión de las lenguas indoeuropeas por Europa y el Oriente Próximo y Medio, en base a una selección de topónimos antiguos cuya distribución coincide con ese amplísimo territorio. Como ya se ha señalado arriba

¹⁴ A veces también se ha dado cierta laxitud desde el lado fonético, ante la «evidencia» de la semejanza semántica y funcional de algunas correspondencias; así, durante decenios muchos indoeuropeístas han admitido como cognados los futuros latino en *-bo-* e irlandés en *-f-*, a pesar de la imposibilidad de la correspondencia fonética.

¹⁵ L. Trask, 1996, 220 ofrece un buen ejemplo de parecidos azarosos entre el hawaiano y el griego clásico. No los tomamos por cognados, porque sabemos por información externa e interna (carencia de correspondencias fonéticas recurrentes) que no son lenguas emparentadas: ¿Qué pasaría si no tuviéramos este conocimiento y solo dispusiéramos de muy poca información?

(§ 3e,f), descubre topónimos que responden a dos patrones distribucionales: el patrón A con una gran densidad en el Próximo Oriente, que actúa como foco de expansión hacia Europa y la India, contiene topónimos de etimología indoeuropea; lo relaciona con la difusión démica debida a la neolitización. El patrón B tiene su foco de irradiación en la Península Ibérica y su área de expansión es Europa atlántica hasta la zona báltica. Hay también topónimos de etimología indoeuropea. Su expansión se debe, probablemente, a la recolonización posterior al periodo glacial.

El principal escollo que hallo en la aceptación de las conclusiones presentadas reside en el procedimiento para el establecimiento de los elementos toponímicos equiparables, de los «cognados toponímicos», por llamarlos de alguna manera. El método es claramente acumulativo, es decir, de acopio de todo tipo de topónimo que en ese inmenso territorio (desde el Finisterre atlántico hasta la India inclusive) presente un parecido externo suficiente, primando a poder ser los hidrónimos, pero incluyendo también otras clases como nombres de ciudad, orónimos y étnicos.

Las explicaciones históricas alternativas a partir de las numerosas lenguas que se hablaban en el mundo antiguo faltan o escasean, limitadas solo a algunos casos hispanos. Eso hace que no se plantee como posibilidad que los topónimos que contienen un elemento *Ars-* (que van desde el nombre ibérico de Sagunto, *Arse*, hasta el topónimo *Arsagalitae* de la India, pasando por los del Norte de África y el nutrido conjunto procedente de Armenia y el imperio persa) puedan deberse a otras muchas razones independientes y relacionadas más bien con la historia de las lenguas locales respectivas.

Ante la posibilidad de una explicación alternativa, que se ensaya en el caso de la toponimia en solar vasco, donde se discute si muchos topónimos de aspecto indoeuropeo pueden ser celtas o más bien indoeuropeos arcaicos, se prefiere sistemáticamente esta última opción. Así se llega sesgadamente a proporciones muy elevadas de topónimos indoeuropeos arcaicos frente a célticos. Para ello, se fuerzan o violentan a veces los datos comparativos, como en el caso del hidrónimo *Deva*, cuya adscripción al celta es irreprochable, objetando que el paso del diptongo IE *-ei- a -e- es también indio e itálico. En el caso del indio es cierto, aunque se trata de un paso bastante tardío (en una secuencia *-ei- > -ai- > -ē-, ya que se da la alternancia morfofonológica entre las dos últimas variantes, y el iranio muestra la fase intermedia -ai-: persa ant. *daiva-*, avéstico *daēuua-*); pero el itálico, como tal grupo, es conservador del diptongo como lo muestra claramente el osco en multitud de ejemplos; solamente el umbro y el latín han monoptongado de modo independiente y en un momento muy tardío. Pensar, por tanto, que *Deva* pueda adscribirse a un indoeuropeo arcaico, que se haya mantenido intacto, es imposible, porque esperaríamos **deiwā*; que haya evolucionado a *dēwā*, como en celta, pero en otra lengua indoeuropea desconocida es gratuito y va en contra del principio de la navaja de Occam, que nos pide que no creemos entes innecesarios para explicar datos que pueden explicarse sencilla y económicamente por lenguas conocidas y que resultan ser, además, el estrato lingüístico más moderno y por tanto más productivo. El peso de la prueba recae en quien quiere demostrar la existencia de una tal lengua indoeuropea concreta y específica, para la cual deberá indicar un territorio y unos rasgos fonéticos y morfológicos individuales¹⁶.

