

INSTITUTO	SOCIOAMBIENTAL
data	10, 09, 98
cod.	A2D 00034

1. Nombre del Proyecto

Elaboración de una Base de Datos Lingüísticos y cartografiables del Centro Colombiano de Estudios de Lenguas Aborígenes (BDCCELA) con fines científicos comparativos y aplicados - Resultados y perspectivas

2. Antecedentes

2.1. Dispositivo General

Desde noviembre de 1991 en el Centro Colombiano de Estudios de Lenguas Aborígenes se viene desarrollando el proyecto de una Base de Datos lingüísticos y cartografiables sobre diferentes lenguas colombianas. Se ha propuesto adelantar el trabajo lingüístico a partir del diseño, conceptualización, formulación y elaboración de cuatro tipos diferentes de cuestionarios: Fonético-Fonológico, Lexical, Gramatical y Sociolingüístico. Se decidió trabajar en cada cuestionario por separado, dada la complejidad de la tarea; se comenzó el proyecto con el Cuestionario Fonético-Fonológico, el cual estuvo listo para su aplicación en el segundo semestre de 1992 (Cf. Anexo 1).

Después de escoger treinta y dos (32) puntos fuentes de información sobre la diversidad lingüística colombiana (ver Mapa N° 1), se aplicó en cada uno de estos puntos el cuestionario Fonético-Fonológico que, como se anotó más atrás, fue conceptualizado y diseñado en el marco de este proyecto. Se han tomado como base los puntos en donde se encuentran investigadores del CCELA; para el caso del cuestionario fonético-fonológico y después de una primera recolección de datos, se convocó a una reunión de todo el grupo de investigadores para examinar problemas de formulación, ambigüedades, dificultades de recolección, traslapes, etc.. Aunque este proyecto podría parecer como meramente de documentación, la discusión sobre la formalización de datos para su entrada en la base ha permitido un trabajo científico colectivo muy valioso en y para el CCELA; además es preciso decir que las dificultades de manejo conceptual y/o terminológico, se minimizaron gracias a los debates propiciados sobre estos temas.

Los datos fonéticos y fonológicos entregados fueron sometidos a estudio por dos lectores (lingüistas idóneos) para garantizar un manejo sistemático de las respuestas; los

informes de lectura fueron remitidos al investigador comprometido y si se precisaban correcciones, éste las hacía. Luego fueron introducidos en el Sistema de Gestión de Datos (cf.2.3.2). Una vez digitados se ha entrado en un proceso de revisión (en pantalla o impresión) por cada uno de los investigadores, con el fin de verificar la correcta transcripción.

A continuación se presenta la lista actual de investigadores y lenguas del BDCCELA:

