

## Euskal Corpus Onomastikoa - Corpus onomástico vasco. Un modelo de base de datos para la gestión de información onomástica

Patxi Galé García

DOI: 10.2436/15.8040.01.249

### Resumen

El *Corpus Onomástico Vasco* es un proyecto que puso en marcha el año 2005 la Real Academia de la Lengua Vasca / Euskaltzaindia. El proceso de estandarización del euskera está en marcha desde la década de 1960. Además de los problemas habituales que esas labores de actualización causan en la lengua común, hallamos otros obstáculos propios del léxico onomástico que hacen aun más difícil la estandarización de este campo. El euskera ha accedido hace poco al estatus de lengua administrativa. Hasta ahora, los topónimos y apellidos patrimonio de esta lengua los hemos visto casi siempre escritos según las normas gráficas de los idiomas fuertes y cultos que nos rodean, y además ha sido obligatorio utilizarlos así. El léxico onomástico tiene, propiamente, un cierto vínculo especial con la identidad de la gente: hablamos sobre todo del nombre de los lugares que conforman el escenario de la vida de las personas y, más aun, de los apellidos que facilitan la identificación de personas y linajes. Si hasta ahora hemos visto siempre escritos de un modo el nombre de nuestro pueblo, nuestro barrio, nuestra casa o —lo que es más íntimo— nuestro apellido, el difícil objetivo de la estandarización consiste en modificar ese hábito, estableciendo su grafía eusquérica “elevada” o “culta”. El *Corpus Onomástico Vasco* es el proyecto que pretende reunir toda la información necesaria para ello en una única base de datos. Esa información tiene dos pilares principales. En primer lugar, hay que recopilar los datos onomásticos contenidos en publicaciones y trabajos de tipo archivístico, histórico, geográfico, filológico, literario, etc., perfectamente datados y fielmente transcritos. En segundo lugar, hay que recoger el uso oral, seleccionando informantes cualificados y recopilando los topónimos orales lugar por lugar, pronunciados en los idiomas utilizados en cada zona geográfica (euskera, castellano, francés). Simultáneamente, esos mismos informantes nos darán noticia de la naturaleza y la situación precisa de los lugares nombrados. Con la ayuda de unos estrictos criterios de normativización, y teniendo a la vista toda esa información —y solo de este modo— puede fijar la Academia cuál es la grafía moderna estándar de un topónimo o un apellido eusquérico.

\*\*\*\*\*

El *Euskal Corpus Onomastikoa* o *ECO* (‘Corpus Onomástico Vasco’) es un proyecto que la Real Academia de la Lengua Vasca / Euskaltzaindia puso en marcha en el año 2005, y que consiste en una base de datos diseñada para ser el contenedor de información onomástica procedente de investigaciones en torno a la disciplina. No se trata por tanto de un *proyecto de investigación* en sentido estricto: Euskaltzaindia no pretende con ello “investigar” la onomástica vasca, sino conocer y recopilar las investigaciones que se hacen en ese campo. El objetivo de cada investigador puede ser dispar: aclarar las etimologías de los topónimos o apellidos, buscar las formas arcaicas de nombres de lugares, estudiar la evolución que los topónimos han tenido a lo largo del tiempo, rastrear el origen, significado o tradición de los nombres de pila... El *ECO* no se limita a recoger estas informaciones, pero su objetivo último es más bien de carácter normativo. Se trata de conocer todos los datos disponibles sobre cada nombre, apellido o topónimo, para poder así establecer con cierta garantía su forma eusquérica estándar.

### Estandarización del euskera y onomástica

Por contextualizar el proyecto dentro de la propia historia de la lengua vasca, podemos comenzar afirmando —sin demasiado temor a equivocarnos— que esa historia viene a ser sencillamente la de su propia agonía. Ciertamente se trata de una agonía larga, pero no por ello resulta ser menos agonía. De hecho, desde las primeras noticias que tenemos del euskera,

su extensión y su presencia social no han hecho otra cosa que disminuir hasta prácticamente nuestros días. Y no podemos hablar de cambio en esa tendencia hasta que el euskera ha acometido su proceso de estandarización.

Aunque a lo largo de su historia sí ha habido algunos tímidos intentos de acuñar unos dialectos literarios, el proyecto serio y “definitivo” de establecer un euskera literario común no se acometió hasta la década de 1960, bajo la dirección del conocido lingüista y académico Koldo Mitxelena. Con la salvedad de algún que otro autor ilustrado y los hombres de iglesia, el euskera había estado limitado hasta entonces al mundo rural iletrado. Por esa razón es frecuente hallar en la documentación histórica la expresión “lengua vulgar vascongada”, dando por supuesto que lo culto era el uso del castellano o del latín, reservado a las élites alfabetizadas del país. Téngase en cuenta que, además, cuanto más atrás vayamos en el tiempo, la presencia de las lenguas cultas era menor en términos sociológicos: la inmensa mayoría de la población era vascohablante monolingüe, pero la lengua “deseada” era la de esas minorías cultas o ilustradas. No es de extrañar, pues, que incluso hasta los vascos nacidos en la primera mitad del siglo XX no sólo desearan aprender castellano y que sus hijos lo hicieran, sino que además se avergonzaban de que se les notase que eran vascohablantes: el acento vascongado era para ellos sinónimo de aldeanismo y debía disimularse. No sólo había que aprender castellano, sino que era bueno también no saber euskera.