Igualmente son frecuentes los análisis morfológicos interesados, en los que ante la posibilidad de varias interpretaciones se prefiere aquella que redunde en beneficio de la hipótesis: A pesar de que *Nerva* puede ser analizado como raíz *ner-* + suf. *-wo/ā-* sin mayores problemas, con paralelos en otros

¹⁶ No es el momento de discutir en detalle los rasgos de esa posible lengua precéltica italoide supuesta por Villar para Hispania (p. 104 ss.), aunque rápidamente se puede decir que algunos de ellos están basados sobre material no indoeuropeo (*Arse* y *Salduie*) y otros admiten

explicaciones tradicionales perfectamente válidas (como *Siculus*, con *l* velar frente a *Sicilia* con *l* palatal a partir de una forma Sikeliva, sin necesidad de pensar que estamos ante un sufijo *-ulo-*).

puntos de la Céltica, se prefiere pensar que en su parte final existe el apelativo **uba-* «río», identificable así mismo en el topónimo *Iluberis* (obtenido del étnico *Iluberitani* de Plinio), que se analiza a su vez de manera muy particular como *Il + uberi*¹⁷, cuando todos los especialistas en ibérico lo analizan como *Ilu(m/r-)-ber(r)i-*.

Por otro lado, tampoco queda clara en ocasiones la exactitud de la correspondencia fonética. Un ejemplo entre muchos nos lo ofrecen los topónimos formados sobre la base **kard-*, derivada de la raíz **kar-* «duro» mediante sufijo en dental hallado en germánico (gót. *hardus*, ingl. ant. *hard*) y griego, *κάρτα* adv. «fuertemente». La dental originaria, según estos cognados, es **-t-*, conservada en griego y sonorizada en germánico por la ley de Verner, un cambio fonético exclusivo del germánico. Los topónimos aducidos por Villar (p. 128 con mapa) presentan todos ellos una forma *Card-/Cord-*, con *-d-* sonora en todo el extenso territorio desde Persia hasta Hispania. Habría que imaginarse una variedad de lengua indoeuropea, que en el foco originario, es decir en Anatolia, antes de la difusión hubiera sonorizado el grupo *-rtV-* > *-rdV-* y lo hubiera extendido por todo el ámbito de la oecumene, antes de la formación de las lenguas indoeuropeas conocidas que piden una forma original con *-rtV-* conservada. Esa lengua también presentaría una alternancia vocálica *a/o*, ya que en Hispania tenemos *Carduae* y *Corduba*. Por otro lado, los topónimos en *kors-*, *kord-* (mapa en p. 85), atribuidos a la raíz **k'er-*, presentan formas sin satemización en los territorios orientales, de modo que hay que concluir que los iraníes se han asentado en sus territorios históricos después de que se hubiera producido este cambio fonético en su gramática. Es decir, tiene fuertes implicaciones para la historia de los grupos indoeuropeos históricos¹⁸.

Por último, los rasgos fonéticos y gramaticales de este proto-indoeuropeo responsable de estos topónimos tan extendidos como antiguos no quedan expresamente definidos, aunque muestren una apariencia, en mi opinión, moderna. Sería lógico esperar testimonio o indicio de las fases más arcaicas recuperadas por la comparación indoeuropeística tradicional, como la presencia de larinales, un ablaut con representación de sus diferentes grados, presencia de sonantes vocálicas; entre los comportamientos morfológicos, se esperaría la inexistencia del femenino y de las formaciones directamente relacionadas, como los temas en *-ā*, cuya formación en indoeuropeo sabemos que es secundaria.

En conclusión, a este procedimiento de establecimiento de cognados toponímicos se le puede hacer la misma crítica que se suele hacer a las propuestas comparativas denominadas de «mass comparison» o «comparación masiva», las cuales pretenden justificar el parentesco lingüístico de muchas lenguas no relacionadas según los parámetros estándares en base a parecidos superficiales en un limitado número de palabras de vocabulario básico. Igualmente, en el método empleado por Villar se reúnen topónimos (preferentemente hidrónimos que funcionarían como el léxico básico en las comparaciones entre apelativos) en atención a su parecido externo a partir de zonas muy extensas que abarcan más de un continente. No se hace ningún ensayo por eliminar los riesgos de contaminación ni atender a explicaciones particulares en el ámbito de cada una de las diferentes áreas toponímicas

