

Supplements of the Anuario de Filología Vasca "Julio de Urquijo", LIII

TOOLS FOR LINGUISTIC VARIATION

Gotzon Aurrekoetxea, Jose Luis Ormaetxea
(eds.)

eman ta zabal zazu



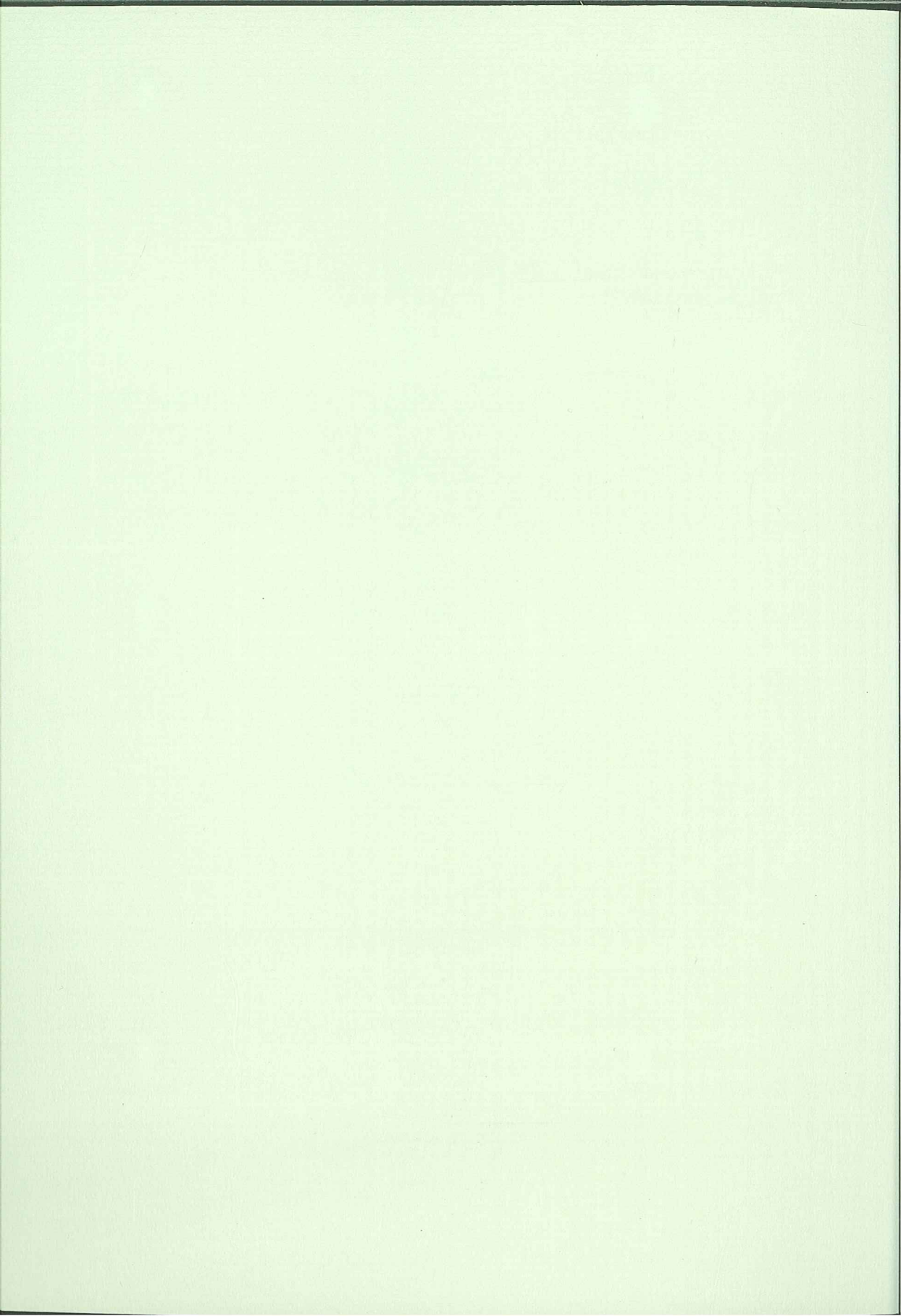
Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Bilbao

