HAL
open science

El Atlas Lingüístico Vasco (EHHA)

Gotzon Aurrekoetxea

To cite this version:

Gotzon Aurrekoetxea. El Atlas Lingüístico Vasco (EHHA). 1, M. Aurnague, M. Roché (dir.) - Editions Atlantica, pp.63-71, 2002. artxibo-00071353

HAL Id: artxibo-00071353

https://artxiker.ccsd.cnrs.fr/artxibo-00071353

Submitted on 23 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Aurrekoetxea2002a
Argitaratua in: M Aurnague / M. Roché (Dir.), *Hommage À Jacques Allières-1 Domaines Basque Et Pyrénéen,* Atlantica, Biarritz (2002), 63-71

El Atlas Lingüístico Vasco (EHHA)

Gotzon Aurrekoetxea
Universidad del País Vasco

1. Introducción

Durante la confección y los trabajos de recogida de datos del Atlas Lingüístico Vasco (en adelante EHHA), nos hemos preguntado en infinidad de ocasiones sobre el sentido de una recogida de datos de esta envergadura, o sobre el alcance de un altas lingüístico.

Pensamos que el objetivo último de una recogida de datos tan amplia no puede ser únicamente la publicación en forma atlas lingüístico. La explotación de los datos recogidos a de ser más amplia, para que pueda ser una fuente de información tanto para fonetistas, como para los fonólogos, "gramáticos" interesados en morfología o aspectos sintácticos, lexicógrafos, etnólogos, antropólogos...

No podemos quitarnos de la mente la poca o nula utilización que ha tenido hasta nuestros días la inmensa mayoría de los atlas lingüísticos. Son conscientes de ello, incluso los mismos autores. Así podemos citar al maestro de los dialectólogos tolosanos J. Séguy (1973: 13) o al hispanista M. Alvar (1984: 112-113) cuando afirmaba:

"Los atlas abruman por la inmensa cantidad de datos que encierran y si no facilitamos su consulta, quedarán como gigantescos archivos no utilizados. Desgraciadamente, tal es la situación en el mundo hispánico".

Pero no son ni los únicos, ni los primeros ni los últimos en lamentarse de tal situación. A decir verdad, la infrautilización de los atlas se deriva de la dificultad de su consulta: datos transcritos únicamente en alfabeto fonético (en muchos casos diferente en cada uno de ellos), falta de elaboración lingüística de los datos, falta de elaboración cartográfica de los mismos, falta de índices de formas obtenidas, etc.

Tanto la elaboración lingüística de los datos recogidos como la cartografía utilizada en los atlas lingüísticos necesita una profunda

renovación. Así lo ha entendido el equipo del EHHA partiendo de unos presupuestos claros:

a) Los datos no se publicarán solamente en alfabeto fonético; necesitan una elaboración lingüística.
b) Los mapas deben tener una elaboración cartográfica.
c) El tamaño de la publicación en formato libro debe de ser sensiblemente más reducido.
d) Debe conllevar la publicación de índices de formas, en aras a una consulta rápida y relativamente cómoda de las respuestas obtenidas.

2. Elaboración lingüística de los datos

El primer aspecto nos conduce a una elaboración lingüística de los datos. El equipo del atlas vasco ha profundizado mucho en la elaboración lingüística de los datos, entendiendo ésta como una interpretación. Creemos totalmente necesaria una interpretación de los datos para que las personas que accedan a la información tengan a mano la más certera interpretación sin tener que aventurar una.

Entendemos que la persona más indicada para hacer una correcta interpretación es aquella que ha tomado parte en la recogida de los datos. Nadie sabe mejor el estado en que se encuentra el material recogido. Esta necesaria interpretación, a veces difícil, tiene sus riesgos, pero éstos se acentúan aún más cuando se deja en manos de personas sin contacto directo con los datos. Por ello propugnamos que los datos recogidos no sean publicados sin una interpretación.

Somos conscientes, no obstante, de las reticencias de muchos dialectólogos autores de atlas lingüísticos a aceptar propuestas de este estilo. Pero la dialectología se debe adaptar a los tiempos y ponerse a la altura de las demás ciencias a la hora de ofrecer los datos.

