

Cartografiado automático y bases de datos geolingüísticas

*Francisco Moreno Fernández
José Enrique Moreno Fernández
Antonio García de las Heras*

Universidad de Alcalá

Atlante es una aplicación informática desarrollada por la Universidad de Alcalá (España) para el cartografiado automático y la gestión de bases de datos geolingüísticas. La aplicación permite realizar de forma automatizada las labores de gabinete más importantes en la elaboración de atlas lingüísticos. Se trata de un programa en Visual Basic para Windows '95 o versiones posteriores que requiere, al menos, un procesador Pentium con 16 M de RAM y 10 M de disco duro.

1. INTRODUCCION

Desde 1996, la Universidad de Alcalá ha venido haciendo un notable esfuerzo para conseguir una aplicación informática capaz de cartografiar los materiales registrados previamente en una o más bases de datos. Este esfuerzo ha permitido incorporar nuestro trabajo a una línea de investigación muy cultivada en otros centros de investigación, tanto de Europa como de América. Pensamos, por ejemplo, en los sistemas informáticos desarrollados para el *Atlante lingüístico del ladino dolomitico/centrale e dialetti limitrofi*, de la Universidad de Salzburgo¹; en la informatización del *Atlante linguistico della Sicilia*, realizada

¹ Véase R. BAUER, "L'informatizzazione dell'Atlante linguistico sonoro ALDI", *Lingüística*, xxxii (1992), pp. 197-212.

en el Centro de Estudios Filológicos y Lingüísticos Sicilianos (Palermo)², o en los trabajos del *Linguistic Atlas of the Middle and South Atlantic States*, de la Universidad de Georgia (EE.UU.)³.

En el dominio hispanohablante, la primera obra que se ha adentrado en el terreno de la automatización ha sido el *Atlas Lingüístico y Etnográfico de Cantabria*⁴, dirigido por Manuel Alvar, aunque también hay noticias de la informatización de los materiales del *Atlas Lingüístico de Hispanoamérica*. La aplicación desarrollada en la Universidad de Alcalá se denomina *Atlante* y la intención de estas páginas es describir someramente su utilidad y características⁵.

2. OBJETIVOS DE ATLANTE

El principal objetivo del programa *Atlante* es la automatización de las labores de gabinete que conducen a la confección de un atlas lingüístico, así como el máximo aprovechamiento de toda la información lingüística que contiene una obra de estas características. Para el cumplimiento de tal objetivo se precisan unas bases de datos en las que queden registrados todos los materiales y la elaboración de diversas clases de índices que permitan una rápida búsqueda de cualquier información requerida por el investigador. Además, se necesita un mecanismo que permita trasladar a un mapa cualquiera los datos lingüísticos contenidos en las bases de datos, ya sea en forma de transcripción fonética, ortografía ordinaria o mediante símbolos.

Pero la finalidad de la aplicación *Atlante* no es solamente –y no es poco– llegar a una redacción informatizada de atlas lingüísticos que van a ser presentados en papel, sino ofrecer la posibilidad de trabajar y manipular la información lingüística contenida en ellos directamente en medios informáticos. En otras palabras, se ha buscado un medio que automatice la elaboración convencional de los atlas lingüísticos y que permita presentar un atlas o bien encuadernado o bien en soportes informáticos como el CD-ROM o la consulta *on line*.

² Véase A. PENNISI, "Piano di informatizzazione dell' Atlante linguistico della Sicilia (ALS). Fondamenti, strategie, progetto", en G. RUFFINO (ed.), *Materiali e ricerche dell'Atlante linguistico della Sicilia*, vol. V, Palermo, Centro di Studi Filologici e Linguistici Siciliani, 1990, pp. 3-106.

³ Véanse referencias en C.M. ZIAMANDANIS, "Dialectología y ordenadores", en M. ALVAR (dir.), *Manual de dialectología hispánica. El español de España*, Barcelona, Ariel, 1996, pp. 55-62.

⁴ 2 vols., Madrid, Arco/Libros, 1995.

⁵ Véase C.M. ZIAMANDANIS, art. cit., pp. 59-60.