Investigador	Lengua	Familia	Lugar Enc.	Departame.
1 AGUIRRE Daniel	1 Embera-chamí	Chocó	Cristiania	Antioquia
	2 Embera	Chocó	Garrapatas	Valle Cauca
	3 Embera	Chocó	Sambú	Panamá
2 GOMEZ-IMBERT Elsa	4 Tatuyo	Tukano	Piraparaná	Vaupés
	5 Barasana	Tukano	Piraparaná	Vaupés
3 HOYOS Mario	6. Embera-Pcf sur	Chocó	Esmeralda	Ecuador
4 LANDABURU Jon	7 Andoke	Andoke	Q. Aduche	Amazonas
5 LLERENA Rito	8 Embera	Chocó	Jaidukama	Antioquia
	9 Embera	Chocó	Alto Andágueda	Chocó
	10 Kuna	Chibcha	Caiman Nuevo	Antioquia
6 MELENDEZ Miguel	11. Achagua	Arawak	Umapo	Meta
7 MERCHAN Ana Joaquina	12. Kuiba-Maiben	Guahibo	S José Ariporo	Casanare
8 MOGOLLON M. Cristina	13. Barí	Chibcha	Birikaira	N. Santander
9 MONTES Ma. Emilia	14. Ticuna	Ticuna	Amacayacu	Amazonas
10 NIEVES Rocío	15. Páez	Páez	Caldono	Cauca
11 O'FLYNN Carol	16. Criollo (S Andrés)	Criollo	San Andrés	San Andrés
12 ORTIZ Carolina	17. Kogui	Chibcha	Santa Rosa	SN Sta Marta
13 PABON Marta	18- Guambiano-Totoró	Guambiano	Pedregal	Cauca
14 PEREZ Francisco	19. Wayuu	Arawak	Santa Clara	Guajira
15 QUEIXALOS Francisco	20. Sikuaní	Guahibo	Santa Rita	Vichada
16 REINOSO Andrés	21. Piapoco	Arawak	Lag. Colorada	Guainía
17 ROBAYO Camilo	22. Carijona	Caribe	Miraflores	Guaviare
	23. Yuko	Caribe	Iroka	Cesar
	24. Páez	Páez	Munchique-T	Cauca
18 ROJAS Tulio	25. Guayabero	Guahibo	Barranco color.	Guaviare
	26. Tinigua	Tinigua	Macarena	Meta
20 TRILLOS María	27. Damana	Chibcha	Abingüe	SN Sta Marta
	28. Chimila	Chibcha	Issaoristunna	Magdalena
	29. Guambiano-Quizg	Guambiano	Quizgó	Cauca
21 TRIVIÑO Lilia	30. Kubeo	Tukano	Kuduyarí	Vaupés
22 VALENCIA Simón	31. Guambiano-Guam.	Guambiano	Guambía	Cauca
24 ZALABATA Rubiel	32. Ika	Chibcha	Séynimin	SN Sta Marta
	32 Puntos encuesta 24 Lenguas 6 var Embera 3 var Guambiano 2 var Páez.	Fam.: 6 Arawak: 3 Lenguas (9) Caribe : 2 Lenguas (2) Chibcha: 6 Lenguas (7) Chocó : 1 Lenguas (2) Guahibo: 3 Lenguas (4) Tukano: 3 Lenguas (18) Indepen: 5 Lenguas: Andoke, Guambiano, Páez, Ticuna, Tinigua Criollo: 1 (2)		

El proyecto tiene dos componentes importantes:

2.2. Componente lingüístico: para el montaje de esta parte se ha trabajado con un equipo permanente de concepción y ejecución compuesto por Jon LANDABURU, Miguel

Angel MELENDEZ y Tulio ROJAS CURIEUX, un comité local de investigadores del CCELA y el grupo general de investigadores asociados a este proyecto. Primero se diseñó el Cuestionario Fonético-Fonológico, el cual consta de 18 rubros (con diferente número de preguntas cada uno) y está acompañado de un Glosario que permite resolver las principales dudas que se generen en su diligenciamiento. Con este cuestionario se pretende tener un acercamiento detallado a las estructuras sonoras de las lenguas estudiadas en este proyecto; interesa conocer la realidad fonética de cada una de ~~las~~^{ellas} lenguas y la organización de los fonemas en términos de rasgos pertinentes para poder hacer estudios comparativos que permitan establecer características areales y de familias lingüísticas. Este cuestionario ya ha sido aplicado en la mayoría de puntos de encuesta (23 de los 32 - Cf. Anexo 2).

Después del Cuestionario Fonético-Fonológico se replicó nuevamente la labor preparatoria (diseño, conceptualización, formulación y elaboración) para el Cuestionario Lexical. Este consta de 3.360 palabras agrupadas en tres capítulos (Cf. Anexo 3). En su elaboración se han encontrado más dificultades que en el caso del cuestionario Fonético-Fonológico, dado que la teoría lingüística no dispone, en la actualidad, de una guía lexical general que pueda ser aplicada a diferentes lenguas y que a su vez permita generalizar observaciones a través de las mismas. Por lo tanto, se organizó un primer trabajo a partir de nociones enciclopédicas que permitiera adecuar las expresiones (en español) a nuestro conocimiento del universo léxico de las lenguas amerindias del proyecto. Este cuestionario se encuentra en proceso de aplicación y ya se tienen datos de 24 de los 32 puntos de encuesta (Cf. Anexo 4). Es claro que este cuestionario tendrá cambios importantes en el futuro.