El EHHA ha optado por una dobe publicación de los datos: por una parte publica las respuestas en alfabeto fonético, tal y como han sido transcritas. Por otra parte publica los datos elaborados con una interpretación lingüística y otra posterior cartográfica.

2.1. Los datos fonéticos

Las respuestas obtenidas en alfabeto fonético, se publican en listas ordenadas por provincias o regiones administrativas y por localidades encuestadas ordenadas alfabéticamente.

La transcripción de los datos se realiza utilizando el alfabeto fonético IPA, adaptado a la realidad de la lengua vasca.

Los datos fonéticos son necesarios para fonólogos y fonetistas que requieren la máxima fidelidad del habla transcrita. Los atlas lingüísticos, nacidos en épocas en las que quizás el aspecto fonético primaba en las investigaciones lingüísticas, han tenido hasta nuestros días una dependencia absoluta de la transcripción fonética.

Pero la lingüística no se circunscribe únicamente a la fonética y fonología. Otros campos como la lexicología, sintaxis, etc. trabajan con datos transcritos en alfabetos ortográficos o normativos.

Por ello en el EHHA los datos transcritos e informatizados fonéticamente son objeto de una transformación gradual que va desde una acomodación ortográfica a una modalización o formalización en lemas y superlemas (ver gráfico), de tal forma que se hallen dispuestos a ser explotados desde diversos puntos de vista y en distintas disciplinas.

2.2. Datos en alfabeto ortográfico

El punto de partida de toda la interpretación lingüística de las respuestas obtenidas es su forma fonética. La transformación de los datos fonéticos a ortográficos se realiza por medio de una tabla de transliteración que se halla incrustada en la base de datos, de tal forma que la introducción o corrección de una información fonética es guardada automáticamente en alfabetos y lugares diferentes.

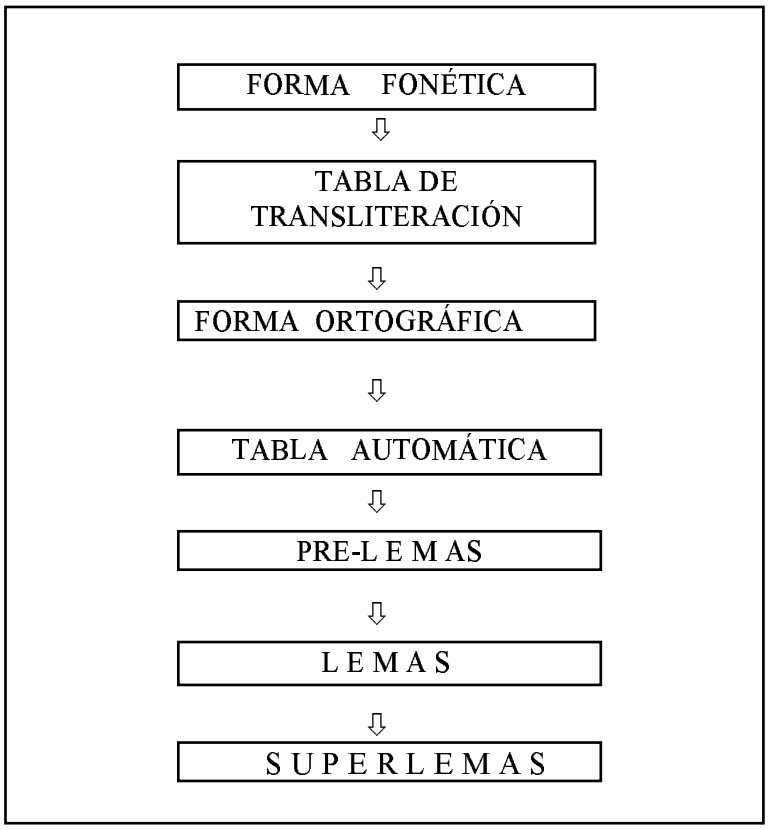
Esta transformación implica una pérdida de información. Pero lo que se pierde en exactitud fonética se gana en comodidad de consulta. Los rasgos que se pierden (nasalización de las vocales, distinción entre la pronunciación normal y cuidada, distintas realizaciones de aberturas vocálicas, relajamiento de los sonidos, sonorización de las sibilantes, etc.) son rasgos no pertinentes fuera del campo de la fonética.