Aunque el proceso de estandarización de la lengua no ha estado exento de escollos, disidencias, discusiones, tendencias y demás peleas, hoy no se puede poner sinceramente en duda su rotundo éxito. La larga agonía de la lengua que mencionábamos más arriba, sólo se ha ralentizado y ha cambiado de tendencia a partir de ese paso. La referencia a un modelo de lengua vasca culta, nos ha llevado en la actualidad a la percepción ya generalizada de que saber euskera no es una desgracia, sino un mérito cultural.

Cabe añadir que también el cambio político operado a finales de la década de 1970 vino a ayudar en este proceso de recuperación de la lengua, con diferencias ciertamente significativas en unos u otros territorios político-administrativos en que se divide Euskal Herria. El euskera, además de ser ya idioma administrativo cooficial, ha llegado prácticamente a cualquier rincón del espectro cultural: todos los niveles educativos, la literatura, el periodismo, los medios de comunicación, algunas administraciones públicas, etc.

Sin embargo, el pasado que he pretendido dibujar antes, ha dejado pesados lastres en la situación de la lengua. Estamos hablando de un idioma antiguo, dividido en dialectos hablados que apenas han tenido cultivo escrito y cuya disgregación se ha visto potenciada por el hecho de vivir relegado al uso privado familiar o, como mucho, local. Y sobre aquella exigua base literaria dividida, Euskaltzaindia ha ido construyendo, con el apoyo entusiasta de no pocos euskaldunes, escritores, periodistas, profesores, etc., ese modelo de euskera culto que hoy en día se conoce como *euskara batua* o ‘euskera unificado’.

Ese gran paso que la Academia ha conseguido, sin embargo, ha sido contestado en algunos ambientes de marcada tendencia populista o “etnografista”, por varias razones. Primero, el esfuerzo de alfabetización que deben realizar los hablantes de los dialectos extremos del euskera (el vizcaíno y el suletino) es considerablemente mayor que el que tienen que hacer los de los dialectos centrales, puesto que el estándar se ha basado sobre todo en estos últimos. En segundo lugar, el apego a lo que “hemos venido utilizando” hasta ahora, es mucho más rentable en términos de esfuerzo, que la voluntad de alfabetizarse y adquirir el modelo culto que nos abre las puertas a la radio-televisión, la prensa o la literatura. Por esas razones, en algunos lugares han surgido posiciones contrarias a esa estandarización, una especie de “nostalgia del aldeanismo” que aún hoy es el marco en que se encuentran cómodos muchos vascohablantes, incluidas algunas firmas reseñables del mundo de la cultura vasca.

El procedimiento de estandarización de toda lengua, y también en el caso del euskera, suele consistir básicamente en priorizar unos recursos léxicos, sintácticos, morfológicos y

fonológicos de entre los disponibles en el uso escrito de los diversos dialectos. Ello, claro está, conlleva el abandono de otras variantes dialectales o locales, cuyos usuarios no siempre aceptarán de buen grado. En el campo del léxico, la tendencia general es priorizar las formas más extendidas o tradicionales y, en consecuencia, abandonar las variantes dialectales de esas mismas palabras. Sin embargo, se suele aceptar el uso de vocablos diferentes utilizados en uno u otro dialecto literario, indicando en tal caso su filiación dialectal “menor”. Pero el campo de la onomástica va más allá. El modelo culto adoptado para esta parte de la lexicología es respetuoso con las formas dialectales, incluso cuando se trata de variantes formales de la palabra estándar priorizada: Por ejemplo, los apellidos *Aréizaga*, *Arizaga*, *Aréchaga*, *Hariçague* y *Haritchague* son variantes de una misma composición léxica que significa ‘lugar de robles’, cuyos componentes “estándar” serían *haritz* ‘roble’, *Quercus robur*, y el sufijo locativo *-aga*. La forma más antigua del sustantivo inicial entre las que observamos en los apellidos citados, es *hareitz*, de donde proceden las variantes vivas en la actualidad: *aritz* sin aspiración en los dialectos centrales, *haritz* con aspiración en las hablas septentrionales, *haritx* con aspiración y palatalización en los dialectos orientales y *aretx* sin aspiración y con palatalización arcaica en los dialectos occidentales. Como consecuencia de todo ello, los apellidos citados pueden escribirse en euskera estándar respectivamente *Aritzaga*, *Areitzaga*, *Aretxaga*, *Haritzaga* y *Haritxaga*.