Bilbo

2010



EL PROCESAMIENTO INFORMÁTICO DE LOS MATERIALES DEL *ATLAS LINGÜÍSTICO DE LA PENÍNSULA IBÉRICA* DE TOMÁS NAVARRO TOMÁS

Pilar García Mouton
ILLA-CCHS (CSIC)

Abstract

Estas páginas dan noticia de la puesta en marcha de un proyecto intramural del CSIC para elaborar y editar definitivamente, en soporte informático, los materiales del Atlas Lingüístico de la Península Ibérica, que Tomás Navarro Tomás dirigió en la primera mitad del siglo XX.

These pages announce the launch of the internal CSIC project to produce and definitively edit, in digital format, the materials from the Linguistic Atlas of the Iberian Peninsula, directed by Tomás Navarro Tomás during the first half of the 20th century.

Palabras clave: *Procesamiento informático del ALPI, Geolingüística peninsular s. XX.*

Key words: *ALPI Information Technology, Peninsular Geolinguistic XXth century.*

1. Introducción

Redacto estas páginas en calidad de coordinadora del proyecto intramural del Consejo Superior de Investigaciones Científicas (CSIC)¹ para elaborar y editar los materiales del *Atlas Lingüístico de la Península Ibérica* (ALPI).

Los colegas conocen bien las características de este atlas, que fue un proyecto emblemático del Centro de Estudios Históricos de la Junta para la Ampliación de Estudios, dirigido por Tomás Navarro Tomás e impulsado por Ramón Menéndez Pidal, en la línea de los atlas de gran dominio que se plantearon en Europa en la primera mitad del siglo XX. Desde que se tomó la decisión de llevarlo a cabo hasta que empezaron los trabajos de campo pasaron muchos años, pero las encuestas del ALPI ya estaban prácticamente terminadas en las partes correspondientes a las hablas astur-leonesas, aragonesas, castellanas y catalanas cuando la Guerra Civil española las interrumpió.

¹ Proyecto intramural del CSIC, de referencia 200410E604; Investigadora principal: Pilar García Mouton; Título: *Elaboración y edición de los materiales del Atlas Lingüístico de la Península Ibérica* (ALPI).

Para la zona castellanohablante los encuestadores fueron Aurelio M. Espinosa hijo, Lorenzo Rodríguez Castellano, Aníbal Otero y Manuel Sanchis Guarner; para la zona gallegoportuguesa, Aníbal Otero y Rodrigo de Sa Nogueira, sustituido primero por Armando Nobre de Gusmão y, después, por Luís F. Lindley Cintra, y para la zona de hablas catalanas, Francesc de Borja Moll y Manuel Sanchis Guarner.

Años después, acabadas las encuestas pendientes, el CSIC publicó un tomo de *Fonética* con 75 mapas en el año 1962, pero después los trabajos de edición se abandonaron hasta hoy. En los últimos años David Heap, que había localizado los materiales de encuesta, fue colgando fotocopias de ellos en internet, y de esa manera les volvió a dar vida en cierta forma.

2. Nuestro proyecto de elaborar y editar los materiales del ALPI

En el verano del año 2007 el CSIC se planteó la posibilidad de reiniciar la edición del ALPI y decidió encargar un nuevo proyecto para elaborar y editar los materiales del atlas a un equipo, coordinado por Pilar García Mouton (CSIC), integrado por Inés Fernández Ordóñez (Universidad Autónoma de Madrid), David Heap (Universidad de Western Ontario), María Pilar Perea (Universidad de Barcelona), João Saramago (Centro de Lingüística de la Universidad de Lisboa) y Xulio Sousa (Instituto da Língua Galega de la Universidad de Santiago de Compostela). El CSIC asumió como propio el proyecto en el mes de marzo de 2010, como uno de sus proyectos intramurales, después de haberlo sometido a una estricta evaluación internacional. El resultado previsto no mantendrá la forma del primer volumen del ALPI, en soporte papel, porque se concibe como un atlas interactivo de libre acceso que se podrá consultar desde un geportal del CSIC.