¹⁷ Como paralelos modernos se citan los nombres de lugar *Ubera*; el documentado en Bergara se explica perfectamente a partir del vasco *ubera* «vado», de clara etimología (< *(h)ur «agua» + *be(he)ra* «abajo»)

¹⁸ Una visión que propugna una indoeuropeización tan extensa y tan antigua debe contar con varios procesos posteriores para dar cuenta de las lenguas históricamente atestiguadas. En este sentido, véase la propuesta reformada de C. Renfrew 2003, según la cual la primera expan-

sión indoeuropea hacia el este habría sido a cargo de los «Early Steppe Proto-Indo-Europeans» que con el tiempo darían lugar a los tocarios, mientras que los iraníes tendrían su origen en otra segunda migración posterior, después de haber participado en la liga balcánica prehistórica. Esta propuesta, aunque más fina que su primera y conocida formulación de 1987, se enfrenta también a problemas lingüísticos considerables, sobre todo en Europa occidental.

y lingüísticas históricamente conocidas, de modo que los emparejamientos formales a gran distancia pueden perfectamente ser el resultado de semejanzas azarosas, de análisis morfológicos erróneos y aceptación de segmentos arbitrarios, máxime cuando no se extreman los cuidados en la observancia de las correspondencias fonéticas y no se tienen en cuenta las historias lingüísticas particulares de los territorios históricos en los que se documentan cada uno de los topónimos. Añádase a ello la especial naturaleza de la onomástica, como ha quedado señalada arriba en § 5.5, y tendremos un cuadro enormemente especulativo, donde se dan todos los ingredientes básicos que debilitan cualquier propuesta comparativa.

7. La población y la lengua del País Vasco se inscriben en este cuadro general de la indoeuropeización antigua de Europa occidental. Ahora bien, un proceso masivo datado en época neolítica, aunque presente los problemas arriba señalados, sería mucho más asumible que una indoeuropeización anterior, de época paleolítica, y responsable de la recolonización de la Europa atlántica tras el Dryas reciente. Yo no tendría, en teoría, nada en contra de una indoeuropeización parcial más antigua que la asumida tradicionalmente; solo que concibo la difusión démica del neolítico no como unilingüe sino multilingüe, pudiendo ser el antecesor del vascuence una de estas lenguas neolíticas¹⁹, como muchas otras más que han desaparecido en procesos de sustitución posteriores. Ya hemos visto, sin embargo, que una cosa es admitir la posibilidad de un cuadro tal y otra, muy diferente, probarlo de modo razonablemente seguro. También sería perfectamente válida la hipótesis de un antecesor vasco paleolítico autóctono o residente en el occidente europeo, que hubiera sufrido una aculturación neolítica.

La indoeuropeización paleolítica se me antoja, sin embargo, enormemente cargada de problemas. Admitiendo los «cognados toponímicos» de Villar, observamos que no hay diferencias lingüísticas entre el indoeuropeo paleolítico y el neolítico²⁰ (excepto la distribución territorial), a pesar de que habría una separación mínima de cinco milenios, lo cual sería muy difícil de admitir para un lingüista histórico.

Por otro lado, aun admitiendo la recolonización postglacial de Europa según la teoría de M. Otte, es arbitrario pensar que fuera estrictamente monolingüe, como el mismo Villar sugería en el pasaje citado arriba (§ 4.4). Y sin embargo, Villar se esfuerza en su último capítulo en intentar armonizar los mapas genéticos de A. Torroni (sobre la distribución de los Hg V, H1 y H3 del ADNmt) con la toponimia de patrón B.

En ese intento de armonización hallo algunas contradicciones, que paso a señalar someramente:

- a. en el plano puramente genético no está claro el papel del Hg V en la población vasca; mientras que los datos aportados por Torroni *et al.* 1988 indicaban un porcentaje alto en la población guipuzcoana (no correspondido de cerca ni por vizcaínos y alaveses)²¹, lo que le llevó a

¹⁹ Iba a decir «babilónicas», como guiño a la tradición erudita renacentista vasca y española.

²⁰ El vocalismo /u/ de los topónimos en *burdo-* y *burso-* le parece problemático, pero admite su relación con otros que presentan vocalismo en /a/: *Bard-uli*, *Barsa*, etc. No tiene problemas en admitir la indoeuropeidad de *alant-*, *lauro-*, *wak-* y *lut-*, topónimos pertenecientes a este patrón distribucional B (cf. p. 145).