Sin embargo, los partidarios de la tendencia “etnografista” citada antes, verán en esta aceptación la puerta abierta para otras variantes populares mucho más recientes. Normalmente no ocurre en apellidos, a cuya grafía se suele tener un gran apego aunque se trate de una forma castellanizante o afrancesante. Pero en toponimia y en nombres de pila está muy generalizada. Por poner un ejemplo, nadie ve necesidad de modificar la grafía de los apellidos *Iturreta*, equivalente de ‘fontanal’ o ‘fontaneda’, o *Urizar* que viene a significar ‘villa vieja’. Sin embargo, los caseríos de donde son originarios esos apellidos, pueden denominarse en la actualidad “popularmente” *Itxúrta* y *Uxárre* e, incluso, lucir en su fachada un letrero que muestre precisamente esa pronunciación popular: algo parecido a rotular *Güerba* en vez de *Huelva* o *Tàrgue* en vez de *Tàrrega*. Eso es lo que más arriba denominábamos “nostalgia del aldenanismo”: en ciertos ambientes escribir de este modo goza de cierto predicamento, aunque no creemos que quienes defienden ese modelo gráfico sean conscientes de que, en definitiva, están negando a la lengua aquel paradigma culto que necesita construir para ser una lengua “normal” y tener alguna opción a sobrevivir.

En el extremo contrario, los nombres, apellidos y topónimos eusquéricos que los investigadores hallan en la documentación escrita, figuran casi siempre adaptados a la normativa gráfica de las lenguas administrativas “cultas” de cada momento y lugar o, dicho de otro modo, disfrazadas con la grafía del idioma oficial correspondiente, las modas o estándares de cada época e, incluso, con las costumbres del escribano de turno en cada administración y momento histórico. Desde una óptica que vamos a llamar “historicista”, se negará incluso la existencia de topónimos vascos como *Lazkao* u *Hondarribia* porque apenas se hallan en la documentación escrita. Para ello, evitarán recordar que el euskera nunca ha tenido estatus de lengua administrativa y que sus topónimos siempre aparecen citados *Lazcano*, *Hondar(r)abie*, *Fontar(r)abie*, *Fuenterrabía*... Básicamente, la documentación histórica donde se suelen rastrear datos onomásticos vascos, estará redactada según la época y el lugar, en latín, en castellano, en occitano o en francés. No faltarán quienes defiendan que la grafía estándar debe ser la que se ha utilizado en esas fuentes documentales, donde el euskera es literalmente una lengua inexistente o “imaginaria”.

Visto de este modo, parece que la elección de una grafía normativa para las formas onomásticas vascas va a estar en el debate entre priorizar lo hablado o priorizar lo escrito. Desde la Comisión de Onomástica, se ve con claridad que ambos datos tienen su valor y que en cada tipo de testimonio debemos buscar unos u otros rasgos que nos ayuden a tomar las

decisiones más acertadas y acordes con el proceso general de estandarización del idioma. Resumiendo: con los datos documentales y con los que las encuestas obtengan en trabajo etnográfico de campo, la labor de la Comisión de Onomástica de Euskaltzaindia consiste en decidir cómo han de escribirse y utilizarse esos nombres propios en euskera estándar actual, en euskera culto.

En ese proceso solemos distinguir con insistencia dos fases bien diferenciadas: el trabajo académico consiste en aplicar a esos datos onomásticos una serie compleja de criterios ortográficos, de modo que se establezca la forma normativa actual de cada nombre propio, sea este topónimo, nombre de pila, apellido, etc.

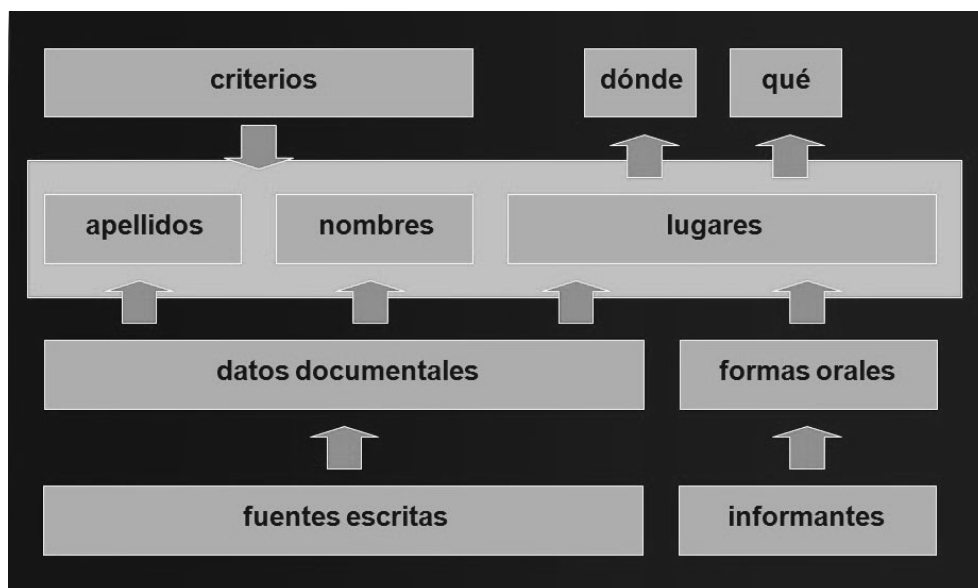
Esta tarea solemos denominarla “normativización”, contraponiéndola expresamente a la fase siguiente, que llamamos “normalización” y que vendría a ser el proceso colectivo de asunción de esos nombres propios como “normales” en el uso cotidiano. Este paso queda más allá de las posibilidades de la Academia y necesita, de algún modo, del refrendo administrativo y social más o menos explícito. Sería deseable que las instituciones y los medios de comunicación asumiesen esa labor y esa responsabilidad como propia con mayor convencimiento y decisión, siempre en la medida de sus posibilidades y de sus competencias correspondientes. Por ejemplo, la señalización viaria ejerce una enorme labor de normalización. Un rótulo recién puesto, que muestra una grafía aún extraña para los usuarios pero normativamente correcta, puede crear problemas o cierta contestación los primeros días, pero en un tiempo mucho más breve de lo que pensamos, conseguirá que la forma rotulada obtenga carta de naturaleza entre los usuarios habituales. Enseguida podremos oír que tal o cual lugar se han llamado así “de toda la vida”.