2.3. Componente computacional:

2.3.1. Consultas preliminares y configuración del equipo: organizamos reuniones y contactos con unidades especializadas de la Universidad de Los Andes (Programa Hermes, Centro de cómputo, departamento de Sistemas de la Facultad de Ingeniería) y otros especialistas en tratamiento de datos y organización de mapas en computador. Después de realizar estas consultas se decidió por un ambiente de trabajo en Macintosh y para tal efecto se adquirió un Computador IISI el cual debió ser sustituido por un Power Macintosh 7200/90 con las siguientes especificaciones técnicas: Unidad

Principal : Procesador PowerPC 601, Memoria: 16 MB RAM y 500 MB de disco duro, Unidades de disco: Unidad de disco flexible de alta densidad de 1,4 MB SuperDrive, Unidad de disco rígido interna SCSI Apple, CD-ROM drive, Generador de Sonido: Chip de sonido (análogo y/o digital); Interfaces: Un puerto Apple Desktop Bus (ADB), Puerto de video, Ranura de expansión interna, dos puertos serie RS-232/RS-422, Interfaz SCSI, Puerto de salida de sonido estereo. Este equipo se conectó con un Monitor Color Macintosh 16" con Tarjeta de Video Macintosh 8•24.

2.3.2. Organización del Sistema de Gestión de Datos: una vez definido el ambiente de trabajo y para organizar el sistema de tratamiento de los datos en el computador, se decidió aprovechar de la mejor manera las aplicaciones y las ventajas que el equipo nos ofrece. Se ha preferido una pantalla en color de 16" dado que buena parte de la información se presentará en mapas que permitan una rápida visualización de la complejidad lingüística, pues el tener una pantalla muy pequeña dificultaría esta tarea; esta pantalla permite una óptima utilización de las aplicaciones dada la posibilidad de tener varias ventanas abiertas y si es necesario mantener diversas aplicaciones activas.

Se ha privilegiado la presentación de los datos en mapas (generales, areales o por familias) ya que ello permite establecer áreas de localización de un determinado rasgo o proceso, así como las posibles expansiones e influencias, tanto en la parte fonética-fonológica, como en la parte lexical. Esta forma de trabajo permite avanzar rápidamente en la definición de procedimientos de comparación, sea porque se defina uno o varios rasgos, sea porque se decida comparar y/o superponer varios mapas, o porque se decida tomar datos de diferentes mapas y colocarlos en uno nuevo, etc. Para tal efecto se ha trabajado en función de los Cuestionarios existentes. Con antelación se consideró el navegar libremente entre las tarjetas con miras a facilitar no sólo la consulta de la Base sino su utilización como herramienta de trabajo en posteriores investigaciones.

Cuestionario Fonológico: este Sistema se ha establecido con base en la aplicación HYPERCARD para Macintosh. Para entrar a la Base de Datos se han previsto diferentes formas: por Familias Lingüísticas, por Lenguas, por Investigadores, por Datos lingüísticos o por mapas (generales, genéticos o areales). También se ha previsto una forma de navegar dentro de la Base que responda a las diferentes necesidades de los

usuarios del programa; así, se puede pasar de una tarjeta de respuesta en una lengua determinada al mapa genético o al areal según quiera el investigador; además es posible buscar una lengua particular dentro de un stack definido. Se han organizado dos tipos de tarjetas de datos lingüísticos: una tal como está presentado el cuestionario en el formato diligenciado por los investigadores y otra en mapas (generales, areales, genéticos). Cada una de las preguntas del cuestionario Fonético-Fonológico está relacionada con trece mapas (uno general de Colombia, siete mapas genéticos y cinco mapas areales) en los que se encuentra la respuesta. Esto convierte a la Base de Datos en una valiosa herramienta no sólo de presentación de resultados de investigación, sino que mediante la posibilidad de correlacionar los mapas se abre una interesante e ilimitada perspectiva investigativa y científica.