A la hora de determinar el atlas-tipo al que deseamos llegar mediante nuestra aplicación, hemos usado como referencia uno de los modelos más ricos y complejos de los que conocemos: el modelo de los atlas lingüísticos españoles dirigidos por Manuel Alvar⁶. Se trata de atlas, como es bien sabido, que incluyen mapas lingüísticos, mapas etnográficos y mapas lingüístico-etnográficos, en los que la información lingüística se ofrece, unas veces, mediante símbolos –como es frecuente en otras obras informatizadas–, pero la mayoría de ellas en transcripción fonética con respuestas de extensión variable en cada punto, con los datos distribuidos dentro de los mapas, fuera de los mapas (en los márgenes) y en láminas complementarias, en los que se alternan tipos de letras y caracteres de diversa naturaleza. La transcripción fonética se presenta, además, siguiendo el alfabeto de la *Revista de Filología Española*.

Las características de las obras que nos han servido de modelo exigían la preparación de una aplicación sumamente flexible, capaz de registrar materiales en transcripción fonética y en ortografía ordinaria; de alternar alfabetos diferentes en el mismo texto y de redactar automáticamente mapas muy densos, pero dando al investigador, al mismo tiempo, la oportunidad de ajustar la presentación definitiva de los mapas a sus gustos y necesidades. Todo ello complementado con el desarrollo del alfabeto fonético mencionado, pero enriquecido con una larga lista de caracteres fonéticos imprescindibles para la transcripción de numerosos rasgos de las hablas hispánicas.

Las bases de datos deben ser capaces de permitir usos y manipulaciones muy diversas, desde la simple consulta de cualquier cadena de caracteres, hasta la elaboración de índices –incluidos índices de formas inversas–, la incorporación de archivos sonoros⁷ y gráficos, o el cálculo de distancias lingüísticas entre puntos (dialectometría).

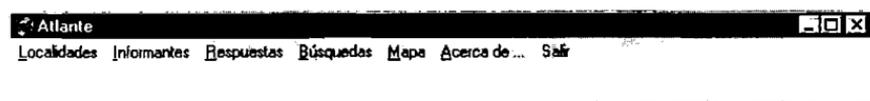
3. CARACTERÍSTICAS DE *ATLANTE*

Atlante es una aplicación preparada para ejecutarse bajo sistemas operativos de 32 bits –en particular para Windows'95 o versiones posteriores– utilizando Visual Basic v.4.0.; la formación y el manejo de la base de datos de *Atlante* se realiza mediante “*Microsoft Access Jet Engine*”. Para la utilización de este programa es aconsejable disponer de un procesador Pentium 100 o superior,

⁶ Véase M. ALVAR, *Estudios de geografía lingüística*, Madrid, Paraninfo, 1991.

⁷ Véase R. BAUER, “Dialettologia computazionale di atlanti linguistici: la dimensione sonora”, en M.T. Romanello e I. Tempesta, *Atti del xxvii Congresso della Società di linguistica italiana*, Roma, Bulzoni, 1995, pp. 155-167.

con 16 M de RAM como mínimo (aunque lo recomendable es contar con 32 M) y suficiente espacio en el disco duro para almacenar tanto la base de datos como los ficheros BMP de 256 colores que se vayan desarrollando al elaborar los mapas.



Atlante organiza su funcionamiento en la utilización de tres bases de datos específicas y su proyección gráfica en un mapa. Las bases de datos que se crean y manejan son las siguientes:

- *Base de localidades.* Se accede a ella desde el menú principal, una vez abierto el programa. Allí se deben introducir –y es donde van a figurar– el nombre de la localidad, el código con el que la identificaremos, los nombres de los informantes y las coordenadas relativas que cada localidad ha de tener en el mapa. En esta base es posible añadir otros datos adicionales, según las características de cada atlas (por ejemplo, el nombre dialectal de los naturales de cada localidad).
- *Base de informantes.* También se accede a ella desde el menú principal, una vez abierto el programa. En esta base figuran todos los datos que se consideran pertinentes relacionados con los informantes: edad, profesión, nivel de instrucción, naturaleza de los padres, lecturas, viajes, tiempo dedicado a ver y oír la televisión o la radio, etc.
- *Base de respuestas o de información lingüística.* Como en los casos anteriores, accedemos a ella desde el menú principal, una vez abierto el programa. Los datos lingüísticos que aquí se registran pueden aparecer en ortografía ordinaria (en cualquiera de los alfabetos de que disponga el ordenador o computadora) o en transcripción fonética. Para transcribir los datos fonéticos en los campos de la base destinados a ello, se puede utilizar una fuente “TTF” para windows, previamente instalada en el directorio “Windows/Fonts”. La Universidad de Alcalá ha preparado una fuente de estas características (afe.ttf) para la representación de signos fonéticos ajustados a las pautas establecidas en la *Revista de Filología Española*, aunque también es posible utilizar cualquiera de las versiones disponibles en el mercado para el Alfabeto Fonético Internacional.