En esta cuestión de la normalización en su sentido sociológico, también hay que resaltar que el objeto de investigación de la onomástica, se trate de nombres de pila, apellidos o topónimos –y en ese orden– conlleva inevitables vínculos identitarios. Cualquier hablante demostrará mucha mayor sensibilidad ante el cambio de grafía de su nombre, su apellido o del nombre de su pueblo o barrio, que ante el cambio de grafía de un sustantivo común. Por esa razón, parece que la Academia no puede limitarse a emitir listados de formas onomásticas como decisiones ya tomadas “desde arriba”. Resulta fundamental poner en marcha mecanismos de información y labor didáctica para que todos los usuarios interesados se sientan parte del proyecto de recuperación lingüística en que estamos inmersos. Tanto Euskaltzaindia como muchos proyectos locales de investigación toponímica realizados estos últimos años, han dado pasos en este sentido y las dinámicas participativas no deben olvidarse nunca, puesto que constituyen la garantía de éxito más firme para la normalización de nuestras formas onomásticas en su grafía correcta.

Como decíamos al principio, a fin de que los trabajos de normativización de la Comisión de Onomástica de Euskaltzaindia tuviesen la mayor garantía posible de estabilidad y “definitividad”, en el año 2005 se puso en marcha el proyecto denominado *Euskal Corpus Onomastikoa*, consistente en la creación de una base de datos que pudiese reunir la mayor cantidad posible de información onomástica. En este sentido, la propia Comisión era consciente de que no se podían tomar decisiones normativas con criterios más “afectivos” y “populistas” que científicos, puesto que los datos en muchas ocasiones indican que son erróneas algunas formas gráficas a las que hemos llegado a tomar cierto apego.

### **Estructura del corpus**

Hechas estas aclaraciones que espero sirvan para contextualizar el proyecto y dar una idea de sus objetivos, pasemos a describir sucintamente su estructura. El *ECO* se articula en una serie de tablas de datos interrelacionadas, cuya estructura responde con cierta lógica a la que presentamos a continuación:



En la base de la estructura tenemos un corpus abierto de posibles fuentes de datos. Estas fuentes son de diferente naturaleza, atendiendo a los datos que nos ofrecen. Hay fuentes documentales archivísticas que pueden ser expurgadas directamente o a través de publicaciones de cartularios, archivos municipales, etc.; otras son publicaciones de carácter histórico-geográfico de más o menos antigüedad, como diccionarios geográficos, estudios históricos, etc.; también hay gran cantidad de estudios de carácter filológico en torno a topónimos, apellidos o nombres de pila, cuyo objetivo suele ser la averiguación de su composición y significado; últimamente también van viendo la luz estudios toponímicos, sobre todo monografías de ámbito municipal, que alcanzan un gran nivel de precisión y profundidad... Estos trabajos han llegado a nosotros en diversos formatos (archivos con o sin catálogo, bibliografía en general, revistas especializadas, hemeroteca, páginas web, etc.) y, como decíamos, sus datos tienen diferente valor dependiendo de la antigüedad, naturaleza y calidad de cada fuente.

En segundo lugar, claro está, hay que generar un corpus de datos procedentes de esas fuentes. Los datos deberán estar fechados y referenciados con precisión, de modo que cualquier usuario pueda consultar directamente en la fuente original cada testimonio recogido. En el caso de los topónimos, es muy importante recoger también la información que cada fuente ofrezca sobre la situación y naturaleza del lugar identificado, de cara a poder situarlo en un mapa o a identificarlo con su denominación actual, que no necesariamente va a ser la misma que nos ofrece la fuente expurgada.

