



Modelamiento de una base de datos espacial para el Atlas Lingüístico-Etnográfico de Colombia

Spatial database modeling for the Atlas Lingüístico-Etnográfico de Colombia

Johnatan E. Bonilla

INSTITUTO CARO Y CUERVO
COLOMBIA
johnatan.bonilla@caroycuervo.gov.co

Julio Alexander Bernal Chávez

INSTITUTO CARO Y CUERVO
COLOMBIA
julio.bernal@caroycuervo.gov.co

Recibido: 28-VIII-2017 / **Aceptado:** 22-V-2019

DOI: 10.4067/S0718-09342020000200346

Resumen

El proyecto ALEC Interactivo tuvo como objetivo el modelamiento de una base de datos espacial para el diseño y desarrollo de un Sistema de Información Geográfica y un Atlas Web con el fin de salvaguardar el patrimonio lingüístico del país y brindar acceso y divulgación a los materiales del Atlas Lingüístico-Etnográfico de Colombia (ALEC). En tal sentido, el artículo presenta una propuesta metodológica para el diseño y desarrollo de una base de datos espacial que permite el procesamiento informático, digitalización y sistematización de los datos léxicos, etnográficos y los materiales suplementarios del ALEC.

Palabras Clave: Atlas Lingüístico-Etnográfico de Colombia, Sistemas de Información Geográfica, Base de datos espacial, geolingüística.

Abstract

This article presents a methodological proposal for the design and development of a database that allows computer processing, digitization and systematization of lexical, ethnographic and supplementary materials of the Atlas Lingüístico-Etnográfico de Colombia. Our project aims to safeguard the linguistic heritage of Colombia and provide access and dissemination to the ALEC materials, through the modeling of a spatial database for the design and development of a Geographic Information System and a Web Atlas.

Key Words: Atlas Lingüístico-Etnográfico de Colombia, Geographic Information Systems, Spatial Database, geolinguistics.

INTRODUCCIÓN

El Atlas Lingüístico-Etnográfico de Colombia (ALEC) (Instituto Caro y Cuervo, 1982-1983) es una de las fuentes esenciales para el estudio del español del país en diversos campos como la dialectología, la lexicología y la fonética. Además, es considerado muestra representativa del patrimonio cultural y lingüístico de Colombia. El tamaño de la obra impresa, compuesta por seis tomos, cada uno de 50 x 35 cm, ha hecho que la circulación del material se vea reducida a los anaqueles de bibliotecas y a la consulta y análisis por parte del público especializado. Con el interés de conservar y brindar acceso y divulgación a la obra, desde el año 2015, la línea de Investigación en Lingüística de Corpus y Computacional (LICC) del Instituto Caro y Cuervo (ICC) en el marco del proyecto ALEC Interactivo tiene como objetivo el desarrollo de la versión digital de los materiales del ALEC con el apoyo del Núcleo de Investigación en Datos Espaciales (NIDE) de la Universidad Distrital “Francisco José de Caldas” (UDFJC).

Si bien la metodología de recolección de datos y la presentación en formato de atlas corresponde con las propuestas de la geolingüística tradicional, el ALEC se constituye en un Corpus en la medida en que está compuesto por un conjunto estructurado y documentado de materiales elicitados en función de criterios lingüísticos explícitos (Sinclair, 1996). Teniendo en cuenta que uno de los objetivos de la lingüística de corpus es brindar herramientas para la recolección, sistematización, análisis y explotación de muestras de lengua y que el desarrollo de los medios tecnológicos ha permitido, no solo almacenar y codificar datos, sino ponerlos a disposición de un público más vasto, el proyecto ALEC Interactivo comprendió, en su primera fase, la formulación y aplicación de una propuesta metodológica para la digitalización de la información contenida en todos los materiales del ALEC (mapas, suplementos, fotografías, ilustraciones y audios) y su estructuración y modelamiento en una base de datos espacial. Lo anterior, con el interés principal de servir de base para consolidar un Sistema de Información Geográfica y desarrollar una versión digital-interactiva denominada ALEC Digital.

En dicho sentido, el objetivo de este artículo es presentar la metodología utilizada para el procesamiento informático, digitalización y sistematización de los datos léxicos, etnográficos y materiales suplementarios del ALEC en una base de datos espacial. Cabe aclarar que el presente artículo no pretende ofrecer aportes teóricos, sino que es de carácter únicamente metodológico. En la primera parte se presenta un acercamiento descriptivo a los materiales que componen el ALEC, generalidades conceptuales sobre el manejo de datos espaciales, de metadatos y el desarrollo de bases de datos espaciales con posibilidades de aplicación específica para grupos de datos lingüísticos y etnográficos de dimensión diatópica. La segunda parte expone la propuesta metodológica que incluye los criterios tenidos en cuenta para el tránsito de

la información del soporte análogo a digital, la descripción de los metadatos para la descripción y organización de la información y el modelamiento de la base de datos espacial y, por último, se presentan las posibilidades de uso y explotación de la base de datos espacial, en el estado actual, de la versión web del ALEC.

1. Marco teórico

1.1. *Materiales del Atlas Lingüístico-Etnográfico de Colombia*

La investigación del Atlas Lingüístico-Etnográfico de Colombia (ALEC) inició en 1954 con el objetivo de investigar los dialectos y subdialectos del español hablado en Colombia (Flórez, 1961). Durante veintidós años (1956-1978) se llevó a cabo la recolección de información lingüística y etnográfica, participaron 23 investigadores del Departamento de Dialectología del Instituto Caro y Cuervo y 2.234 informantes de 264 localidades distribuidas en el territorio colombiano. Como resultado de la investigación fueron publicados, entre 1982 y 1983, seis tomos que contienen 1.696 láminas con 1.523 mapas lingüísticos, etnográficos y mixtos, muestras de habla espontánea, información adicional de tipo etnográfico sobre las temáticas tratadas, ilustraciones y fotografías. Según Flórez (1983), en los mapas lingüísticos¹ se registran los nombres que los informantes dieron a los conceptos por los cuales se preguntó en las encuestas y los mapas etnográficos muestran las áreas o la difusión geográfica de ‘cosas’ u objetos de la vida material popular.

La metodología de recolección de información del ALEC fue la encuesta directa con base en un cuestionario que inicialmente tenía “8.058 puntos, repartidos en 7.300 de léxico, 430 de fonética y 328 de morfosintaxis” (Flórez, 1983: 29). El cuestionario fue reducido a 2.000 preguntas tras ser sometido a pruebas en Pacho, Cundinamarca en enero de 1956 y en 17 localidades del departamento de Bolívar. Después de la evaluación del ejercicio de prueba y de revisar la pertinencia de varias de las preguntas referentes a cosas de la vida material inexistentes en las localidades, en 1961 se imprimió un cuestionario reducido con 1.348 preguntas que sería aplicado en el resto de las localidades hasta la conclusión del trabajo de campo.