Presentación de los datos en Mapas: se ha trabajado en un mapa general de Colombia, en mapas por familia lingüística y en mapas por área. El mapa general de Colombia permite una visión global de los datos relativos a las lenguas en la Base; los mapas areales presentan la diversidad lingüística según cinco áreas predefinidas (Norte, Llanos Orientales, Amazonas, Occidental y Andina-suroccidental); los mapas por familias presentan los datos de las lenguas individuales en su correspondiente familia lingüística (Arawak, Caribe, Chibcha, Chocó, Guahibo, Tukano y lenguas independientes).

Cuestionario Lexical: Luego de analizar los avances que se han encontrado en el diseño del Sistema para el anterior cuestionario y tratando de resguardar la versatilidad y sencillez de manejo de la Base, se se ha buscado continuar con la misma aplicación (HYPERCARD) e introducir algunos cambios en los campos y la manera de manejar la información. El Sistema de Gestión de Datos para el Cuestionario Lexical está en proceso de concepción. Se ha previsto que en este caso ya no se tenga una entrada por Cuestionario sino un archivo con los diferentes ítems que pueden ser consultados; una vez solicitado un ítem, la respuesta aparecerá en un mapa general y de allí se podrá navegar a los genéticos o areales.

Dentro de los mapas areales y genéticos se diseñaron campos específicos para la presentación del ítem en cada lengua de la Base y se han previsto campos que podrán

ser desplegados a voluntad por el usuario, dentro de un cartucho, de tal manera que se puedan visualizar al mismo tiempo -o sucesivamente- tanto la segmentación morfológica del ítem como las notas relativas a ítems relacionados y/o la identificación del referente.

Se ha pensado organizar dos tipos de tratamiento de los datos lexicales: uno en la aplicación FileMaker y el otro en Hypercard. FileMaker es una aplicación que permite manejar Bases de Datos y encontrar las informaciones requeridas de una manera rápida y flexible. La organización de los campos y las disposición de las tarjetas en este tipo de documentos se pueden realizar de una manera ágil y acorde a las necesidades del trabajo. Por su parte, Hypercard permite las facilidades de presentación de resultados tal como ya se ha diseñado para el Sistema de Gestión del Cuestionario Fonético-Fonológico.

La puesta en marcha de este tipo de cuestionario ha introducido discusiones respecto de las posibilidades de establecer parámetros de comparación entre lenguas tan disímiles como las presentes en la Base de Datos.

El siguiente es el estado actual de la Base de Datos en su Sistema de Gestión de Datos:

BASE DE DATOS - CCELA

BDCCELA

10-Nov-95

	Stack (pila)			
	Elab.	Nombre	Número Tarjetas	Número Fondos
Información General				
Home-BDCCELA	SI		42	
Home 2	SI		5	
Home 3	SI		5	
Infos	SI		236	
Mapas Generales	SI		24	
Sub-Total Inf. General			312	
Cuest. Fonético-Fonológico				
A. Formulario Fonét-Fonolog.				
1.Fonemas Consonánticos	SI	idem	129	4
2.Fonos Contoides	SI	idem	231	8
3.Reglas de Realización C	NO			
4.Fonemas vocálicos	SI	idem	129	4
5.Cuadro fonos vocoides	SI	idem	40	5
6.Reglas de Realización V	NO			
7.Procesos fonológicos	SI	idem	64	2
8.Neutralización	SI	idem	96	3
9.Frecuencia y rendimiento	SI	idem	32	1
10.Distribución defectiva	SI	idem	32	1
11.Fluctuación de fonemas	SI	idem	32	1
12.Fonología expresiva	SI	idem	32	1
13.Sílaba	SI	idem	224	7
14.Rasgos suprasegmentales	SI	idem	288	9
15.El morfema	SI	idem	160	5
16.Morfología	SI	idem	192	6
17.Otros tipos de alomorfos	SI	idem	32	1
18.La palabra	SI	idem	96	3
Sub-Total Form Fonét-Fonol.			1809	61
B. Mapas Generales	SI	Mapas 1,2 Fonemas Con.	669	45
Sub-Total Mapas Generales			669	45
C. Mapas Familias	SI	Familia Arawak	669	44
	SI	Familia Caribe	669	44
	SI	Familia Chibcha	669	44
	SI	Familia Choco	669	44
	SI	Familia Guahibo	669	44
	SI	Familia Tukano	669	44
	SI	Familia Independientes	669	44
Sub-Total Mapas Familias			4683	308