Por otra parte, y de cara a la obtención de datos etnográficos orales, en primer lugar hay que conformar un listado de informantes locales a los que entrevistar. En cada localidad suele ser conocido quién o quiénes son los que mejor conocen los montes, las costas, los ríos del término municipal. Normalmente son personas mayores, que han ejercido lo que ahora llamamos “modos de vida tradicional” y que debido a ello tienen un conocimiento del territorio mucho mayor que nosotros. La extinción de esos modos de vida y la irremediable desaparición paulatina de esas personas, hace que a diario se pierdan para siempre muchos testimonios toponímicos y muchos datos acerca de lugares cuyos nombres quizá sí conocemos, pero que pueden ocultar para siempre su ubicación precisa o su naturaleza geográfica concreta.

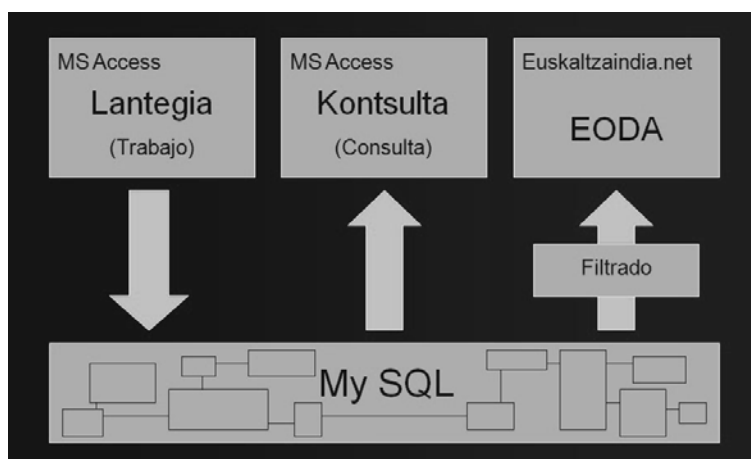
En efecto, los datos a obtener de estos informantes son tanto las propias formas onomásticas –en general topónimos–, como los datos referentes a los lugares por ellos denominados: entidad geográfica, situación precisa, amplitud, cambios habidos, etc. Los

datos orales también han de estar recogidos con precisión y referenciados con exactitud. Normalmente las entrevistas se suelen conservar grabadas en formato digital y resulta conveniente que la referencia posibilite la audición del dato. Aunque en algunos trabajos se suelen transcribir los datos en notación fonética, nosotros creemos que es más práctico hacerlo en grafía fonemática o alfabética corriente, pues esto facilita la transmisión de la información, que además así resulta más comprensible para posibles usuarios interesados en ello. Sin embargo, sí es de alto interés la recogida de los acentos, pues el sistema de acentuación es muy diferente en euskera y en castellano y este dato puede resultarnos de gran valor para establecer la pronunciación culta correcta de las formas onomásticas eusquéricas.

Con estos datos, tenemos que ir estableciendo cuál o cuáles son las grafías cultas de cada uno de los apellidos, nombres de pila o topónimos. Para ello, es necesario establecer una serie de criterios lingüísticos que describan los fenómenos que observamos que suceden de modo recurrente en esos términos onomásticos, y que nos indiquen qué solución gráfica debe adoptarse en cada caso.

Al contrario de lo que se cree comúnmente, la adopción de esos criterios no es “previa” al trabajo de normativización. Muy al contrario, los fenómenos ante los que hay que buscar soluciones gráficas se van detectando a medida que se realiza ese trabajo. Igualmente la recurrencia o la extensión geográfica de un fenómeno fonológico, va siendo tanto más clara cuanto más amplio es el corpus analizado. Por esta razón, esos criterios se han ido formulando y ampliando a la par que se acometían investigaciones de cierto volumen. La reformulación de criterios no debe entenderse con simpleza como voluntad de molestar a los usuarios de la lengua, o como resultado de la arbitrariedad y desorientación de los investigadores. Muy al contrario, la formulación continuada de proposiciones sobre los fenómenos observados, su contraste y posible falsación, y su consiguiente reformulación, constituyen la ruta más saludable y fiable de todo quehacer investigador y, en consecuencia, de todo avance científico.

Volviendo a la base de datos, pasemos a comentar algunas características técnicas de la misma.



En primer lugar, quizá resulte de interés indicar que las tablas que contienen la información se encuentran almacenadas en el software libre para la gestión de bases de datos llamado *My SQL*, si bien para la mayoría de los trabajos se utilizan aplicaciones o formularios *MS Access*. Estos formularios están contenidos en una aplicación denominada *Lantegia* (‘factoría’, ‘lugar de trabajo’), que permite acceder a todos los rincones de la base de datos, incluida la propia estructura, así como intervenir en sus registros (editar, añadir, quitar, etc.).