La información léxica y etnográfica recolectada, desde su planeación en los cuestionarios y en la representación cartográfica en los seis tomos finales (1982-1983), está dividida en dieciséis campos semánticos relacionados con la cotidianidad de los hablantes encuestados y la vida en el campo, a saber: el cuerpo humano, el vestido, la vivienda, la alimentación, la familia y el ciclo de vida, las instituciones y vida religiosa, las festividades y distracciones, el tiempo y el espacio, el campo --los cultivos y otros vegetales, industrias relacionadas con la agricultura, ganadería, animales domésticos, animales silvestres, oficios y empleos, transporte y embarcaciones y pesca. Sumado a esto, encontramos preguntas concernientes a aspectos fonéticos y gramaticales, específicamente, onomatopeyas en el Tomo II y variación fonética de vocales y

consonantes en diversos contextos de palabra y fenómenos gramaticales en el Tomo VI.

Los seis tomos del ALEC tienen información adicional que, para efectos del tránsito de la versión análoga a la digital, puede considerarse multimedia, a saber: muestras escritas de habla espontánea: partituras: ilustraciones; fotografías; y en el suplemento del Tomo III, dos discos en acetato, que contienen audios de juegos y cantos de velorio de las costas caribe y pacífica de Colombia. Tanto los contenidos gráficos de los tomos (imágenes, ilustraciones y partituras), como los audios del suplemento del Tomo III, se encuentran digitalizados. Todos los tomos del ALEC fueron escaneados a gran resolución en la Biblioteca Nacional de Colombia y entregados al ICC en versión digital, en imágenes de alta resolución en formato TIFF por cada lámina; el suplemento también fue escaneado y los discos de vinilo digitalizados en formatos WAV y MP3.

La investigación cuenta también con 1.112 archivos de audio, que hacen parte de los Corpus Lingüísticos del ICC (CLICC), además de 16.803 fotografías. Los audios del ALEC fueron recogidos en medio magnético durante la investigación y digitalizados por la Fundación Patrimonio Fílmico Colombiano en el año 2014. Los audios contienen relatos, canciones, encuestas fonéticas y entrevistas, cada uno dura de 30 a 60 minutos. Respecto a las fotografías, en el año 2012 fue digitalizado el fondo de negativos fotográficos del ALEC. Durante el proceso se concluyó que todas las localidades visitadas fueron fotografiadas y corresponden a un guión único, que se adapta y desarrolla según las circunstancias locales, en el que todos los objetos correspondientes a las preguntas y todos los informantes fueron fotografiados. Por último, el ALEC tiene un Manual (Flórez, 1983) que guarda la información referente a aspectos contextuales y metodológicos de la investigación; la organización y lectura de mapas; y datos descriptivos acerca de los encuestadores, localidades e informantes.

1.2. Datos espaciales y datos no espaciales

El ALEC fue desarrollado bajo los principios básicos de la geografía lingüística y dialectología de mediados del siglo XX. La geografía lingüística, según Coseriu (1956: 29) es un “método dialectológico y comparativo” que se encarga de registrar en mapas las formas lingüísticas encontradas a través de las encuestas directas, en una red de puntos espaciales de un territorio determinado. Entre tanto, la dialectología da razón de las variedades de una lengua sobre el espacio geográfico, partiendo del análisis lingüístico en alguno de los niveles de la lengua (fonético, morfológico o léxico), apoyado desde la geolingüística y la sociología (Montes Giraldo, 1995). Sin embargo, para la época de concepción del ALEC la geografía lingüística como método, y la dialectología como disciplina de análisis, se centraban en la dimensión diatópica, puesto que trabajaban a partir de la ubicación de datos lingüísticos en espacios geográficos delimitados.

Cualquier grupo de datos que esté directa o indirectamente relacionado con una ubicación en la superficie de la tierra es considerado un dato espacial (Lamigueiro, 2014). La referencia de los datos espaciales puede estar dada por coordenadas (latitud y longitud) o por sistemas de referencia (convenciones). Surve, Bhaskar y Nehru. (2014) exponen que los conjuntos de datos espaciales normalmente tienen como dato primario un punto, línea o polígono que hace referencia a algún sistema de redes específico. Por el contrario, los datos no espaciales, que también son conocidos como atributos o características, representan información independiente de toda consideración geométrica porque no guardan relación con alguna ubicación concreta en el espacio.

Toda la información del ALEC es multidimensional y correlacionada, ya que puede ser definida como un conjunto de datos espaciales indirectos, ligados a un dato espacial primario, llamado localidad. Cada localidad representa un punto dentro del polígono diseñado para el territorio encuestado y cuenta con un sistema convencional de referencias. Los datos no espaciales indirectos son, en principio, cada una de las variantes lingüísticas o etnográficas encontradas en las localidades. Las variantes más comunes fueron cartografiadas, es decir, puestas sobre el mapa o polígono mediante convenciones. Las variantes menos comunes fueron georreferenciadas, a partir de un sistema de notación de letras y números que se refieren a las localidades. De modo idéntico ocurre con el resto de la información que contiene los seis tomos del ALEC, su suplemento y los fondos no publicados (corpus oral y fotográfico). Las adiciones de texto, fotografías, ilustraciones y dibujos, grabaciones, partituras, muestras de habla espontánea etc. fueron vinculadas al dato espacial primario, mediante el mismo sistema de georreferenciación de las variantes no cartografiadas.

En el Manual del ALEC (Flórez, 1983) encontramos datos no espaciales, es decir, atributos o características, tales como: las variables sociales de los informantes y los cuestionarios contestados, el tiempo de vinculación de los encuestadores, las fechas exactas de trabajo de campo, etc., pero cada uno de estos datos también está vinculado con el dato espacial primario. Por ejemplo, cada informante pertenece a un punto específico en el mapa lingüístico, por lo tanto, los atributos o características sociales que se puedan describir permiten la representación en el mapa y una lectura multidimensional de la variación lingüístico-etnográfica. Las variables sociales de género, edad, ocupación o proveniencia pueden ser analizadas mediante la superposición de capas de datos y revelar correspondencia o distancia con alguna variante lingüística en el plano espacial.