	Stack (pila)			
	Elab.	Nombre	Número Tarjetas	Número Fondos
D. Mapas Zonas	SI	Zona Norte	669	44
	SI	Zona Llanos Orientales	669	44
	SI	Zona Amazonas	669	44
	SI	Zona Occidental	669	44
	SI	Zona Andina Sur-Occidental	669	44
Sub-Total Mapas Zonas			3345	220
Sub-Total Mapas			8697	573
Sub-Total Cuest. F-F			10506	634
Gran Total BDCCELA			10818	

3. Objetivos

Se trata de la concepción, elaboración, montaje y control de una Base de Datos lingüísticos que además se representan en diferentes tipos de mapas (genéticos, tipológicos o areales).

3.1. Generales

La herramienta que se encuentra en proceso de construcción permitirá:

- Acceso rápido a datos lingüísticos estandarizados y numéricamente importantes relativos a la diversidad lingüística considerada.
- Representación en mapas de la distribución geográfica de estos datos para facilitar la investigación tipológica, comparativa e histórica.
- Publicación o difusión de partes de estos datos, cartografiados o no, en función de las necesidades de las investigaciones (lingüística, arqueológica, histórica, antropológica, etc.), de las organizaciones indígenas, de las agencias del Estado, de los programas educativos, etc.

Se trata de un programa de carácter instrumental que busca volver más accesibles y utilizables los datos existentes con el fin de responder a múltiples finalidades (científicas y técnicas) inducidas por la realidad multicultural.

3.2. Específicos

Presentamos los objetivos que tenemos para el desarrollo del trabajo. Dado que nos hemos propuesto trabajar en cuatro tipos de cuestionarios y solamente hemos avanzado en dos (Fonético-Fonológico y Lexical) hemos decidido reorganizar nuestro trabajo para poder avanzar en el análisis conjunto del material recopilado, diseñar el Sistema de Gestión de Datos para la parte léxica y proyectar los otros formularios propuestos. En ese sentido tenemos los siguientes objetivos específicos:

3.2.1. Cuestionario Lexical

- Conformación de un equipo de trabajo responsable de la revisión de los Cuestionarios Lexicales recogidos.
- Evaluación del proceso de recolección de datos, dificultades teóricas, prácticas y metodológicas.
- Organización del Sistema de Gestión de Datos para el tratamiento de los datos lexicales:
 - Sistema de consulta en Filemaker
 - Sistema de mapas en Hypercard
- Selección de un número de palabras para grabar y colocar como ilustración sonora en la base. Este trabajo será de carácter experimental con miras a tener no sólo el material sino una muestra lo más representativa posible de la realidad fonética de cada una de las lenguas.
- Organización de un seminario general de todos los investigadores asociados al proyecto para trabajar sobre y con el Cuestionario Lexical - Escoger un campo de trabajo y establecer posibles rasgos y patrones de comparación (genética y/o areal) - Discutir sobre los problemas teóricos y prácticos de la traducción entre lenguas con diferente conceptualización sobre el mundo.

3.2.2. Cuestionario Gramatical

- Conformación de un equipo de trabajo responsable de la concepción y discusión correspondiente a la investigación gramatical requerida.
- Avanzar en la clarificación de la problemática que un cuestionario de este tipo representa en la lingüística general y en la lingüística amerindia

- Organización de cuestionarios sobre: Tipos de Predicación, Estructuras Actanciales, Tipos de Oración, Negación, etc.
- Realización de un Seminario general de discusión y ejercicios de aplicación del Cuestionario Gramatical a las diferentes lenguas.
- Diseñar y experimentar un Sistema de presentación y gestión de los datos gramaticales en el Sistema General de Gestión de Datos de BDCCELA.