En segundo lugar, se dispone de otra aplicación *MS Access* muy similar a la anterior, llamada *Kontsulta* (‘consulta’), desde la que también se puede acceder a toda la información, pero sólo para visualizarla o consultarla. También puede obtenerse el informe en papel de

cualquier registro de la base de datos, con toda la información asociada que contenga. El enorme volumen de datos que debe gestionar la aplicación, hace que su uso por vía telemática resulte por ahora un tanto lento y tedioso.

Por último, la Academia ofrece parte de este *corpus* al público en general, a través de su página web ([www.euskaltzaindia.net](http://www.euskaltzaindia.net)) en el área denominada *EODA* (*Euskal Onomastika Datutegia* ‘base de datos de onomástica vasca’). Los datos de la web se actualizan trimestralmente y vienen a ser una versión limitada y filtrada del *ECO*, pues sólo se muestran aquellos registros que han recibido el visto bueno de la propia Comisión de Onomástica o del Pleno de la Academia.

En cuanto al volumen o cantidad de datos que la base de datos contiene en la actualidad, vienen a ser aproximadamente los siguientes: Hay registradas un total de 2.744 fuentes escritas de distinta naturaleza, tamaño y calidad. El grado de vaciado es también variable, pues algunas están a la espera de expurgo, otras han sido despojadas parcialmente y sólo unas cuantas lo han sido de modo exhaustivo. Procedentes de esas fuentes escritas, se han obtenido y registrado hasta ahora un total de 1.533.801 testimonios documentales. En cuanto a los informantes entrevistados, registramos un total de 5.108 personas o grupos, que han proporcionado hasta la actualidad 309.652 testimonios orales. Hay fichados 523 criterios de normativización, referentes a los distintos campos de la onomástica, que se han aplicado con mayor o menor precisión al establecer las grafías de apellidos, nombres o topónimos. Por último, en cuanto a “registros principales”, contamos 10.800 apellidos, 3.196 nombres de persona y 228.927 lugares.

Como hemos dicho más arriba, sólo una pequeña parte de estos registros principales se muestra en la página web al público, a través del *EODA*. Son 9.480 apellidos (el 87,8 % del total), 3.138 nombres de pila (el 98,2 %) y 6.310 lugares (tan sólo el 2,8 % de los lugares fichados).

## Visualización de los datos

Con el fin de ayudar a la mejor comprensión del sistema de recopilación de los datos y su ordenación, mostramos a continuación, a modo de ejemplo, unas fichas de la aplicación de consulta (muy similares a las de la aplicación de trabajo). El objetivo de esas fichas es poder acceder a todos los datos asociados a cada registro, entendiendo por registro un nombre de pila, un apellido o un lugar.

Menú principal: Contiene seis submenús. Los tres de la zona inferior permiten acceder de diversos modos a algunos datos auxiliares o secundarios, como las fuentes documentales, las formas documentales o los criterios de normativización. Los tres de la parte superior, posibilitan la búsqueda de fichas de registros principales del *corpus*, es decir, de cualquier apellido, de cualquier nombre de pila o de cualquier lugar.

Estas fichas “principales” permiten visualizar toda la información asociada a un apellido, nombre de pila o lugar. Vémoslas.

### Deiturak (apellidos)

The screenshot shows the 'EUSKAL CORPUS ONOMASTIKOA DEITURAK' interface. The search term 'deitura: BELTZUNTZE' is entered. The results are displayed in a table with columns for 'era', 'iturrria', and 'datuak'. The table lists various historical and administrative forms of the surname, such as 'belzunce (1571) Lg. Dedic.: Belzunce. Ap. Belzunceco-Iaun-, 'Monsieur de Belsunce'', 'gulielmo arnalt de belzunce, gulielmo ard[...] (1638) O.NUV. II, IV. kap., 109. orr. [0626. orr.]', and 'belzunce (1921-1925) AZK.CPV, 0863. zkia., II. lib., 1004. orr. [10. lib., 092. orr.]'. Below the table, there are fields for 'arautzea', 'data', and 'normalizatuas'. The interface also includes a search bar at the bottom with 'Registro: 2818 de 10800' and 'Sin filtrar'.

La ficha de apellidos está presidida por la grafía normativa correspondiente (lema de la ficha) y muestra además las formas equivalentes habituales en las lenguas oficiales administrativas (español y francés), es decir, las formas castellanizantes o afrancesantes del apellido registrado. En la primera de las pestañas se pueden ver, en orden cronológico, las formas en que el apellido se ha documentado a lo largo de su historia.

En la segunda pestaña se muestra el criterio o criterios de normativización que se ha aplicado para decidir la grafía aprobada.

En la tercera pestaña se muestran otros registros con los que puede estar en relación el apellido lema de la ficha. Puede tratarse de otros apellidos (como en este caso ocurre con la grafía *Belsuntze*, variante igualmente correcta del apellido), pero también de nombres de pila o topónimos.

En el caso de los topónimos relacionados, puede obtenerse más información, de modo que se nos indique qué es y dónde está ese lugar. En este caso, *Beltzuntze* es un lugar del valle navarro de Juslapena (eusk. Txulapain).