2.1. Cuestiones metodológicas

Es evidente que desde que se hicieron las encuestas del ALPI hasta ahora han cambiado muchas cosas, pero aquí resulta especialmente apropiado referirse a dos de ellas. En primer lugar, han cambiado radicalmente la cultura y las hablas de los territorios estudiados y, en segundo lugar, también ha cambiado mucho la metodología geolingüística en general. Ahora bien, ninguna de las dos circunstancias merma el interés del ALPI, más bien al contrario: los atlas lingüísticos se refieren siempre a una época determinada que no tiene por qué coincidir con la de quienes los utilizan; lo contrario los convertiría en obras casi efímeras. De ahí que sean precisamente las fechas en las que se recogieron sus materiales las que hacen del ALPI un archivo histórico fiable y segmentado. Como ya señaló el mismo Navarro Tomás (1975: 14) en el prólogo a su libro *Capítulos de geografía lingüística de la Península Ibérica*, “Noticia histórica del ALPI”: “Por virtud principal de su información fonética, el ALPI es como una especie de acta documental del carácter y fisonomía del habla popular de la Península en los años inmediatamente anteriores a la guerra civil. La honda conmoción producida por esta guerra en todo el país, y el movimiento de población ocasionado después por motivos económicos y sociales, habrán modificado sin duda alguna las líneas del ALPI, lo cual acentúa su interés como testimonio de valor histórico.”

La segunda circunstancia, la relativa a la actualización de su metodología, sólo sería exigible a un atlas que se planteara hoy, no a uno ideado a principios del siglo xx. Con algunas características propias, el ALPI es un atlas equiparable a los similares de su época. Y, finalmente, no conviene olvidar que, a estas alturas, no disponemos de ningún otro atlas de gran dominio peninsular en el que se puedan contextualizar los distintos atlas zonales para establecer comparaciones productivas. En este sentido, buscando algún otro paralelo, se podría aducir que en las últimas décadas se están publicando el *Atlante Linguistico Italiano* (1995-1999), cuyas encuestas se hicieron después de la primera gran guerra; acaba de aparecer el tomo IV del *Atles Linguistic del Domini Català* de Joan Veny y Lúdia Pons i Griera (2009), a partir de unas encuestas relativamente lejanas en el tiempo; se están digitalizando otros atlas históricos, y algunos colegas, como Hans Goebel (2002), revisitan con metodología dialectométrica el *Atlas Linguistique de la France* (ALF) o el clásico atlas italo-suizo, el *Sprach- und Sachatlas Italiens und der Südsweiz* (AIS).

Además de estas razones científicas, existen otras de tipo ético que el CSIC ha tenido en cuenta a la hora de apoyar este proyecto, fundamentalmente el compromiso de poner a disposición de la comunidad científica y de los hablantes en general un patrimonio que les pertenece y saldar así una deuda histórica con unos investigadores que pusieron un gran esfuerzo profesional en los trabajos de este atlas (García Mouton 2007).

2.2. Descripción del proyecto

Nuestro proyecto tiene como objetivo elaborar y editar los materiales del ALPI utilizando las posibilidades que la tecnología informática proporciona actualmente a los trabajos geolingüísticos, no sólo para el cartografiado automático, sino también para realizar búsquedas de todo tipo en el corpus completo de datos lingüísticos y etnográficos que conforman un atlas de estas características. Se trata de hacerlas sobre un fondo georreferenciado que cubra las necesidades fundamentales de cualquier proyecto geolingüístico. Una vez concluido, se podrán realizar búsquedas más o menos avanzadas y obtener resultados puntuales, pero también mapas —mapas clásicos (con la respuesta recogida en cada uno de los puntos de encuesta), mapas por áreas o por isoglosas, mapas simbólicos— y cualquier otro tipo de elaboración que esté previsto en la herramienta informática que se está diseñando para tratar los datos y en la estructura del geoportal que los va a acoger. De este modo, cabe esperar que las mismas circunstancias que supusieron un grave retraso en la publicación del ALPI a la larga acaben beneficiando la difusión de estos materiales históricos, que finalmente se podrán consultar en la forma más novedosa.