De esta forma las respuestas pueden ser consultads tanto en alfabeto fonético como en ortográfico.

2.3. Pre-lemas

Estos datos ortográficos, aunque cómodos para la lectura, presentan, sin embargo, dificultades para su consulta, por la gran cantidad de variantes que presentan. Supongamos que queremos saber si en el EHHA ha recogido la palabra "zizare" para el concepto 'lombriz / ver de terre (Lumbricus terrestris)'. La consulta automática nos presentará la palabra "zizare" y las localidades en las que se ha recogido. Pero no se nos aparecerán variantes como "zizere", "zizari", "zizeri", "zizai" u otras parecidas. Falta la elaboración de las respuestas para consultas ya sea en una base de datos, ya sea en el index de un libro.

ESQUEMA DE LA ELABORACIÓN DE LOS DATOS



Esta elaboración lingüística comienza con la tabla automática por la que se filtran los datos transcritos en alfabeto ortográfico. Se trata de un procedimiento que, instalado en la base, despoja a las respuestas de las características que no importan en la elaboración posterior, tomando así el aspecto típico del lenguaje escrito. Son los pre-lemas, es decir, datos transcritos siguiendo las convenciones al uso de la escritura estandard.de la lengua vasca, excepto en algún detalle.

Resumiendo, la informatización de una respuesta en alfabeto fonético conlleva indirectamente, sin ninguna participación del lingüísta, una forma ortográfica lograda por medio de una tabla de transliteración y una forma estandar, mediante el filtro de la tabla automática.

2.4. Lemas

Si hasta este punto las distintas versiones de las respuestas han sido fruto de unas artimañas automáticas que toda palabra tiene que soportar, el siguiente paso, en cambio, es responsabilidad directa del investigador.

Los lemas se eligen atendiendo a criterios y procedimientos parejos utilizados en trabajos de lexicología: misma etimología, despojo de vestigios gramaticales (morfemas de declinación)... Criterios que no han sido utilizados en sentido restrictivo. Es decir, el dialectólogo no actúa como lo hace un lexicólogo al decidir una entrada de diccionario, sino que trata de conservar los rasgos más interesantes desde el punto de vista de la geolingüística.

Más aún, tiene que actuar a veces teniendo en cuenta el punto de vista de otras disciplinas interesadas en los materiales y preguntarse por la forma más idónea para que sean accesivibles y utilizables.

Nuestra preocupación por la consulta automática de la base de datos es constante. La lematización facilita la consulta de las distintas variantes de una palabra. En una consulta del tipo léxico, por ejemplo, si lo que interesa saber es si un ítem lexical ha sido recogido o no, la forma más simple sería comenzando por la forma escrita estándar de ese vocablo. La respuesta de la base debería proporcionarnos si tal vocablo ha sido recogido y en cuántas o en qué localidades.

Los lemas son formas, atestiguadas o no, que se presentan en lenguaje natural, utilizando variantes que pueden ser verbalizados por los informantes. Es decir, desprovistas de todo signo gráfico que indique cierto grado de abstracción.

2.5. Superlemas

La elección de los superlemas obedece a criterios lingüísticos, pero está fuertemente condicionado a la cartografía diseñada para el EHHA.

Los superlemas son la base de la cartografía diseñada para el atlas vasco. A cada uno se le asigna un color. Como el máximo de colores de cada mapa es 16, los superlemas por cada pregunta no podrán exceder de 16. Limitación que a más de uno se le antojará bondadosa, pero obliga en más de una ocasión a elevar el grado de abstracción.

Los superlemas pueden, y de hecho la gran mayoría lo son, presentar formas del lenguaje natural, como los lemas. Es decir, formas lingüísticas que son o pueden ser verbalizadas por los informantes. Pero cuando la variedad idiomática es grande el dialectólogo puede echar mano de formas más abstractas utilizando signos gráficos, para recoger en una misma forma variantes fonológicas o morfológicas, especialmente en palabras compuestas.