## Ponte izenak (nombres de pila)

The screenshot shows the 'EUSKAL CORPUS ONOMASTIKOA' interface for 'PONTE IZENAK'. The main search field contains 'izena: XABIER'. Below this, there are fields for 'sexua: Gizonezkoa', 'hipokoristikoa: ', and 'jaieguna:'. A tabbed interface shows 'dokumentazioa', 'irizpideak', 'izendegia', and 'erlazioak'. The 'izen kideak' section lists 'Javier (gazt.)' and 'Xavier (fr.)'. The 'historia:' field is empty. The 'oharrak:' section contains a note about the name's origin. The 'arautze maila:' is 'Euskaltzaindiaren araua', 'data:' is '1997/01/31', and 'normalizatu:' is . The 'era' section contains a list of references with columns for 'era' and 'iturria datuak'. The 'Aplikazioa: Patxi Galé' is noted at the bottom. The footer shows 'Registro: 3086 de 3196' and search options.

La ficha de nombre de pila tiene parecida estructura, mostrándose también otras informaciones propias de este campo: sexo, si se trata de un diminutivo o hipocorístico, fecha de celebración santoral, etc. Igual que en los apellidos, puede además indicarse el equivalente castellano o francés del nombre registrado y también la documentación recopilada sobre él en orden cronológico. Hay otros campos más “abiertos” que permiten la recopilación de datos históricos y notas acerca del nombre de pila o el santoral.

Los datos documentales de cualquiera de las fichas, permiten consultar la fuente concreta de la que procede cada dato recogido, pues la fuente aparece identificada por una abreviatura que no siempre conoceremos. También puede accederse a la ficha completa del registro documental que, en ocasiones, contiene además el contexto en que la forma documentada ha sido hallada.

En la segunda pestaña, como hemos dicho para los apellidos, también se halla registrado el criterio de normativización aplicado al nombre.

La tercera pestaña se titula *izendegia* ‘nomenclátor’ y contiene el texto aclaratorio que ese registro lleva en la publicación académica titulada *Euskal izendegia* ‘diccionario de nombres de pila’. El objetivo inicial de esta pestaña es precisamente servir de instrumento de trabajo (en la aplicación de trabajo, claro está) para la actualización y reedición del libro, que en la actualidad se encuentra agotado y en proceso de revisión.

La última pestaña muestra los registros relacionados con el que mostramos. En este caso, el nombre de varón *Xabier* está relacionado con el nombre de mujer *Xabierra*, con el apellido *Xabier* y con los lugares llamados así en Navarra (término municipal y su capitalidad).

## Toponimia (nombres de lugar):

La ficha de lugar es la referente a la toponimia y la más compleja.

Además del lema o denominación eusquerica principal, vemos su gentilicio correspondiente en euskera (*hendaia* ‘hendayés’) y su codificación geográfica, así como el genérico geográfico estándar que se le debe aplicar (*udalerria* ‘municipio’).

El campo *altitua* ‘altitud’ puede tener cierta relevancia para los montes, puertos de montaña, collados, etc.

En todo registro se indica dónde se encuentra situado el lugar designado. En este caso, el municipio *Hendaia* se sitúa en la comarca de la costa labortana (*Lapurdi itsasegia*), en el país vasco-francés.

El sistema en que se ve la documentación es idéntico al de las fichas de apellidos o nombres de persona.

La segunda pestaña nos ofrece el dato de recogida oral correspondiente al topónimo registrado, dato que no parece tan relevante en el caso de los apellidos o nombres de pila, donde no se recoge sistemáticamente. Igual que con las fuentes documentales, también se puede consultar el dato del informante entrevistado que ha proporcionado el topónimo.

La tercera pestaña nos indica las denominaciones que se dan o han dado a ese lugar. No se trata de todas las grafías posibles, sino de aquellas que se han utilizado o utilizan en algún idioma u ocasión concreta. En este caso vemos que la denominación oficial (e igualmente la denominación afrancesante) es *Hendaye*. Este campo es el que se utilizará para realizar búsquedas, pues además de grafías castellanizantes (*Guipúzcoa* < *Gipuzkoa*) y afrancesantes (*Urrugne* > *Urruña*), contiene también otras históricas (*Gaztelu-Elexabeiti* ~ *Arteaga*), anticuadas (*Leire* ~ *Leyre*), populares (*Pardixerria* ~ *Urraulgoiti*), etc.

La siguiente pestaña es la referente a los criterios de normativización aplicados, idéntica a la de apellidos o nombres.

La quinta pestaña indica la geoposición, donde se almacenan las coordenadas precisas de situación del lugar designado. Se pueden recoger tanto las coordenadas UTM con referencia al *datum* que se desee, como las coordenadas geográficas en grados y decimales, formato que facilita enormemente el almacenamiento, así como todo tipo de cálculos.