Como se ha dicho, la edición del atlas está planteada casi exclusivamente en soporte informático a través de la red, a partir de una base de datos georreferenciada alojada en un servidor del CSIC, con las posibilidades que los actuales Sistemas de Información Geográfica (SIG) proporcionan para convertir la futura web del ALPI en una IDE (Infraestructura de Datos Espaciales). Esto significa que sus resultados podrán ser confrontados con otros cualesquiera que hayan sido tratados de forma similar. El soporte papel se reservará, en todo caso, para algunos mapas que puedan servir testimonialmente como muestras representativas institucionales.

2.3. Plan de trabajo

Se trabajará con los cuestionarios correspondientes a los 527 puntos de encuesta del ALPI, que se distribuyen así: 53 para Galicia; 93 para Portugal; 78 para Asturias, León y Extremadura; 90 para las dos Castillas y Albacete; 71 para Andalucía y Murcia; 40 para Navarra y Aragón; y 104 para Andorra, el Rosellón, Cataluña, Valencia y las islas Baleares. Cada cuestionario está dividido en un *Cuaderno I*, con unas preguntas sobre el informante y la localidad y unas notas de orientación fonética, más el cuerpo del cuestionario propiamente dicho, con 411 cuestiones de fonética, morfología y sintaxis, y un *Cuaderno II*, con cuestiones de léxico que alcanzan hasta la pregunta 828 en la versión sintética IIG, mientras que la versión IIE amplía muchas de ellas con varias subpreguntas.

2.3.1. Digitalización y herramienta para introducir datos

Para elaborar los materiales se están abordando actualmente dos cuestiones previas: la digitalización de los cuestionarios —que se ha hecho básicamente en el Instituto da Lingua Galega (ILG) de la Universidad de Santiago de Compostela bajo la responsabilidad de Xulio Sousa— y el diseño de una herramienta informática específica —que estamos elaborando en el Centro de Ciencias Humanas y Sociales del CSIC con el apoyo de la Unidad TIC (Tecnologías de la Información y la Comunicación), especialmente con el de su responsable, Juan Carlos Martínez, y el de Ángel Díaz del Castillo, y la ayuda de la Unidad SIG (Sistemas de Información Geográfica), especialmente la de su responsable, Isabel del Bosque, y la de Carlos Fernández Freire—. La digitalización está prácticamente resuelta, algo que estimamos fundamental desde el primer momento porque va a proporcionar imágenes de los materiales de mucha calidad para trabajar, con la ventaja de tenerlas permanentemente disponibles en pantalla y de poder ampliarlas, unas imágenes que resultan imprescindibles por las características específicas de este proyecto.

Al encarar el proceso de elaboración de los materiales, había que partir del hecho de que los encuestadores del ALPI utilizaron muy a conciencia el alfabeto fonético de la *Revista de Filología Española* (AFE), un alfabeto creado por el propio Tomás Navarro Tomás, publicado en la revista en 1915, y que hoy cuenta con una larga tradición en el mundo hispánico. Ese alfabeto reflejaba el interés fonético de su autor, y de su época, con un lujo de detalles que hoy puede parecer prolijo en exceso y, por ese motivo, puede reducir su consulta por parte de los especialistas que no están familiarizados con él.

A la hora de abordar la edición del ALPI, se planteaba un dilema científico: conservar la estricta transcripción que tanto preocupó a su director y sus encuestadores o intervenir, simplificándola, para hacerla más asequible y volcarla al Alfabeto Fonético Internacional (AFI), lo que aseguraría al ALPI una difusión mayor y posibilitaría la recuperación ágil de los datos en un sistema de búsqueda complejo. Creemos haber encontrado una solución equilibrada que va a permitir conservar la transcripción fonética original de los encuestadores —porque siempre se dará la posibilidad de acceder a la imagen digitalizada de las respuestas originales manuscritas— y conseguir búsquedas ágiles a partir de su conversión al Alfabeto Fonético Internacional, que facilitará la elaboración de los datos.