1.3. Metadatos y metadatos espaciales

Los metadatos son un tipo de datos cuyo propósito primario es describir, definir y/o etiquetar otros datos (Nadkarni, 2011). A partir de la definición de metadatos es posible que datos espaciales y no espaciales puedan ser identificados y consultados

mediante diversos criterios de búsqueda y selección. De manera similar al registro del catálogo de una biblioteca, los metadatos documentan quién, qué, cuándo, dónde, cómo y porqué de un recurso de datos. Los metadatos espaciales se diferencian de los no espaciales en la medida que describen mapas, archivos, imágenes y cualquier otro recurso de datos basado en la localización.

Si bien los metadatos del ALEC serán descritos de forma detallada en el apartado 3.2, a modo de ejemplo podemos observar que en la obra impresa del ALEC hay diversos datos susceptibles de descripción y de asignación de metadatos. En cada tomo la información distribuida por campos semánticos a los que se asignó un número específico de láminas y mapas. Mediante la asignación de los metadatos: Tomo, Campo Semántico, Lámina y Mapa, la jerarquización permite realizar diversas consultas: un tomo mediante el identificador numérico, establecer un grupo limitado de láminas y mapas dentro de un tomo mediante la selección de un campo semántico o limitar directamente la búsqueda a un número de lámina o título de mapa.

Los metadatos del Manual del ALEC (Flórez, 1983) son específicos para descripción de la información sobre los encuestadores, hablantes y localidades y todos son susceptibles de georreferenciación al guardar relación con los lugares de la encuesta. Por lo tanto, la información del manual permite la delimitación de relaciones entre características sociales como el nivel escolar, oficio o procedencia y geográficas como altura sobre el nivel del mar y vías de acceso.

Para el caso de los fondos no publicados, Corpus Oral del ALEC y fondo fotográfico, contamos con la localidad como metadato principal. De la misma manera, cada audio y fotografía ha de ser descrito en términos de su contenido y condiciones de recolección para que su ubicación sea posible mediante más metadatos por parte de los usuarios.

1.4. Bases de datos espaciales y Sistemas de Información Geográfica

Durante la época en la que fue concebido el ALEC, los mapas en papel eran el medio principal para sintetizar y representar información geográfica. Ahora, las nuevas tecnologías en torno a la recolección y digitalización de datos espaciales y la creciente necesidad de manipulación interactiva y análisis de estos datos ha generado la necesidad de implementar una tipología de software especializada conocida como Sistemas de Información Geográfica (SIG). Rigaux, Scholl y Voisard (2001) explican que además de ser una herramienta cartográfica para producir mapas, los SIG almacenan datos espaciales, recuperan y combinan estos datos para crear nuevas representaciones del espacio geográfico y proveen herramientas para el análisis y la simulación.

Los SIG permiten la administración eficiente de bases de datos con información compleja y diversa, como la del ALEC, mediante el modelamiento y estructuración de bases de datos espaciales (BDE). En inicio, se requiere del diseño de Sistemas Gestores de Bases de Datos (SGBD), que construyen sistemas encargados de definir la estructura de almacenamiento de los archivos y los mecanismos de manejo de la información. Sin embargo, un SGBD común sólo puede manipular conjuntos de datos simples, mientras que una Base de Datos Espaciales (BDE) tiene un componente cartográfico que da la posibilidad de administrar los datos geográficos del ALEC. De allí surge la necesidad de extender y adaptar la tecnología de los SGBD a las particularidades de las BDE.

La diferencia entre las BD convencionales y las BDE está dada por la naturaleza de la información que se contiene. Las BDE almacenan entidades del mundo real que tienen componentes espaciales y no espaciales de forma estructurada y sin duplicidades (Bosque, Fernández, Martín-Forero & Pérez, 2012). Los Sistemas de Gestión de Base de Datos Espaciales (SGBDE) ofrecen en su modelamiento y sistema de consulta compatibilidad con diferentes tipos de datos espaciales (puntos, líneas, polígonos), proporcionando la abstracción fundamental para el modelado de la estructura de las entidades geométricas en el espacio, así como sus relaciones, propiedades y operaciones (Güting, 1994). Teniendo en cuenta lo dicho, para la digitalización de los datos del ALEC fue esencial un SGBDE, ya que ofrece las capacidades de manejo de datos espaciales y cuenta con la tecnología de base suficiente para consolidarse como SIG; soporta los datos espaciales y no espaciales en su implementación; y provee de indexación y algoritmos eficientes para el relacionamiento y análisis de la información.

A mediados de la década de los setenta inicia la inclusión de tecnología computacional para el almacenamiento de datos de investigación lingüística y la producción de atlas lingüísticos (Hoch & Hayes, 2010). La primera publicación realizada sobre el uso de BDE y modelamiento computacional fue realizada por Alan Richard Thomas en 1980 en el trabajo *Areal Analysis of Dialect Data by Computer: A Welsh Example*. En este texto, Thomas presentó un temprano ejemplo de las posibilidades de uso de un SIG para medir la correlación espacial en datos computarizados de su investigación lingüística. Thomas logró el almacenamiento y visualización de múltiples atributos lingüísticos mediante la asignación de valores numéricos al uso de palabras del galés en diferentes regiones sobre un mapa, pero con las limitaciones propias de la época, como la difícil asignación de caracteres especiales o la imposibilidad de administrar grandes cantidades de texto (Hoch & Hayes, 2010). Años después, inspirado en los trabajos de Thomas, Pederson diseñó una BDE para el almacenamiento computarizado y la posterior visualización del *Linguistic Atlas of the Gulf States*. El trabajo de Pederson (1996) supuso un adelanto en el uso de herramientas SIG ya que muchos de sus desarrollos solo estarían disponibles en una

interfaz gráfica hasta casi una década después. En cuanto al contexto hispanohablante, el Atlas Lingüístico y Etnográfico de la Provincia de Santander (Alvar, 1977) fue el primero que utilizó un diseño de BDE. La base de datos desarrollada como un producto del proyecto facilitó la creación de mapas por demanda y la preparación de índices utilizados en la interpretación de atlas lingüísticos (Alvar, 1991).

Desde estos primeros intentos, hasta la fecha se han publicado diversos proyectos de atlas lingüísticos con base en tecnologías de BDE y SIG establecidos en dos vías: a) atlas lingüísticos diseñados desde su inicio mediante el modelamiento de BDE y el uso de herramientas SIG y b) atlas lingüísticos que en principio fueron análogos y que mediante el procesamiento informático hoy cuentan con una versión digital. En la primera vía encontramos trabajos como el *Linguistic Atlas of the Middle and South Atlantic States* (Schneider & Kretzschmar, 1989), consolidado recientemente en línea como parte del macro proyecto *Linguistic Atlas Projects* (www.lap.uga.edu); *Ethnologue* (www.ethnologue.com); el Atlas Lingüístico y etnográfico de Castilla La Mancha (García Mouton & Moreno Fernández, 2003) y *The World Atlas of Language Structures Online*. Mientras que como ejemplos de procesamiento informático de atlas análogos para su posterior presentación digital se encuentran el *Sprach und Sachatlas Italiens und der Südschweiz* (Jaberg, Jud & Scheuermeier, 1928) y el proyecto 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.