Los signos gráficos utilizados en la creación de los superlemas son el paréntesis, le guión, el signo "+" y el signo "/". Por lo tanto, los superlemas pueden presentar aspectos diferentes::

2.5.1. Superlema en lenguaje natural. Es decir, una variante que puede o no ser atestiguada y que alberga el aspecto de una entrada de diccionario.

2.5.2. Una forma abstracta que engloba más de un variante

2.5.2.1. Si las variantes se distinguen por la presencia o no de un sonido, y si se quieren incluir en un superlema, se utiliza el signo gráfico de la paréntesis. Supongamos que para el concepto 'saltamontes / sauterelle' se han recogido diversas palabras; un tipo de ellas lo consfiguran los lemas *ote* y *othe.* Si el editor se encuentra ante la necesidad de agrupar las variantes citadas puede optar por la forma *ot(h)e* como superlema.

2.5.2.2. Si las variantes se distinguen por la presencia o no de un morfema se utilizan los signos **(-).** En este caso dicho superlema alberga lemas que pueden estar constituidas únicamente por un lexema base únicamente o por un lexema base y morfema suplementario; es decir, se encontrarán respuestas o variantes simples y compuestas. Un ejemplo de lo que estamos hablando se encuentra entre las respuestas dadas al concepto 'telaraña / toile d'araignée'. Un tipo de respuestas han sido agrupadas en los siguientes lemas: *aramu, aramontoki, aramusara, aramosara, aramosare, aremusara* y *aramutela.* El editor se ha visto en la necesidad de recoger todos ellos en un único superlema por la gran

variedad de formas recogidas. Para ello ha tenido en cuenta dos factores: el lexema base *(aramu, aramon-, aramo-, aremu-, aramu-)* y por otra la posible aparición de un morfema suplementario, que es triple en este caso *(-sare / -sara, -toki* y *-tela).* La posibilidad que un superlema pueda tener o no un morfema suplementario se expresa por medio de "**(-)**". En este caso el superlema será **aramu (-)**, con lo cual queremos indicar que dicho superlema engloba a respuestas que tienen solamente el lexema base y a respuestas que contienen dicho lexema base más otro morfema.

Este signo gráfico puede aparecer tanto detrás como delante del lexema base. La misma pregunta recoge dos respuestas que han sido recogidas en dos lemas *(telaraña y arañetela).* Lemas que han sido el punto de partida del superlema **(-) araña (-).**

2.5.2.3. Otro tipo de respuestas es la siguiente: un morfema base (idéntico en todas las respuestas) más varios morfemas suplementarios (todas las respuestas deben contener alguno, pero no siempre es el mismo). Estos morfemas pueden situarse en algunos casos delante y en otros detrás. Un caso de este tipo de respuestas se encuentran en la pregunta 'libélula / libellule' en variantes que contienen el morfema **infernuko** *(inpernuko zaldi, infernuko errexin).* Si el editor estima que las dos respuestas deben formar parte del mismo superlema tiene a mano el artilugio **(+)**, creando el superlema **infernuko (+)**, que indica que en las dos respuestas aparece el morfema "infernuko" y que dicho morfema está acompañado en los dos casos por un morfema diferente.

2.5.2.4. Una cuarta posibilidad de crear un superlema es integrar en la misma las dos variantes de una respuesta. En este caso se utiliza el signo "/". Por ejemplo, en la pregunta 'aguijón / dard' se han recogido entre otras las respuestas *lantza* y *lantzet*. Estas dos respuestas se integran en un mismo superlema, **lantza/et**.

2.5.3. La limitación a 16 superlemas, imprescindible para la utilización racional de los colores, se convierte en un obstáculo en aquellas preguntas que presentan respuestas que no se pueden englobar en 16 tipos o superlemas.

Era una posibilidad un tanto remota a primera vista, pero que la gran dialectalización de la lengua vasca hace posible en más ocasiones de las que hubieramos deseado. En estos casos se ha recurrido a tipologizar un superlema, denominado **bestelakoak** (otros), que aparecerá en aquellas preguntas que conocen más de 16 superlemas. En la mayoría de los casos son respuestas de una aparición, respuestas dudosas, respuestas de una tipología

especial, préstamos léxicos no asimilados... En todos los casos el superlema **bestelakoak** aparece desgranado en las notas al mapa correspondiente.