La cuestión del respeto a la transcripción original no deja de ser básica. Las siguientes palabras de Navarro Tomás (1975: 19) prueban la importancia que los investigadores del ALPI daban a su minuciosa transcripción: “El valor demostrativo de estos mapas [...] es mérito de la detallada precisión de sus transcripciones. Una investigación realizada con menos rigor analítico y con transcripción menos estrecha, habría resbalado por encima de muchos de estos pormenores tan significativos para el cabal conocimiento de la materia. [...]. Los aparatos mecánicos sólo prestan una ayuda relativa. El oído convenientemente ejercitado sigue siendo el instrumento más perfecto. Muchas veces, al repasar en la oficina una cinta magnetofónica, se echa de menos la presencia de la persona que la inscribió.”

2.3.2. *La simplificación de las transcripciones originales*

Antes que nada hay que tener en cuenta que los mismos editores del tomo I del ALPI, de acuerdo con Navarro Tomás, decidieron proceder desde el principio a una primera simplificación de las transcripciones de sus encuestas. Por el *Epistolario* que acaban de publicar Vicent García Perales y Santi Cortés (2009: cartas 135, 138, 157, etc.) sabemos que existió una “plantilla” elaborada con la ayuda del maestro, plantilla que trajeron de su viaje a Nueva York Manuel Sanchis Guarnier y Lorenzo Rodríguez Castellano y que no se ha conservado. La única solución pasaba por tratar de reconstruirla confrontando la transcripción de los cuestionarios y los resultados impresos. Decidimos trabajar después en otra etapa más en la simplificación, para establecer una tabla de equivalencias entre los signos empleados con los del AFI, el alfabeto en el que se van a introducir los datos. Para poder hacerlo con garantías, antes de implementar la tabla de equivalencias fonéticas en la herramienta informática, se pidió ayuda a varios especialistas en la fonética de los distintos dominios lingüísticos —Amélia Andrade, para las hablas portuguesas; Francisco Dubert, para las gallegas; Ralph Penny, para las hablas asturleoneras; Daniel Recasens, para las catalanas y Juana Gil, para las castellanas— que trabajaron un tiempo con los materiales. Los días 7 y 8 de mayo de 2009 el equipo de nuestro proyecto se reunió con ellos, gracias a la hospitalidad de la Universidad Autónoma de Madrid,² y allí quedó establecida la correspondencia que se ha decidido adoptar entre los signos fonéticos del Alfabeto de la *Revista de Filología Española* de los originales y los signos del Alfabeto Fonético Internacional, que es mucho más sencillo. Partimos de la base de que, como elaboradores y editores del ALPI, no entramos a juzgar unas transcripciones heredadas, ni intervenimos en la transcripción original más que para establecer las equivalencias con el AFI, en nuestro interés de evitar que se pueda llegar a perder cualquier matiz imprescindible para la caracterización de las hablas estudiadas. Publicaremos con el atlas la tabla de equivalencias y, de todas formas, quienes consulten el ALPI, si lo desean, podrán *ver* literalmente la transcripción original de cada respuesta en cada uno de los cuestionarios.

² Gestionada por Inés Fernández Ordóñez.

2.3.3. *El diseño de la herramienta para la introducción de los datos*

Hecho lo anterior, empezamos el diseño de la herramienta informática —que parte de una base de datos relacional— para introducir los datos. La herramienta, aún sin terminar, está bastante avanzada: incluirá la imagen digitalizada de las respuestas en transcripción AFE, los distintos teclados que permitirán introducir esas respuestas en transcripción AFI —con una ayuda permanente—, la grafía en las distintas lenguas iberorrománicas, las imágenes y los contenidos etnográficos, y tendrá funciones para la premarcación fonética, morfológica, sintáctica y léxica previsibles en cada una de las cuestiones. En otra vertiente, dispondrá de controles de acceso personalizados y jerarquizados para los distintos colaboradores —ya que se trabajará en ella a través de la red desde distintos sitios—, y también de funciones para optimizar el control y la corrección del trabajo.