²¹ En la Tabla 2 (p. 1142) adjudica a los guipuzcoanos un 20% de Hg V, mientras que Bertranpetit *et*

al. les adjudicaban un 11,1%. Las poblaciones con mayor índice son: Cerdeña septentrional 10,4%, Marruecos beréber, 11,1%, Cataluña 26,7% y los Saami con 40,9%. Para Torroni, la elevadísima proporción de los Saami se debería a un efecto fundacional relativamente reciente, que aún se mantiene en sus mitocondrias, mientras que sus cromosomas Y muestran un flujo asiático acusado.

- postular como foco para la recolonización europea una población vasca originaria, los datos procedentes del ADNmt del conjunto familiar de los restos humanos de Aldaieta (s. VI-VII d. C.) no muestran este Hg. Consiguientemente Izagirre & De la Rúa (1999), basándose en la carencia antigua y en la gran disparidad actual entre diversos territorios vascos, achacan los datos a fenómenos de deriva génica en poblaciones pequeñas y dudan de las conclusiones debidas a movimiento de población.
- b. Villar hace una interpretación diferente: admite que la distribución de los Hg V, H1 y H3 se debe a migración poblacional a partir del refugio peninsular. Estos Hg no son exclusivos de los vascos, sino de todas las poblaciones del Norte peninsular, especialmente de los Cantabros pasiegos. En esos lugares la toponimia indoeuropea es normal, pero como los vascos también participaron de la recolonización logra un modo para conciliar los datos contradictorios: 1. como consecuencia del análisis y de la interpretación de los topónimos —cuyo estudio detallado ahora no puedo abordar—, la población antigua vasca era indoeuropea, 2. participaron en la difusión tanto de los Hg señalados como de los topónimos del patrón B, y 3. la población vasca histórica se debe a una migración de población vascofona desde el norte de los Pirineos al País Vasco español y Navarra en época post-romana y alto-medieval²².
- c. Ahora bien, si algún indicio arqueológico existe sobre esta presencia de gentes ultrapieninsulares en época alto-medieval es la interesante necrópolis de Aldaieta en Nanclares de Gamboa (Álava)²³. Los estudios genéticos realizados a partir del ADNmt conservado en sus restos muestran algunos rasgos difíciles de encajar en ese cuadro. Por un lado Alzualde *et al.* 2007, 155, manifiestan que «this population consists largely of autochthonous individuals from north of the Iberian peninsula» en atención a la presencia de ciertos haplotipos muy específicos que solamente se encuentran en la población actual del norte peninsular. Sin embargo, en otro estudio comparativo del ADNmt antiguo del País Vasco Alzualde *et al.* 2005 incidían sobre la diferencia existente entre las muestras de los yacimientos del neolítico final (San Juan ante Portam Latinam, Longar y Pico Ramos) y las de Aldaieta: mientras aquellos presentaban un aspecto genético más «neolítico», los restos de Aldaieta se encuadraban mejor en el contexto actual de la Europa atlántica. Pero en atención a los Haplogrupos mitocondriales sintomáticos del carácter paleolítico de la población vasca según Torroni, —es decir, alta presencia de H y escasa de J—, todo ese ADNmt antiguo (tanto el neolítico como el de Aldaieta) se aleja por igual de los niveles actuales. Con ello el panorama aparece mucho más complejo que el que resulta de la utilización de los datos modernos y los intentos de unir tal o cual haplogrupo con una determinada migración pierden fuerza.
- d. Villar concluye, por tanto, que el dominio originario, o al menos suficientemente antiguo, de la lengua *debía estar* en zona aquitana al norte de los Pirineos, ya que no lo estaba en la península. Pero es una conclusión que deja sin probar y, si bien se mira, contradictoria con sus supuestos metodológicos y los datos genéticos utilizados. No tiene sentido que la recolonización de Europa en época postglaciar se hiciera con habitantes del norte peninsular, de Cantabria, del País Vasco español y Navarra —todos ellos indoeuropeos— y no participaran

²² Esta última suposición se vería falsada si fueran auténticos los grafitos en lengua vasca aparecidos en 2006 en el yacimiento de Iruña-Veleia, que los arqueólogos responsables de la excavación datan entre el s. III y s. v d. C. Los hallazgos están en estudio y no pueden

utilizarse por el momento como argumento para ninguna postura; véase Gorrochategui 2006b.