Afe.ttf*muestra de algunos caracteres Afe*

žšžšɾɿŷłśśśž!"šlŷřđŷɿɿcɸezđɸgɦŋŋłŋžɦŋbŋmɿø
 ęɛɿɸəuáéãẽĩæeiɿŷ

De estas tres bases de datos, sin duda la más compleja, desde un punto de vista técnico, y la más rica, desde una perspectiva lingüística, es la de *respuestas*. Técnicamente, no se trata de una sola base sino de un conjunto de bases, que será más o menos amplio, según la cantidad y la variedad de los contenidos que en ella se quieran incluir: no es lo mismo un atlas fonético, que un atlas general; no es igual un atlas de 300 mapas que uno de 3.000, evidentemente. Al hablar de una base de respuestas tan sólo pretendemos facilitar la comprensión de la estructura del programa.

Las bases de datos que componen lo que hemos llamado la *base de respuestas* podrían ser tantas como niveles lingüísticos se quieran recoger en el atlas. Así, si un atlas sólo incluye información fonético-fonológica y léxica, serían dos las bases necesarias; si se trabaja, además, con un nivel morfológico o sintáctico —o de cualquier otro tipo—, se contaría con otras tantas bases de datos. Por otra parte, cada una de las bases, según el número de unidades que la compongan, puede estar subdividida en “tablas” a las que denominamos *campos*; de este modo, la base de “Fonética” puede subdividirse en un campo de “Fonética vocálica” y otro de “Fonética consonántica”, por ejemplo, y la base “Léxico” puede quedar dividida en tablas o campos como “Partes del cuerpo”, “Vida agrícola”, “Fenómenos atmosféricos”, etc. Dentro de cada campo figuran las respuestas recogidas a propósito de determinados objetos, acciones o conceptos. Cuanto mayor sea el número de unidades que componen cada base, más necesaria se hace la subdivisión en tablas o campos, para garantizar así un manejo ágil y rápido de la base.

La introducción de los datos en la *base de respuestas* se hace a través de una ventana dispuesta de forma similar a un formulario; algunos datos han de registrarse en ortografía ordinaria y otros en transcripción fonética. Para introducir la información correspondiente a cada campo, es preciso seleccionar en el formulario una serie de opciones referidas al nivel lingüístico, la localidad, el informante y el nombre del propio campo. Las posibilidades de selección de cada uno de estos aspectos (nivel lingüístico y nombre del campo que interesan, código de la localidad, nombre del informante) se despliegan automáticamente en el formulario, eliminando el riesgo de error. Esto es posible gracias a que en el formulario contamos con una lista actualizada de todas las localidades y todos los informantes incluidos anteriormente en

Edición de las respuestas

Niveles: Léxico (Hombre) Campos: Animales del campo
 Localidades: AB 103
 Informantes: Eugenio Blasco Zolve

Aparecidos hasta el momento:

- laqertíxe
- laqertíhäh
- relentíxe

Pregunta: lagartija
 Ortografía: lagartija Superíndice
 Fonética: laqertíxe Subíndice
 Tipo

nuestras bases. Así, pues, la ventana de la *base de respuestas* muestra, en principio, información sobre la localidad, el informante, el nivel lingüístico y el campo a los que corresponden los datos lingüísticos que allí se van a introducir y que posteriormente se van a consultar, si se desea.

La información lingüística que se introduce en esta ventana se registra tanto en ortografía ordinaria como en transcripción fonética; se procede así, entre otras razones, para facilitar las búsquedas de secuencias de caracteres. Si se maneja información referida a las conjugaciones verbales, también es posible contar con una lista de los verbos cuyas formas se van a introducir o consultar.

Para introducir la información lingüística, esto es, las respuestas que los informantes han dado a un cuestionario, disponemos de dos tipos de ayuda: por un lado, el formulario ofrece un listado con las respuestas registradas en cada campo, evitando introducir una y otra vez los mismos datos, con el consiguiente ahorro de tiempo y la reducción de errores innecesarios; por otro lado, pulsando las teclas F1 y F2 obtenemos una ayuda gráfica del alfabeto fonético que estamos utilizando y su distribución en el teclado.

En resumen, la ventana en la que se introducen y consultan las respuestas recogidas en las encuestas geolingüísticas ofrece, en todo momento, información actualizada sobre el lugar, informante, nivel lingüístico y campo a los que

corresponden unas formas lingüísticas que se presentan en transcripción fonética y ortografía ordinaria.