De acuerdo con la experiencia de procesamiento informático del Atlas Lingüístico de la Península Ibérica (ALPI), los SIG se están aplicando para la visualización de los datos lingüísticos en un entorno geográficamente referenciado, es decir, como nexo de conexión de la base de datos relacional con una base de datos espacial, posibilitando no sólo el cartografiado automático a través de diferentes tipos de representaciones, sino también la realización de búsquedas complejas en el corpus completo geolingüístico y etnográfico (García Mouton, 2013). Las tecnologías de información geográfica se usan también en el ALPI para generar una aplicación interactiva en Internet (alpi.csic.es), a través de un geoportal con servicios web basados en localización geográfica, con la guía de los estándares del *Open Geospatial Consortium* (OGC) y combinados con búsquedas sobre la base de datos temática (Heap, 2003).

Si bien existen adelantos en bases de datos y desarrollos geoespaciales para la administración y análisis de datos lingüísticos en la bibliografía encontrada y en la web no se encontraron modelos abiertos para replicar o implementar en los trabajos del ALEC. Cada uno de los atlas lingüísticos desde su elaboración en papel o digital plantea características específicas que obligan a diseñar modelos ajustados con los tipos de información y el estado de las investigaciones. Por estas razones fue necesario

modelar desde cero, con el apoyo de un grupo de ingenieros expertos en geodesía, una BDE como se expone en el siguiente capítulo.

2. Metodología para el modelamiento de la BDE del ALEC

Para el diseño de la BDE del ALEC se utilizó el modelo de entidad-relación. En dicho modelo se adopta una representación más natural del mundo real que consiste de entidades, los atributos de éstas y las relaciones entre esas entidades. Cada objeto puede ser comprendido como una entidad que se identifica con precisión (por ejemplo, una persona específica o un lugar) y la relación es la asociación que existe entre ellos (persona -lugar de nacimiento) (Fuentes, 2012). Estas entidades están compuestas por varios atributos, que vienen a ser sus propiedades. Por ejemplo: la entidad alumnos, tendrá los atributos nombre, #pasaporte, nacionalidad, fecha de nacimiento, etc. Los atributos también reciben el nombre de columnas en la terminología de BD. Entre los atributos, habrá uno o un conjunto de ellos que no asegura la unicidad de una fila, a dicho atributo, o conjunto de atributos, se le llama clave de la entidad. En el caso de los alumnos, sería el #pasaporte. En toda entidad siempre hay al menos una clave que, en el peor de los casos, estará formada por todos los atributos de la tabla. Ya que puede haber varias claves y necesitamos elegir una, lo haremos atendiendo a estas normas: a) Que sea única. b) Que se tenga pleno conocimiento de ella. c) Que sea pequeña, ya que será muy utilizada por el SGBD (Fuentes, 2012).

Este modelo fue seleccionado porque permite representar de manera sencilla los elementos que componen el ALEC y sus relaciones; permite conexión con el componente geográfico; hace posibles las consultas, la manipulación de los datos y el análisis. Además, el modelo es compatible con el diseño e implementación del SGBDE PostgreSQL con su extensión geográfica PostGis, seleccionados para la administración de la BDE del ALEC.

Recientemente, en el desarrollo de BDE se han utilizado modelos no relacionales (NoSQL), para la gestión de datos no estructurados, de gran tamaño y escalabilidad. Los datos no estructurados, son datos en bruto y sin organización alguna, para su almacenamiento no requieren estructuras fijas, tales como tablas. El modelo NoSQL no fue aplicado al ALEC ya que los datos geoespaciales del atlas tienen una estructura fija, explícita desde la definición de sus metadatos, y era indispensable conservar las interrelaciones entre ellos. Si bien los modelos NoSQL ofrecen la gestión de un mayor volumen y variedad de datos con mayor agilidad, no necesariamente brindan garantías de fiabilidad de los datos (Ramírez & Herrera, 2013) y en ocasiones el mecanismo que ofrecen los gestores para la recuperación de información dificulta la realización de consultas espaciales. Para ello resultaría necesario indexar los datos espacialmente, pero generalmente el rendimiento que se logra es inferior al que puede alcanzarse con un sistema relacional.

El método general utilizado para el desarrollo de la BDE y SIG del ALEC fue el método del ciclo de vida propuesto por Guevara (1992). En general, el método consiste en un conjunto de actividades que los analistas, diseñadores y usuarios realizan para desarrollar e implementar un sistema de información y comprende seis pilares básicos en el siguiente orden: 1). Investigación preliminar, 2). Determinación de los requerimientos del sistema, 3). Modelamiento de la BD 4). Desarrollo del software, 5). Prueba de los sistemas, 6). Implantación y evaluación. El presente artículo da cuenta de los tres primeros pasos del método que se relacionan directamente con el desarrollo de la BDE. El tercer paso fue desarrollado en dos etapas, a saber: la definición de objetos y metadatos, y la definición de objetos y relaciones.

2.1. Investigación preliminar

El primer paso consistió en la lectura y revisión de todos los materiales del ALEC. Sumado a esto, se desarrollaron talleres prácticos sobre manejo de datos del ALEC para los investigadores y asistentes de investigación, con participación de los ingenieros encargados del desarrollo de la BDE y SIG. Los temas tratados fueron: conteo y clasificación de variantes lingüísticas, análisis de variantes dialectales y uso de datos del manual, entendidos como factores que pueden incidir en el comportamiento de las variantes (fecha de fundación, clima, vías de acceso a las localidades etc.).

2.2. Análisis de requerimientos

Una vez los investigadores e ingenieros estuvieron familiarizados con los materiales, realizaron jornadas de trabajo conjunto para establecer las necesidades del proyecto en cuanto a los requerimientos del sistema y establecer los criterios básicos para el tránsito de la información de medio análogo a digital.

Como resultado del análisis de requerimientos, se optó por el uso de software libre por las ventajas que éste ofrece en cuanto a costos, total independencia de un proveedor, facilidad de personalización, aseguramiento de la durabilidad de la información y posibilidades de migración.

Asimismo, se tuvo presente la necesidad de que la BDE sirva como base para otro tipo de desarrollos: diccionarios, geoportales, juegos, aplicativos web o materiales con fines pedagógicos. En general, se requirió la conservación de los contenidos del ALEC sin que estos perdieran el sentido lingüístico y etnográfico, pero con la posibilidad de generación de mapas personalizados, análisis y cruce de búsquedas con la información complementaria descrita en el punto 2.1.