²³ No es el único testimonio, pero sí el más elocuente, según los estudios de Azkarate 1999.

en la empresa sus inmediatos vecinos nordpirenaicos del Béarn, de la Bigorre, del Gers, etc., que formaban parte del mismo refugio²⁴. Consiguientemente estos deberían ser también indoeuropeos.

8. Soy consciente de que el panorama presentado en este trabajo es un tanto pesimista con respecto a las relaciones entre genética de poblaciones y lenguas, y especialmente en el caso de topónimos sucedáneos de lenguas. Y sin embargo, si he dedicado cierto esfuerzo a la tarea es porque estoy convencido de que una cierta colaboración entre las dos disciplinas –con el aporte de la arqueología, que he dejado de lado ahora– es posible. Pero para que se produzcan avances significativos de índole interdisciplinar es necesario el papel de la crítica, que depure en primera instancia las conclusiones precipitadas o poco fundadas en cada una de las disciplinas y asegure los controles en un posible emparejamiento posterior de los datos de naturaleza diferente. A este cometido destructor, no especialmente brillante aunque necesario, más en la estela del *thinking straight* que en la del *thinking big*, he dedicado mi esfuerzo.

Joaquín GORROCHATEGUI

*Instituto de Ciencias de la Antigüedad / Antzinaroko Zientzien Institutua
Universidad del País Vasco / Euskal Herriko Unibertsitatea*

BIBLIOGRAFÍA

- ALZUALDE, A. *et al.*, 2005, «Temporal Mitochondrial DNA Variation in the Basque Country: Influence of Post-Neolithic Events», *Annals of Human Genetics* 69, 665-679.
- ALZUALDE, A. *et al.*, 2006, «Insights into the «Isolation» of the basques: mtDNA lineages from the historical site of Aldaieta (6th-7th centuries AD)», *Am. J. Phys. Anthropol.* 130, 394-404.
- ALZUALDE, A. *et al.*, 2007, «Influences of the European Kingdoms of Late Antiquity on the Basque Country: An Ancient-DNA Study», *Current Anthropology* 48:1, 155-163.
- ARNAIZ VILLENA, A. & ALONSO GARCÍA, J., 2000, «The Usko-Mediterranean languages», en: A. Arnaiz Villena (ed.), 205-246.
- ARNAIZ VILLENA, A. (ed.), 2000, *Prehistoric Iberia. Genetics, Anthropology, and Linguistics*, New York: Kluwer Academic/Plenum Publ.
- AZKARATE, A., 1999, *Necrópolis tardo-antigua de Aldaieta (Nanclares de Gamboa, Álava)* [Memorias de Yacimientos Alaveses, 6], Vitoria / Gasteiz: Diputación Foral de Álava.
- BOLNICK, D.A., SHOOK, B.A., CAMPBELL, L., GODDARD, I., 2004, «Problematic Use of Greenberg's Linguistic Classification of the Americas in Studies of Native American Genetic Variation», *Am. J. Hum. Genet.* 75, 519-523.
- BOSCH, E. *et al.*, 2001, «High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between Northwestern Africa and the Iberian Peninsula», *Am. J. Hum. Genet.* 68, 1019-1029.
- CAMPBELL, L., [en prensa], «Languages and Genes in Collaboration: some Practical Matters», en: Comrie, B. (ed.), *Language and Genes*, Cambridge: Cambridge U.P.
- CAVALLI-SFORZA, L.L., MENOZZI, P. & PIAZZA, A., 1994, *The History and Geography of Human Genes*, Princeton.
- GOLLA, V., 2000, «Languages families of North America», en: C. Renfrew (ed.), *America Past, America Present: Genes and Languages in the Americas and Beyond*, 59-72, Cambridge: McDonald Institute for Archaeological Research.
- GORROCHATEGUI, J. & LAKARRA, J., 2001, «Comparación lingüística, filología y reconstrucción del protovasco», en: Villar, F. & Fernández Álvarez, M.^a P. (eds.), *Religión lengua y Cultura prerromanas de Hispania* [VIII Coloquio de lenguas y Culturas paleohispánicas], 407-438, Salamanca.