3.2.3 Cuestionario Sociolingüístico

- Definir los temas que debe contener este tipo de cuestionario
- Avanzar en la elaboración de un Cuestionario Sociolingüístico con miras a ser aplicado en poblaciones indígenas y criollas de Colombia
- Diseñar una guía metodológica para la aplicación del cuestionario
- Explorar las posibles maneras de presentación de los datos para utilización en el Sistema de Gestión de Datos.

4. Impacto Científico

Para describir adecuadamente una lengua debemos tener en cuenta que siempre existirá la oposición entre el léxico y la gramática. El léxico está sometido a las leyes que dicta la gramática; además la conjunción de léxico y gramática es condición para la producción y/o comprensión de las oraciones de la lengua.

En el léxico -dentro de una misma lengua- se reconocen dos límites teóricos (y además variables) como son: léxico total y léxico común. Cualquier frontera entre ellos es arbitraria y de carácter eminentemente práctico; por lo tanto, si el proyecto además de adentrarse en el léxico de cada una de las lenguas consideradas (recordar que en la Base hay lenguas que están siendo estudiadas en varias de sus variantes) logra avanzar en formulaciones de carácter general y que sean válidas para diferentes lenguas, se habrá avanzado un poco más en el conocimiento del funcionamiento del lenguaje humano y en la particularización dentro de cada lengua estudiada. En el componente computacional, el construir una herramienta que permita consultar los ítems lexicales pasando de una lengua a otra, representa una ayuda para la investigación básica y para la investigación aplicada: si además de la consulta tipo lista se tiene la ayuda de los

mapas (general, genético o areal) las posibilidades de encuentro de diversas disciplinas se multiplican.

Avanzar en la dilucidación de estructuras gramaticales dentro de una sola lengua es tarea dispendiosa para el lingüista. La tarea tiene una mayor envergadura cuando se trata de encontrar patrones que permitan proponer esquemas de análisis de estructuras gramaticales bastante diferentes. Ya no se trata de analizar las estructuras sonoras o fonológicas, tampoco se trata del análisis de las palabras, se busca llegar al estudio de las diversas relaciones que enlazan los elementos de la oración (análisis intralengua e interlenguas). El lograrlo representará sin lugar a dudas un avance en la investigación lingüística.

Se está trabajando en la elaboración científica de un proyecto de Atlas que permitiría, con un manejo computarizado de cartografía, la constitución de un banco de datos donde se podrían proyectar muchos de los resultados conseguidos en las investigaciones desarrolladas hasta ahora relativas a las lenguas indígenas y criollas del país. La proyección sobre mapas de datos lexicales, gramaticales y fonético-fonológicos permite constituir áreas de convergencia, focos de dispersión y diversos esquemas topológicos que plantean hipótesis interesantes de muchos puntos de vista (tipología lingüística y comparatismo, difusión cultural, etno-historia y proto-historia, etc.). También pueden servir a encañar más racionalmente la diversificación lingüística frente a necesidades de política lingüística.

Con el desarrollo de este proyecto se avanza en la comunicación entre diversos especialistas de lenguas indígenas de tal manera que se mantenga un diálogo continuo y estable, no sólo respecto de la aplicación de investigación en cada comunidad indígena sino de la funcionalidad de un proyecto lingüístico aplicado a diversas realidades sociales y de un manejo computarizado de los datos recolectados. Así pues, se aspira no sólo a crear conocimiento sino a facilitar el acceso a él y al mismo tiempo propiciar la utilización por las diferentes disciplinas afines, de los datos recopilados.

El diseño de un cuestionario sociolingüístico concebido y estructurado en comunidades indígenas y criollas nos permitirá abordar la discusión de una parte de la problemática

relativa a la utilización de las lenguas en diferentes contextos, conocer los diferentes puntos de vista para abordar el estudio de los códigos utilizados en los diversos actos de habla por los miembros de estas sociedades.

Esta Base de Datos debe convertirse en un futuro no lejano, en una herramienta multimedia que integre datos de diversa índole: fonético-fonológicos, lexicales, ilustración sonora de ciertos ítems, datos sociolingüísticos, etc.; por todo esto nuestro trabajo está aún lejos de concluir.