2.3. Definición de entidades y metadatos

Las entidades que se identificaron en los materiales del ALEC y que sirven para el modelamiento de la BDE fueron descritas mediante la asignación de metadatos. En el Manual del ALEC (Flórez, 1983) se establecieron como entidades Localidades, Cuestionarios, Encuestadores e Informantes. En cuanto a los tomos, se establecieron las entidades susceptibles de relación: Tomo, Campo semántico, Láminas que incluye a los mapas, Convenciones, Símbolos, Variantes Cartografiadas, Variantes no Cartografiadas, Adiciones Cartografiadas, Adiciones no Cartografiadas, Suplemento Georreferenciado, Suplemento no Georreferenciado. A continuación, se describen uno a uno.

2.3.1. Manual del ALEC

Para la digitalización de la información del Manual del ALEC, se diseñaron formularios que cargan de forma directa la información a la BDE por cada una de las entidades que hacen parte de la base de datos distribuidas de la siguiente manera: Localidades, Campos Semánticos, Encuestadores e Informantes. En cada tabla, los metadatos se ubicaron en la fila superior y los datos correspondientes a cada una fueron llenados columna a columna, como podemos observar en el siguiente ejemplo de la Tabla Encuestadores.

a) Tabla Localidades

Tabla 1. Ejemplo de distribución de metadatos en Tabla Encuestadores en la BDE del ALEC.

Idencuestador	Nombres	Apellidos	Años vinculación
1	José Joaquín	Montes	1958-1978
2	Francisco	Suárez Pineda	1958-1972

En el Manual del ALEC encontramos un total de 262 localidades correspondientes a 28 de los 32 departamentos de acuerdo con la división administrativa de Colombia a principios de 1966. Los investigadores del ALEC asignaron a cada localidad un código compuesto por los dos primeros caracteres del nombre del departamento del que hacen parte y un número dado en relación con su ubicación sobre el territorio. El código de la localidad se convierte en el dato espacial primario sobre el cual se georreferencia toda la información del ALEC.

La tabla diseñada para el ingreso de la información tuvo en cuenta la misma codificación. El primer metadato de la tabla, que también es la clave principal, es llamado idLocalidad y corresponde con el mismo código establecido en el ALEC. La clave principal idLocalidad es de vital importancia ya que la información del ALEC tanto de carácter lingüístico-etnográfico como adicional (audios, fotografías,

ilustraciones, muestras de habla espontánea, etc.) se encuentra georreferenciada mediante la misma codificación.

Los siguientes metadatos son Latitud y Longitud. Estos permiten la creación del punto sobre el polígono del mapa de Colombia para la ubicación de la información. Los demás metadatos que describen las localidades son: Año Fundación: que normalmente se refiere al año, pero en casos en los que no se contó con la información exacta, fue reemplazado por enunciados como: ‘Es poblado de fundación reciente (en el siglo XX)’ o ‘Anterior a la conquista española’ (Flórez, 1983); ASNM (altura sobre el nivel del mar): numérica y dada en metros, Temperatura: numérica, en grados centígrados; Área: numérica y dada en kilómetros cuadrados; Población: número de habitantes con relación al censo de 1968 (Flórez, 1983, 73); Comentarios: descripciones topográficas o sociohistóricas del lugar; Actividades Económicas: la información fue normalizada y se ingresaron cada uno de los nombres de las actividades desarrolladas separados por coma a diferencia de como aparece en el Manual, donde se hace una introducción del tipo ‘La economía consiste principalmente en...’; Vías de acceso: se distinguieron las formas de acceso carretera, férrea, aérea, fluvial, marítima con relación a los lugares de acceso; Fecha Encuesta: en localidades que fueron visitadas por periodos de más de un día, las fechas consecutivas se marcaron con un guion (25/12/1975-27/12/1975). En el caso de las fechas no consecutivas o de encuestas suplementarias, fueron separadas con punto y coma (25/12/1975; 06/08/1976); Observaciones: este metadato contiene información adicional dada por los investigadores como el tipo de encuesta aplicado, por ejemplo “primera encuesta con cuestionario reducido” (Flórez, 1983: 169).

b) Tabla Encuestadores

El Manual contiene información acerca de los años de vinculación al proyecto y de cada una de las localidades en la que los 23 encuestadores realizaron trabajo de campo.

La clave principal de la tabla es idEncuestador y es un número entero autoincremental que se asigna de acuerdo con su aparición en el manual del ALEC. Vale resaltar que este orden concuerda también con la cantidad de localidades encuestadas y tiempo de vinculación de mayor a menor. Los demás metadatos son: Nombres y Apellidos que permiten además de la identificación generar hipervínculos a su bibliografía y consulta de textos en la biblioteca del Instituto Caro y Cuervo; añosvinculación: al igual que en el metadato fechaencuesta, de la tabla localidades, se utiliza guion para fechas consecutivas y punto y coma para fechas no consecutivas; todaslaslocalidades: éste metadato se refiere al código de uno o dos caracteres de los departamentos que cubrió en su totalidad el encuestador; localidades: identifica de forma específica cada uno de los lugares del país en los que el encuestador realizó una encuesta; complementarias: se refiere a localidades donde se realizaron encuestas posteriores a las de la primera visita; esto se debía, generalmente a falta de datos o

duda sobre los mismos. La información correspondiente a los metadatos localidades y complementarias fue llenada con base en la codificación del metadato idLocalidades de la Tabla Localidades. Por último, se encuentra el metadato total que se refiere al número de localidades visitadas por el encuestador.

c) Tabla Informantes

Para la selección de los informantes en el ALEC se tuvo en cuenta que fueran nativos de la localidad, tener poca instrucción escolar o ser analfabetas, tener entre 40 y 60 años y ser de origen campesino (Flórez, 1983). Con base en lo anterior se dispuso la recolección de datos de carácter socioeconómico de los hablantes y, a modo de tabla poblacional, características que para el objetivo de digitalización de los materiales funcionan como metadatos. En el Manual también aparecen consignados los cuestionarios de la encuesta que respondió cada informante, lo que permite saber con precisión en cada uno de los mapas quién o quiénes dieron respondieron una pregunta.

De la misma manera que los encuestadores, cada informante se vincula mediante el idLocalidad con la localidad donde fue encuestado. A cada informante se le asignó un identificador mediante el metadato idInformante, que es la clave primaria de la tabla. Este se compone de un máximo de siete caracteres, de los cuales cinco corresponden al metadato Idlocalidad de la tabla localidades y los siguientes al número entero incremental según el orden de aparición en el manual. Por ejemplo, el primer informante de la localidad Acandí tiene como identificador el código Ch01_1, Ch01_ del idLocalidad y 1 por ser el primero de la lista.