La sexta pestaña nos muestra el mapa o mapas en que se halla el lugar registrado, en referencia a las cartografías oficiales de cada zona administrativa (IGN español o francés, cartografías autonómicas o forales, etc.). En cualquier caso, se puede consultar la fuente cartográfica como una fuente documental más.

La última pestaña nos muestra las relaciones con otros registros, que en este caso se limitan al pueblo *Hendaia*, capitalidad del municipio del mismo nombre, y al apellido *Endaia* que, curiosamente, debe escribirse sin hache.

En todas las fichas hemos podido ver en la parte inferior unos campos que nos indican el “grado de garantía” de la normativización. En este caso es ‘norma de Euskaltzaindia’ (*Euskaltzaindiaren araua*), lo que indica que este nombre ha sido aprobado por el Pleno de la Academia.

También figura la fecha de esa normativización, así como su grado (aproximado en todo caso) de normalización en cuanto a uso y arraigo social.

Por último, como ya hemos comentado más arriba, las grafías onomásticas que han sido aprobadas por el Pleno de la Academia o por la Comisión de Onomástica, pueden consultarse en la página web, en el apartado titulado *EODA* (*Euskal Onomastika Datutegia* ‘base de datos de onomástica vasca’).

The screenshot shows the website <http://www.euskaltzaindia.net/onomastika>. The page title is "Onomastika - Euskaltzaindia". The main navigation menu includes: PLAZABERRI, AGENDA, HARREMANAK, BULEGO-HELBIDEAK, EUSKARA, ERAKUNDEA, HIZKUNTZA BALIABIDEAK, IKER EGITASMOAK, JAGON EGITASMOAK, EGITASMO INSTITUZIONALAK, AZKUE BIBLIOTEKA, and ARGITALPENAK. The search bar contains "Azcoitia" and the "Bilatu" button is visible. Below the search bar, there is a section titled "Baliabide zerrenda" (List of resources) with the following items:

- Pertsona-izenek
- Deiturak
- Toponimia (leku izenak)
- Exonimia (atzerriko izenak)
- Argibide-eskaerak

At the bottom of the page, there is a section titled "Euskal Onomastikaren Datutegia" (Basque Onomastic Database), which is circled in red. Below this title, there is a paragraph of text:

*Euskaltzaindiaren Onomastika batzordearen lanak biltzen dira hemen. Lanok lau multzo nagusitan banatzen dira, pertsona izenak, deiturak, toponimia (leku izenak) eta exonimia (atzerriko izenak).*

Basta introducir un dato en la grafía que se quiera, por ejemplo *Azcoitia* con *-c-* (es decir, su forma castellanizante), para obtener la siguiente respuesta:

The screenshot shows the Euskal Onomastikaren Datutegia (EODA) website interface. The browser address bar shows the URL: <http://www.euskaltzaindia.net/index.php?option=...>. The page title is "Euskal Onomastikaren Datutegia (EODA)".

The search results are organized into sections:

- Pertsona-izenak**: A search bar with the input "Azcoitia".
- Deiturak**: A table with two columns: "Ohiko idazkera" and "Egungo idazkera akademikoa". The row shows "Azcoitia" and "AZKOITIA".
- Toponimia**: A table with three columns: "Euskal era arautua", "Azalpena", and "Non". It lists two entries for "AZKOITIA (-A)".
 

Euskal era arautua	Azalpena	Non
AZKOITIA (-A)	Izaera: <i>Hiria, udalerriko auzo nagusia</i> Herritarra: azkoitiar Ofiziala: Azkoitia	AZKOITIA (-A)
AZKOITIA (-A)	Izaera: <i>Udalerria</i> Herritarra: azkoitiar Ofiziala: Azkoitia	UROLA KOSTA
- Exonimia**: A search bar with the input "Azcoitia".

At the bottom of the page, the logo for EUSKALTZAINDIA (REAL ACADEMIA DE LA LENGUA VASCA / ACADEMIE DE LA LANGUE BASQUE) is visible, along with the motto "EKIN ETA JARRAI".

Hay un apellido *Azkoitia* y dos lugares de ese nombre: un municipio de Gipuzkoa en la comarca de *Urola kosta* ('costa del [valle del río] Urola'), así como la villa homónima capitalidad del municipio.

Esperamos que esta sucinta descripción dé una idea de los objetivos, estructura y contenido del 'Corpus Onomástico Vasco' (*Euskal Corpus Onomastikoa*), proyecto básico para la Comisión de Onomástica de Euskaltzaindia y, en definitiva, para la normalización de los nombres propios vascos como una parte más del patrimonio cultural universal.

Patxi Galé

Responsable técnico del proyecto que se describe en la presente comunicación, cuya dirección está a cargo del presidente de la Comisión de Onomástica y académico de número de Euskaltzaindia, Andres Iñigo Ariztegi

Comisión de Onomástica

Euskaltzaindia / Real Academia de la Lengua Vasca

[info@euskaltzaindia.net](mailto:info@euskaltzaindia.net)