5. Impacto Social

En estos momentos en que las lenguas indígenas tienen reconocimiento como lenguas oficiales se precisa avanzar en el conocimiento de cada una de ellas. Las organizaciones indígenas y las agencias del Estado están preocupadas por encontrar elementos que les permitan apuntalar procesos de etnoeducación y revitalización social. Este proyecto en la medida que permite trabajar a cada uno de los investigadores en su lengua y que incita en la búsqueda de perspectivas comunes de presentación de resultados podrá contribuir en estas tareas.

Además, en la medida que se construya la herramienta, nuevas posibilidades investigativas y prácticas se irán abriendo. Se podrá trabajar en elaboración de material para ser utilizado en el ambiente escolar o fuera de él.

6. Aspectos metodológicos

6.1. Marco teórico

La importancia del plano del contenido para el análisis lingüístico ha sido propuesta por muchos estudiosos del lenguaje y las lenguas. Un sistema de significados es intrínseco a la noción de un sistema de lengua; por lo tanto describir una lengua es -entre otras cosas- exponer su sistema de significados. Bloomfield encontró un obstáculo bastante poderoso en su afán de dar a la lingüística el rigor de otras ciencias y a propósito planteó: "Para dar una definición científicamente precisa del significado de todas las formas de la lengua tendríamos que poseer un conocimiento científicamente preciso de cada cosa del mundo del hablante. La amplitud actual del conocimiento humano es pequeñísima comparado con éste. ... El establecimiento de los significados es, por lo

tanto, el punto débil del estudio de la lengua y continuará siéndolo hasta que el conocimiento humano llegue mucho más allá de su estado actual". De cualquier manera abordar el estudio del significado exige tener un acercamiento a los morfemas de la lengua y haber avanzado en la diferenciación entre lexemas y gramemas.

En el Cuestionario Lexical se ha trabajado en la recopilación de formas mínimas libres, lo que corresponde a una definición de palabra propuesta por Bloomfield. No se buscó hacer un inventario de todos los lexemas de la lengua sino realizar un primer acercamiento a una investigación de este tipo. Por tal motivo cuando se propuso la aplicación del Cuestionario Lexical a las diferentes lenguas, hubo claridad en que se iba a tener una muestra del vocabulario de ellas; se trabajó entonces no sólo en la palabra como tal, sino que además del acercamiento al castellano, se consideró como fundamental la segmentación de los diferentes componentes, sean estos lexemas o gramemas.

El vocabulario de una lengua (en el estado de la investigación adelantada) es una totalidad semánticamente articulada, estructurada. De tal manera que en el trabajo adelantado no se trata exclusivamente de la ordenación alfabética del vocabulario, sino de la ordenación del material lingüístico en grupos de palabras que designan un mismo sector de la realidad (recodar que son los lexemas las unidades que por el momento nos interesan) y luego en la medida de la evaluación de los resultados, avanzar en el establecimiento de criterios de ordenación que permitan ir configurando campos semánticos (interna dentro de cada lengua y más adelante una aproximación ordenada entre lenguas). Pero deberá estar presente que en primera instancia nuestra pretensión es de recopilación lexical y posteriormente (en la medida que se profundice) semántica.

El proponerse como otro eje de trabajo la construcción de un cuestionario de tipo gramatical (morfosintáctico), ha pasado en primera instancia por un filtro de discusión: ¿cuál es el interés de un cuestionario de este tipo?, ¿para qué puede ser útil desde un punto de vista científico o aplicado?, ¿cómo superar la dificultad de desarrollos diferentes en el estudio gramatical de las lenguas por parte de los investigadores?, ¿qué nociones gramaticales se privilegian en dicho cuestionario?, en fin, dada la amplitud y dificultad de concebir el entramado de la gramática con fines de selección en un