3. La cartografía lingüística

Los mapas del EHHA suponen una renovación en la cartografía atlantográfica. De mapas con respuestas fonéticas punto a punto, se pasa a mapas de áreas que contienen una interpretación lingüística y cartográfica.

Esta cartografía posibilita un formato más propio de lo que entendemos por un libro y mucho más manejable que los habituales en este tipo de publicaciones. A pesar del formato y de la cartografía elegida cada localidad puede recibir los atributos de 2 respuestas y de una respuesta propuesta[1].

Somos conscientes de los reparos que provoca este tipo de mapas en muchos dialectólogos. Pero creemos imprescindible una interpretación de los datos para una consulta atractiva, rápida y cómoda. Entendemos que la cartografía punto a punto se halla desfasada actualmente y no aconsejable en la mayoría de los casos; su lectura resulta engorrosa y dificultosa, obligando al lector a hacer una interpretación si desea una lectura geolingüística de los datos. La interpretación del editor tiene a nuestro entender más posibiliddes que ser la más idónea de las posibles, por el conocimiento que tiene de la información.

Se trata de mapas de áreas lingüísticas, representadas cada una de ellas por colores. La percepción humana de los colores es limitada. Según los expertos la vista humana presenta dificultades para distinguir más de 16 o 20 colores. Esta limitación humana nos ha llevado a determinar un máximo de colores (16) a utilizar en los mapas.

4. La publicación de los datos

La publicación del atlas vasco se llevará a cabo en dos formatos: formato libro y CD-ROM.

4.1. EHHA-libro. La publicación en soporte papel constará de dos cuerpos diferenciados: respuestas y mapas por una parte; los índices por otra. Los

---

[1] El EHHA ha recogido respuestas producidas por el informante y respuestas propuestas por el encuestador (Aurrekoetxea, 1986), siguiendo los procedimientos utilizados en ALG (Ravier, 1965).

datos de cada pregunta se publican en dos hojas enfrentadas: en las hojas pares se ubican las respuestas en alfabeto fonético ordenadas por regiones administrativas y localidades. Las hojas impares albergan los mapas creados sobre los datos elaborados. La obra incluye dos tipos de índices: índice de superlemas (mapa a mapa), con los lemas y las respuestas correspondientes; índice de todas las respuestas en alfabeto ortográfico ordenados alfabéticamente, con indicación del mapa correspondiente.

4.2. EHHA-CD. El soporte informático incluye por una parte toda la información del EHHA-libro. Pero incluye también toda la base de datos utilizada en la confección de los mapas.

Esta base de datos reúne una información variada y abundante, en dos formatos: campos codificados y en formato texto. La información de los campos codificados es básicamente la que se incluye en formato libro. En la información contenida en los textos se encuentran definiciones dadas por los hablantes, refranes, narraciones, cuentos y canciones populares, creencias, descripciones de diversos objetos, en fin, todo el saber popular que se incluye bajo la denominación de "cultura tradicional".

Pero el no será un CD solo de consulta de los datos elaborados por el equipo del EHHA, el usuario dispondrá también de la posibilidad de crear nuevos mapas ateniendo a criterios personales.

Bibliografía

Alvar, M. (1984). Automatización de índices en los atlas lingüísticos, in: M. Alvar, *Informática y Lingüística,* Málaga: Ágora, p. 107-119.

Aurrekoetxea, G., (1986), Euskal Herriko Hizkuntz Atlasa (EHHA): inkesta metodologia eta ezezko datuak, *Euskera* 31-2, p. 413-424.

Ravier, X. (1965), Le traitement des données négatives dans l'Atlas linguistique et ethnographique de la Gascogne, *Revue de Linguistique Roman* XXIX, 262-274.

Séguy, J. (1973), La dialectométrie dans l'Atlas linguistique de la Gascogne, *RLiR* 37, p. 1-24.