2.4. **Lo etnográfico en el ALPI**

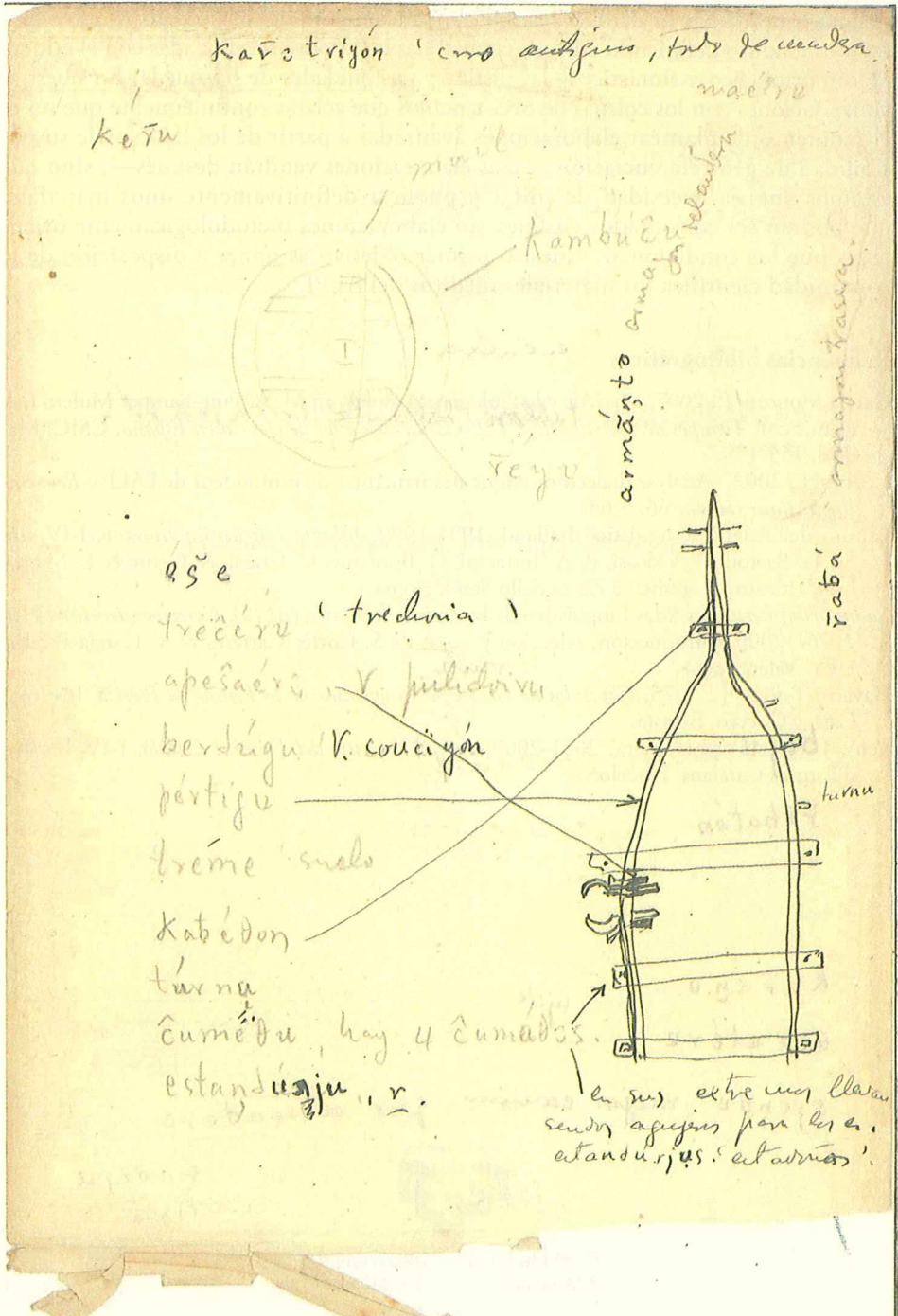
Conviene advertir que el proyecto de edición tiene muy en cuenta el hecho que, junto a las respuestas, en los cuestionarios aparecen con alguna frecuencia informaciones complementarias sobre cultura material y espiritual que los encuestadores anotaron a lo largo de la encuesta, a veces en forma gráfica, con dibujos sobre enseres o utensilios (ver figura en la página siguiente). Está previsto dar un espacio adecuado en nuestro proyecto a estos datos etnográficos, ya que, además de su valor intrínseco, para la mayor parte de la Península, son los últimos recogidos por especialistas antes de la Guerra Civil, circunstancia que los convierte en un testimonio excepcional de una cultura que ya no existe.