4. LA CONSULTA DE LAS BASES DE DATOS

El menú principal del programa dispone de una opción llamada *Búsquedas*, mediante la cual se accede a un sencillo menú. Las búsquedas se realizan automáticamente entre las respuestas registradas en ortografía ordinaria: en el lugar adecuado para ello, se ha de anotar la secuencia de caracteres (una palabra o parte de una palabra) que se quiere encontrar en todos los campos de todas las bases. Efectuado el rastreo, el programa proporciona información precisa sobre el campo, el mapa, la localidad y el informante en los que se ha encontrado cada una de las apariciones de la secuencia buscada. La información se ofrece, para cada uno de los casos encontrados, de la forma siguiente:

Búsqueda: *sma*

Campo: Morfología

Palabra: fantasma

Nº de Mapa: 100

Nº de Lámina: 100

Palabra: fantasma → fantasma, el

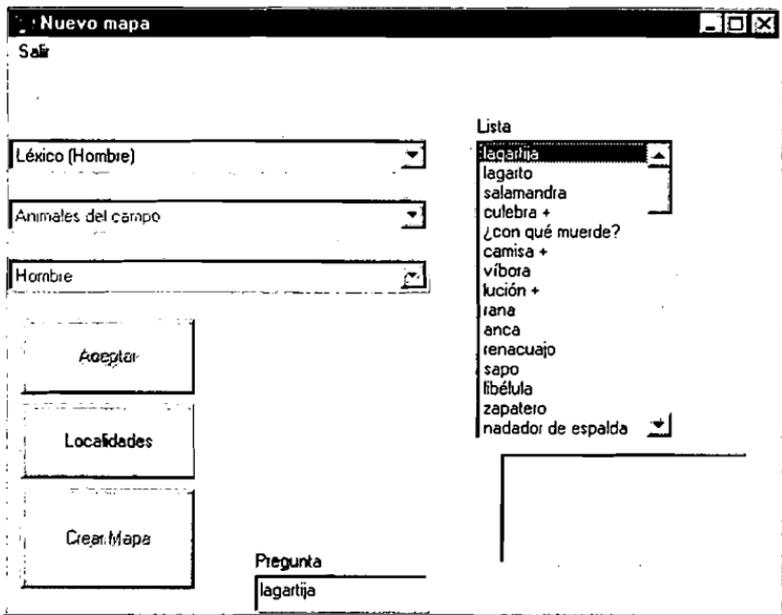
Localidad: M 103

Informante: Pedro Pérez Pérez

Esta información puede ser impresa en papel, para un más cómodo manejo por parte del investigador.

5. CARTOGRAFIADO AUTOMATICO

Las bases de datos lingüísticos permiten que la información en ellas contenida se pueda proyectar de forma horizontal, en mapas y de un modo automático. Para ello basta seleccionar la opción *Mapa* del menú principal.



Cuando se elige la opción de crear un mapa, para lo que previamente debemos haber completado las bases de datos lingüísticas, aparece en pantalla un mapa mudo del territorio de nuestro atlas y varias opciones en un nuevo menú. Las opciones que se presentan son las que se exponen a continuación:

- *Mapa de símbolos*: permite crear un mapa con símbolos asociados a una leyenda, a partir de la información almacenada en la base de datos. Para elaborar estos mapas podemos combinar hasta diez símbolos diferentes, a los cuales se puede asociar cualquier tipo de leyenda escrita con cualquier tipo y tamaño de letra.
- *Texto*: permite insertar cualquier texto en un margen del mapa, con el fin de aclarar cualquier problema sobre el contenido del mapa o para aportar cualquier tipo de información adicional.
- *Archivo*: permite guardar el mapa que se tiene en pantalla; también permite abrir un mapa grabado previamente. Con la opción *Nuevo* se crea un mapa nuevo con las características que se desee, entre las posibilidades que se ofrecen en un formulario. En ese formulario se elige la información lingüística (respuesta) que se quiere representar de entre todas las que se

recogen en las bases de datos, así como las localidades que se quiere representar o no dentro del mapa o en un margen.

6. CONCLUSION

La aplicación *Atlante*, creada en la Universidad de Alcalá, pretende ser un instrumento ágil, flexible y potente para el cartografiado automático y la gestión de bases de datos lingüísticos. Su versatilidad para buscar, cartografiar o consultar información lingüística, lo hace un instrumento capaz de automatizar prácticamente cualquier tipo de atlas lingüístico, desde los más simples a los más ricos y complejos.