²⁴ Hay muchísimos datos de orden etnográfico, incluidos los lingüísticos, que abogan por una estrecha rela-

ción cultural entre ambas vertientes de la cadena pirenaica, desde que tenemos noticia histórica de estas tierras.

- GORROCHATÉGUI, J., 2001, «Planteamientos de la lingüística histórica en la datación del Euskara», *Ciencia y cultura vasca, y redes telemáticas* [XV Congreso de Estudios Vascos], 103-114, Donostia – Iruñea – Baiona: Eusko Ikaskuntza.
- GORROCHATÉGUI, J., 2006a, «Onomástica vasca y aquitana: Elementos para el conocimiento de la Historia Antigua de Navarra», en: Andreu Pintado, J. (ed.), *Navarra en la Antigüedad. Propuesta de Actualización*, 111-134, Pamplona/Iruñea: Gobierno de Navarra.
- GORROCHATÉGUI, J., 2006b, «Los asombrosos hallazgos de Iruña-Veleia», *El Correo* (18/11/2006), 69-70.
- GREENBERG, J. H., 2000, *Indo-European and its closest relatives : the Eurasiatic language family*, Stanford: Stanford Univ.
- HOZ, J. DE, 1999, «Viaje a ninguna parte a través del Mediterráneo; las lenguas que no hablaron ni íberos, ni etruscos, ni cretenses», *Revista de libros*, Madrid.
- IZAGIRRE, N. & C. DE LA RÚA, 1999, «An mtDNA Analysis in Ancient Basque Populations: Implications for Haplogroup V as a Marker for a Major Paleolithic Expansion from Southwestern Europe », *Am. J. Hum. Genet.* 65, 199-207.
- LAKARRA, J. 1996 [1999], «Sobre el Europeo Antiguo y la reconstrucción del protovasco», *ASJU* 30:1, 1-70.
- RENFREW, C., 2002, «Genetics and language in contemporary archaeology», en: Cunliffe, B., Davies, W. & Renfrew, C. (eds.), *Archaeology. The Widening Debate*, 43-76, Oxford: British Academy.
- RENFREW, C., 2003, «Time Depth, Convergence Theory, and Innovation in Proto-Indo-European: ‘Old Europe’ as PIE Linguistic Area», en: Bammesberger, A. & Vennemann, Th. (eds.), *Languages in Prehistoric Europe*, 173-181, Heidelberg: Winter.
- RUHLEN, M., 2000, «The Basque language is included in the Dene-Caucasian language family», en: Arnaiz-Villena, A. (ed.), 197-204.
- SEMINO, O. *et al.*, 2004, «Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area», *Am. J. Hum. Genet.* 74:1023–1034.
- SIMONI, L. *et al.*, 2000, «Spatial Variation of mtDNA Hypervariable Region I among European Populations», en: Renfrew, C. & Boyle, K. (eds.), *Archaeogenetics: DNA and the population prehistory of Europe*, 131-138, Cambridge: McDonald Institute Monographs.
- SIMS-WILLIAMS, P., 1998, «Genetics, linguistics, and prehistory: thinking big and thinking straight», *Antiquity* 72, 505-527.
- TORRONI, A. *et al.*, 1998, «mtDNA Analysis reveals a Major Late Paleolithic Population Expansion from Southwestern to Northeastern Europe», *Am. J. Hum. Genet.* 62, 1137-1152.
- TRASK, R.L., 1996, *Historical Linguistics*, London: Arnold.
- VENNEMANN, TH., 1994, «Linguistic reconstruction in the context of European prehistory», *TPhS* 92:2, 215-285.
- VENNEMANN, TH., 2004, «Sprachgeburth durch Sprachkontakt: Die Entstehung des Englischen», en: Schrijver, P. & Mumm, P.-A. (eds.), *Sprachtod and Sprachgeburth*, [Münchner Forschungen zur historisch-Sprachwissenschaft 2], 21-56, Bremen: Hempen Verlag.
- VILLAR, F. & PRÓSPER, B., 2005, *Vascos, Celtas e Indoeuropeos. Genes y lenguas*, Salamanca: Universidad de Salamanca.
- VILLAR, F., 2005b, «Los indoeuropeos y Europa», en: Carracedo, A. & Pereira-Menaut, G. (coord.), *Xenética e Historia no noroeste peninsular. Unha perspectiva interdisciplinaria*, 203-236, Santiago de Compostela: Consello da Cultura Galega.