Los demás metadatos que describen a los informantes corresponden a Nombre y Apellidos; Edad; Género: se marcó la letra m para masculino y la letra f para femenino, aunque no es explícito en el Manual, el género fue identificado de acuerdo con el nombre e información contextual, por ejemplo, en los oficios se marca la flexión de género: agricultora, minera, costurera etc.; otros metadatos son Ocupación; Nivel Escolaridad; Proveniencia, Proveniencia de los padres, Proveniencia del cónyuge y Viajes. Por último, se relacionan los cuestionarios contestados por el informante mediante el código correspondiente en la Tabla Cuestionarios que contienen un código de dos o tres letras por cada uno de los campos semánticos del ALEC.

2.3.2. Tablas de láminas

Se diseñaron formularios para facilitar el ingreso de los datos de las láminas e información adicional y suplementaria con base en la estructura de datos de la BDE. Los formularios fueron establecidos con el fin de mantener la división de cada una de las entidades, específicamente, la información general de las láminas, las variantes y su

georreferenciación, información adicional y suplementaria. A continuación, se describe cada uno de los formularios de forma detallada.

a) Hoja láminas

Tabla 2. Diseño de tabla con metadatos para ingreso de datos generales de láminas.

Tomo	Campo Semántico	Lámina	Mapa	Título	Nombre Científico	Francés	Inglés	Referencia a otros Atlas
2	AS	206	188	Copetón	Brachypiza pileata	moineau	Crested sparrow	ALEA II, 404

Como podemos observar en la Tabla 2 los metadatos brindan información general sobre la ubicación y el contenido de la lámina. El primer metadato es Tomo y corresponde a un número de uno a seis. Seguido de esto identificamos el metadato Campo Semántico que permite referenciar la lámina con uno de los diecinueve cuestionarios descritos en el punto 2.1. usando la convención que corresponda en la Tabla Cuestionarios; teniendo en cuenta que todo mapa es a su vez una lámina, pero una lámina puede contener un mapa, texto o imágenes. Se establecieron los metadatos Lámina y Mapa que corresponden con un número de tres dígitos y es igual a la numeración dentro de los tomos impresos, cuando el número es de un dígito se agregan dos ceros a la izquierda y cuando es de dos se agrega un cero para tener un total de tres caracteres.

El siguiente metadato es Título cuyo contenido es la entrada principal o título del mapa. En el caso de láminas de texto o imágenes sin título se completó con el número cero para indicar falta de información. Los demás metadatos que describen las láminas son Nombre Científico: sólo en el caso de animales o plantas; francés: traducción al francés del título de la lámina; inglés: traducción al inglés del título de la lámina y, por último, Referencia a otros Atlas que permite identificar la misma entrada en otros atlas lingüísticos ya publicados para 1983 como el Atlas Lingüístico Etnográfico del Sur de Chile (ALESUCH), el Atlas Lingüístico y Etnográfico de Andalucía (ALEA) entre otros.

b) Hoja variantes

Tabla 3. Diseño de tabla con metadatos para ingreso de información de variantes de una lámina.

IdVariante	Variante	Símbolo	Cartografiada
1	guarapo	ES23	1
2	caldo	PA11	1
3	agua (de caña)	CI25	1
4	dulce	0	0
5	jugo	0	0
6	melao	0	0
7	zumo	0	0

En la hoja variantes (Tabla 3) se ingresan y codifican cada una de las variantes léxicas, fonéticas o etnográficas. El metadato principal idVariante es un número entero incremental que se asigna a cada variante, sin importar si está cartografiada sobre el mapa o en la sección de otras respuestas. idVariante fue utilizado para su relación cartográfica con la clave primaria IdLocalidades de la Tabla Localidades. El siguiente metadato es Variante y allí se incluye el texto de cada una de las variantes léxicas, fonéticas o etnográficas. El siguiente metadato es Símbolo. Para la representación de las variantes léxicas y etnográficas en el mapa fueron diseñadas figuras geométricas con diferentes rellenos y formas. A cada uno de estos símbolos, después de su digitalización y vectorización, se les asignó un código para que en las futuras visualizaciones del ALEC en línea sea posible mantener tal distinción.

El último metadato de la hoja variantes es llamado Cartografiada. Este nos permite saber si en principio la variante se encuentra cartografiada sobre el mapa y por lo tanto cuenta con un símbolo o, por el contrario, se encuentra en la sección de otras respuestas. En caso de estar cartografiada se asigna el número uno y en caso opuesto el cero.

Con relación a los datos fonéticos fue necesario realizar una tabla de correspondencia entre los símbolos fonéticos utilizados por los investigadores del ALEC, basados en el alfabeto fonético de la Revista de Filología Española, y el actual Alfabeto Fonético Internacional. Como se observa en la Tabla 5 cada uno de los símbolos fue representado en caracteres digitales utilizando la extensión IPA compatible con el formato de codificación UTF-8, que utiliza símbolos de longitud variable y después, con base en la descripción dada en el ALEC, se estableció su correspondencia en el Alfabeto Fonético Internacional (AFI), con el fin de brindar a los usuarios los dos tipos de búsqueda y representación.

Tabla 4. Tabla de correspondencias ALEC/ARFE a AFI.

ALEC/ARFE	DESCRIPCIÓN	ARFE	AFI	DESCRIPCIÓN
	Bilabial oclusiva sorda	p	p	Oclusiva bilabial sorda
	Bilabial oclusiva sonora	b	b	Oclusiva bilabial sonora
	Bilabial fricativa sonora	ɸ	β	Fricativa bilabial sonora
	Bilabial fricativa sorda	φ	ϕ	Fricativa bilabial sorda
	Labiodental fricativa sorda	f	f	Fricativa labiodental sorda

En la Tabla 5 observamos un ejemplo de ingreso de datos fonéticos, teniendo en cuenta las variantes con su respectiva transcripción en ARFE y AFI, el único cambio realizado con respecto a los datos léxicos y etnográficos es que al haber dos versiones de la variante éstas se separan con barra recta.

Tabla 5. Hoja Variantes con datos fonéticos.

Id Variante	Variante	Símbolo	Cartografiada
1	é 'e	0	1
2	é 'e	0	1
3	e e	0	1

c) Hojas variante cartografiada y variante no cartografiada

Tabla 6. Diseño de hoja variantes cartografiadas.