De nuevo la voz de Navarro Tomás (1975: 12-13) defiende el peso de estos contenidos en el ALPI, que, en ese sentido, fue acusado de ser un atlas desfasado, meramente fonético. Defiende Navarro la atención que habían prestado a lo etnográfico, aunque sus palabras sólo afirmen que el cuestionario seguía una organización temática similar a la del AIS: “Para la sección de léxico fue de gran ayuda el Atlas italo-suizo de Jaberg y Jud, cuyos volúmenes empezaron a aparecer por esa fecha. Adoptamos su organización por temas etnográficos siguiendo el orden de fenómenos atmosféricos, accidentes geográficos, flora, fauna, cuerpo humano, familia, hogar, labores agrícolas, oficios artesanos, herramientas, animales domésticos, etc. Sobre esta base, el ALPI hubiera podido llamarse Atlas lingüístico y etnográfico, como de hecho lo es, aunque no pareciera indispensable indicarlo en el título.”

2.5. **La estructura del geoportal**

Otra parte del proceso de elaboración y edición será la dedicada al diseño de sistemas que faciliten el volcado automático de datos para responder a las búsquedas en línea del geoportal, que permitirán hacer mapas a la carta, búsquedas de todo tipo sobre los datos, enlaces con las referencias previstas a otros mapas sobre los mismos *items* de cualquier otro atlas, acceder a la traducción de las cuestiones a las principales lenguas de cultura, etc.

Estamos aún en las primeras fases. Nos hubiera gustado aportar —como pensamos en nuestros planteamientos más optimistas— el diseño definitivo de la herra-



mienta para introducir datos y avanzar el planteamiento del resto, pero en estos momentos estamos definiendo cuidadosamente las cuestiones relacionadas con el etiquetado morfológico y sintáctico y las distintas posibilidades de búsqueda. En nuestras conversaciones con los colegas de SIG tenemos que aclarar continuamente que no es el momento de plantear elaboraciones avanzadas a partir de los datos y de su posibilidad de georreferenciación —esas elaboraciones vendrán después—, sino que estamos ante la necesidad de editar y publicar definitivamente unos materiales que podrán ser condicionados en línea sin elaboraciones metodológicamente orientadas que los condicionen. Nuestro primer objetivo es poner a disposición de la comunidad científica los materiales inéditos del ALPI.

Referencias bibliográficas

- García Mouton, P., 2007, «La JAE y la filología española», en M. Á. Puig-Samper Mulero (ed. científico), *Tiempos de investigación. JAE-CSIC, cien años de ciencia en España*, CSIC, Madrid, 155-159.
- Goebel, H., 2002, «Analyse dialectométrique des structures de profondeur de l'ALF», *Revue de linguistique romane* 66, 5-63.
- Istituto dell'Atlante Linguistico Italiano, 1995-1999, *Atlante Linguistico Italiano*, I-IV, dir. M. G. Bartoli, G. Vidossi, B. A. Terracini, G. Bonfante, C. Grassi, A. Genre & L. Massobrio, Istituto Poligrafico e Zecca dello Stato, Roma.
- La historia interna del Atlas Lingüístico de la Península Ibérica (ALPI). Correspondencia (1910-1976)*, 2009, introducción, selección y notas de S. Cortés Carreres & V. García Perales, UPV, Valencia.
- Navarro Tomás, T., 1975, *Capítulos de Geografía Lingüística de la Península Ibérica*, Instituto Caro y Cuervo, Bogotá.
- Veny, J. & L. Pons i Griera, 2001-2009, *Atles Lingüístic del Domini Català*, I-IV, Institut d'Estudis Catalans, Barcelona.

Supplements of the Anuario de Filología Vasca "Julio de Urquijo", LIII

TOOLS FOR LINGUISTIC VARIATION

Gotzon Aurrekoetxea, Jose Luis Ormaetxea
(eds.)

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Bilbao Bilbo

2010