IdLocalidad	V1	V2	Signo Convencional	+	N
N_03_	1	2	, Cv		
S_61_	1	2	,		
Bo02_	0				
Bo03_	1	3	; +	guarapo mosto	1

Una vez le era asignado a cada variante cartografiada el metadato IdVariante en la hoja de variantes se pasaba a la hoja variantes cartografiadas para relacionarlo con el metadato IdLocalidad de acuerdo con su orden de aparición de izquierda a derecha.

Por ejemplo (Tabla 6) para la lámina correspondiente a Guarapo (Tomo 1, Lámina 271, Mapa 249) en la localidad N_03_ aparecen en orden de izquierda a derecha las variantes guarapo que se asigna en el campo V1 y caldo en el campo V2. De igual manera que con los símbolos, se realizó el ingreso de los datos correspondientes a las convenciones utilizadas en la cartografía de los datos lingüísticos y etnográficos mediante el metadato Signo Convencional ya que estas no solo sirven para separar variantes en las localidades, sino que también establecen direcciones a fenómenos lingüísticos o género de los informantes; es distinto si aparece entre dos variantes una coma, ya que esto implica polimorfismo, o dos barras inclinadas, que indican que las respuestas fueron dadas por hablantes de diferente género. En el ejemplo de la localidad N_03_ observamos que hay dos signos convencionales, el primero es una coma que separa las formas equivalentes (polimorfismo, sinonimia) y el segundo Cv que determina que la palabra fue recogida en la conversación, es decir, por fuera del cuestionario. Por otro lado, se respeta el orden de aparición de los signos convencionales y se separan por barra recta.

En la localidad Bo03_ además de la coma podemos observar el signo +, este signo que representa una llamada a las notas particulares que, de acuerdo con el ALEC, son notas al pie del mapa que pueden llevar a varios tipos de información adicional. En la Tabla 6 observamos que dicha información adicional se transcribe en el campo llamado + y seguido de éste la letra N que mediante los números 1 (sí) o 0 (no) nos permite esclarecer si la adición fue puesta con negrilla para destacar una entrada léxica o etnográfica.

Para el caso de las variantes no cartografiadas, que sabemos cuentan con georreferenciación, la hoja cuenta con los mismos metadatos a excepción de los campos para adición (+) y negrilla (N) ya que estas variantes no cuentan con información adicional. De la misma forma se ingresa el IdVariante frente al Idlocalidad en su orden de aparición.

d) Hojas adiciones, adiciones georreferenciadas, suplemento y suplemento georreferenciado.

Tabla 7. Diseño de hoja adiciones-suplemento.

IdAdición	Tipo	Título	Texto	Georef
T1L247M249A1	Texto	Bagazo	Bagazo se respondió en casi todas partes. Variantes recogidas: abasto (S60a), gabazo (H 50, B 1.16, S45a, 61a, M 20.2 1. G 30, Ce 3), material (A 52), ripio (A30, 32), sopas (Bo 31, 45), supia (Ca 20)	1

A cada una de las adiciones se les asignó un identificador en el campo del metadato IdAdición (Tabla 7). El identificador se compone de las letras T y número de tomo (1

digito); L y número de lámina (3 dígitos) M y número de mapa (3 dígitos); A y número de adición que es dado de acuerdo al orden de aparición en la hoja y siguiendo la dirección las manecillas del reloj. Seguido a esto aparece el metadato Tipo que se refiere al contenido de la información adicional: Fotografía; Dibujo; Nota: información adicional relacionada al título de un mapa; Aclaración: usualmente explica particularidades en la presentación de los datos, cómo se obtuvo la información o por qué hace falta; Partitura; Véase: nota que envía al suplemento. Seguido de esto se pone el Título y en Texto el rótulo de la adición, o en caso de ser una Nota, el texto completo. Por último, se identificó si la adición contaba con georreferenciación en el campo Georef mediante los números 1 para sí y 0 para no.

En la hoja suplementos se encuentra el metadato IdSuplemento que se compone de las letras ST, el número de tomo del que es suplemento (1 dígito), M y el número de mapa con el que tiene vínculo (3 dígitos), P y número de página del suplemento (dos dígitos), guion bajo y un número de suplemento dado por orden de aparición (dos dígitos), por ejemplo: ST3M001P15_01. Los demás metadatos son los mismos que en la hoja adiciones.

En el caso de que una adición o suplemento estaban georreferenciados se relacionaban con el Idlocalidad en las hojas Adiciones Georreferenciadas y Suplementos Georreferenciados respectivamente. Primero se ingresaba junto al Idlocalidad el identificador asignado a la adición o suplemento, luego la adición o suplemento correspondiente en caso de ser una cadena de texto o rótulo en los campos campo A1, A2 o S1, S2 de acuerdo con el orden de aparición y por último se establecía si la información dada estaba en negrilla o no asignando los valores 1 o 0.

3. Modelamiento de la base de datos espacial del ALEC

Después de la identificación y descripción de la información mediante la asignación de metadatos, el siguiente paso fue la identificación y descripción de sus relaciones. Estas relaciones reflejan las interacciones lógicas existentes entre entidades. De acuerdo con Fuentes (2012) las relaciones pueden ser de tres tipos:

Relaciones 1-1: Las entidades que intervienen en la relación se asocian una a una (Ej.: la entidad HOMBRE, la entidad MUJER y entre ellos la relación MATRIMONIO); Relaciones 1-N: Una ocurrencia de una entidad está asociada con muchas (n) de otra (Ej.: la entidad EMPRESA, la entidad TRABAJADOR y entre ellos la relación TRABAJAR-EN); Relaciones M-N: Cada ocurrencia, en cualquiera de las dos entidades de la relación, puede estar asociada con muchas (n) de la otra y viceversa (Ej.: la entidad ALUMNO, la entidad ASIGNATURA y entre ellos la relación MATRÍCULA).

En el caso de las entidades del ALEC las relaciones son de tipo 1-N y para su identificación se diseñó una matriz en la que se puso cada una de las entidades en relación con las demás con el fin de establecer su relación en términos de uno, muchos, uno o muchos e identificando la obligatoriedad o laxitud de dicha relación mediante los verbos tener y poder.

Una vez analizadas todas las relaciones de las entidades bajo estos términos se procedió al modelamiento de la BDE (Figura 1) para determinar la estructura lógica y de manera fundamental determinar el modo de almacenar, organizar y manipular los datos del ALEC. La herramienta usada para el modelamiento fue *Navicat Data Modeler para PostgreSQL* que permite diseñar de manera visual las estructuras de las bases de datos. Como podemos observar, este modelo resume el nombre de la entidad y sus respectivos atributos en conexión con cada una de las entidades de acuerdo con la matriz de relaciones. Por el carácter espacial de la base de datos observamos en el medio la entidad localidades con sus respectivos atributos y a su alrededor las entidades informantes en relación de uno a muchos, teniendo en cuenta que en una localidad puede haber uno o muchos informantes. Tomo y Campo Semántico es otro ejemplo de las relaciones entre las entidades. Un Tomo puede tener muchos campos semánticos pero un campo semántico no contiene tomos. Para finalizar, en el caso de las relaciones posibles, que no son obligatorias, observamos que la relación entre la entidad Mapa y Adiciones está dada por líneas con espacios, esto se debe a que un mapa puede tener o no adiciones, no es una relación obligatoria.

tránsito del medio análogo a digital de datos producto de estudios geolingüísticos, lo que demuestra la importancia y necesidad del trabajo interdisciplinar.

El tránsito de la información del ALEC del formato análogo a digital y su consolidación en una BDE es un avance en el desarrollo del SIG ALEC² que permitirá realizar funciones de análisis y consultas y a una visualización en línea a modo de Atlas Web³.

El modelo de la BDE está diseñado de tal manera que a futuro es posible ingresar datos de nuevas investigaciones relacionadas con el ALEC, añadir nuevas localidades, informantes y la creación de nuevos mapas, visualizaciones y análisis de la información lingüística y etnográfica.

En el estado actual de la BDE es posible realizar análisis dialectométricos para establecer o comprobar propuestas de división dialectal de Colombia, de la misma manera gracias a la inclusión de los datos del Manual del ALEC es viable hacer aproximaciones multidimensionales a la información lingüística y etnográfica mediante el cruce de búsquedas, por ejemplo, el cruce de variantes lingüísticas con variables sociales de los informantes.

En el mismo orden, la BDE del ALEC permite iniciar el diseño de un diccionario de colombianismos de carácter diatópico. La BDE del ALEC se puede consolidar en desarrollos tecnológicos de diversa índole a través de múltiples medios y plataformas de comunicación, aplicaciones web para celulares, juegos ilustrativos y pedagógicos sobre el español de Colombia, materiales para el aprendizaje del español como segunda lengua y un sinnúmero de opciones que encuentran límite en las posibilidades tecnológicas que, como sabemos, cada día ofrecen nuevas alternativas.

REFERENCIAS BIBLIOGRÁFICAS

- Alvar, M. (1977). El Atlas lingüístico y etnográfico de la provincia de Santander (España). *Revista de Filología Española*, 59(1), 81-118.
- Alvar, M. (1991). *Geografía lingüística*. Madrid: Paraninfo.
- Bosque, I., Fernández, C., Martín-Forero, L. & Pérez, E. (2012). *Los sistemas de información geográfica y la investigación en ciencias humanas y sociales*. Madrid: Confederación Española de Centros de Estudios Locales.
- Coseriu, E. (1956). *La geografía lingüística*. Montevideo: Universidad de la República.
- Flórez, L. (1961). Nota informativa del Atlas Lingüístico-Etnográfico de Colombia (ALEC). *Thesaurus*, 1(1), 77-125.
- Flórez, L. (1983). *Manual del atlas lingüístico-etnográfico de Colombia*. Colombia: Instituto Caro y Cuervo.

- Fuentes, J. (2012). *Bases de datos relacionales: El modelo relacional* [en línea]. Disponible en: http://csrg.inf.utfsm.cl/~jfuentes/_build/html/lectures/week1/lecture1.html
- García Mouton, P. (2013). El procesamiento informático de los materiales del Atlas Lingüístico de la Península Ibérica de Tomás Navarro Tomás. *Anuario del Seminario de Filología Vasca Julio de Urquijo*, 167-174.
- García Mouton, P. & Moreno Fernández, F. (2003). *Atlas lingüístico y etnográfico de Castilla-La Mancha (ALECMAN)* [en línea]. Disponible en: <http://www.linguas.net/alecman/>
- Guevara, J. A. (1992). Esquema metodológico para el diseño e implementación de un sistema de información geográfico. *Geographicalia*, 29, 21-32.
- Gütting, R. H. (1994). An introduction to spatial database systems. *The VLDB Journal-The International Journal on Very Large Data Bases*, 3(4), 357-399.
- Heap, D. (2003). Las encuestas del ALPI a los cuarenta años de la publicación de su primer tomo. En F. Sánchez Miret (Coord.), *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica* (pp. 201-10). Turbinga: Max Niemeyer Verlag.
- Hoch, S. & Hayes, J. J. (2010). Geolinguistics: The incorporation of geographic information systems and science. *The Geographical Bulletin*, 51(1), 23-36.
- Instituto Caro y Cuervo, Departamento de Dialectología (1982-1983). *Atlas Lingüístico-Etnográfico de Colombia*. Bogotá: Instituto Caro y Cuervo.
- Jaberg, K., Jud, J. & Scheuermeier, P. (1928). *Sprach-und sachatlas Italiens und der Südschweiz*. Zofingen: Ringier.
- Lamigueiro, O. P. (2014). *Displaying time series, spatial, and space-time data with R*. Boca Raton, FL: CRC Press.
- Montes Giraldo, J. (1995). *Dialectología general e hispanoamericana*. Bogotá: Instituto Caro y Cuervo.
- Nadkarni, P. M. (2011). *Metadata-driven Software Systems in Biomedicine*. Londres: Springer.
- Pederson, L. (1996). LAMR/LAWS and the Main Chance. *Journal of English Linguistics*, 24(3), 234-249.
- Ramírez, H. & Herrera, J. F. (2013). Un viaje a través de bases de datos espaciales NoSQL. *Redes de ingeniería*, 4, 35-47.

- Schneider, E. W. & Kretzschmar Jr, W. A. (1989). LAMSAS goes SASsy: Statistical methods and linguistic atlas data. *Journal of english linguistics*, 22(1), 129-136.
- Rigaux, P., Scholl, M. & Voisard, A. (2001). *Spatial databases: With application to GIS*. California: Morgan Kaufmann.
- Sinclair, J. (1996). *Preliminary recommendations on corpus typology* [en línea]. Disponible en: <http://www.ilc.pi.cnr.it/EAGLES/corpusyp/corpusyp.html>
- Surve, R. B., Bhaskar Y. & Nehru K. K. (2014). Disparity of spatial and non-spatial data. *International Journal of Advance Foundation and Research in Computer (IJAFRC)*, 1(8), 2348- 4853.

NOTAS

¹ Ejemplo (pregunta): especie de taza grande con poco fondo que sirve principalmente para lavarse las manos y la cara. Respuestas: platón, ponchera, porcelana, aguamanil, etc. (Flórez, 1983: 60).

² <http://atlasweb.caroycuervo.gov.co/>

³ <http://alec.caroycuervo.gov.co/alec/>