cuestionario -hay que recordar que existen varios formularios y tipos de encuestas gramaticales que se han analizado y considerado como pertinentes para el trabajo de campo, pero no para el objetivo del presente proyecto-, se han privilegiado dos focos de atención: el primero, de manera similar al cuestionario lexical, busca desarrollar un instrumento amplio de recopilación de datos gramaticales y se ha denominado "lista de gramemas"; esta lista incluirá un instructivo previo del metalenguaje lingüístico. El segundo, más delimitado, pero no por eso menos exhaustivo, derivado del acopio de información obtenida a lo largo de diez años de investigaciones académicas (Postgrado-CCELA) -no sistematizada en muchos casos, pero sí "latente" en seminarios y trabajos puntuales, o expuesta con claridad en publicaciones (por ejemplo: "Estructuras Sintácticas de la Predicación: Lenguas Amerindias de Colombia", Bulletin de l'Institut Français d'Études Andines, 1994, Tome 23, Nº 3. Trabajo colectivo de varios investigadores del CCELA)- tiene que ver con temas orientadores como la sintaxis de la predicación, las estructuras actanciales, los tipos de oraciones. De esta manera se perfila un engranaje teórico-descriptivo que recoge, por un lado la "lista de gramemas" (integraría aspectos como: forma básica del gramema, alomorfos, caracterización morfológica, categoría gramatical, interpretación semántico-funcional, etc.), y de otro lado, "temáticas gramaticales" específicas como las señalados más atrás (sintaxis de la predicación, estructuras actanciales, tipos de oraciones), las cuales podrían formularse y desglosarse en apartados tales como, por ejemplo, esquemas de oraciones mínimas, diferentes tipos, orden en una oración simple, expresión de la predicación de ubicación, de la predicación de propiedad, expresión del agente, expresión del paciente, manifestaciones por índices intraverbales, por nominales, etc. Esta doble orientación, que es por ahora la más viable, será nuevamente sometida a la discusión interna, pero desde ya se puede decir que abre una gama de posibilidades de trabajo comparativo en frentes "restringidos", como los proyectados a través de la "lista de gramemas", y "ampliados", como los proyectados en las "temáticas gramaticales". No sobra decir que el enfoque teórico privilegiado para esta fase del cuestionario gramatical es el de la sintaxis funcional.

La concepción y elaboración del cuestionario sociolingüístico constituye una de las tareas centrales de la Base de Datos. A diferencia del trabajo de los cuestionarios fonético-fonológico, lexical y gramatical, que en gran medida derivan del análisis a

través de encuestas, el cuestionario sociolingüístico será el resultado del análisis de la lengua en su contexto social. De manera más específica, el habla en la interacción social cara a cara cumple varias funciones dado que un acto de habla "es un fragmento de habla producido como parte de un fragmento de la interacción social" (Hudson 1981:121). En este sentido, la discusión acerca del cuestionario sociolingüístico deberá perfilar cuáles funciones del habla correlacionadas socialmente, son interesantes y posibles de incluir en dicho instrumento. Por supuesto habrá que definir, por lo menos, con relación a los diversos factores sociales relacionados con la diversidad lingüística, las dimensiones del emisor, del receptor y del ambiente, al momento de establecer preguntas pertinentes; por ejemplo, en la dimensión del emisor se sabe que entre los wayú (guajiros) existe un habla masculina y una femenina, este aspecto es crucial y cabría elaborar una pregunta de este tipo en el cuestionario. Con relación al receptor, por ejemplo, sería interesante tal vez preguntar y/o observar, si existen hablas especializadas frente al receptor (habla aniñada, vocabularios especiales de respeto frente a un superior, etc.); o con relación al ambiente, estilos formales o informales. El sexo del hablante, la edad, la sintaxis empleada según la persona a quien se hable, la utilización o no de los pronombres personales en la comunicación cara a cara, las variaciones dialectales, la utilización de diferentes registros en la vida social, niveles de valoración de la lengua por parte de los usuarios, son algunos de los temas que deberán plantearse y evaluarse, entre otros aspectos, en la preparación del cuestionario aludido. Por supuesto que dentro de los planteamientos que abordan la relación del comportamiento verbal y no verbal en la interacción social, habrá que considerar a la Proxémica como campo posible de integración de variables en el Cuestionario. Datos sociolingüísticos indispensables -no siempre bien trabajados en los censos de poblaciones indígenas y criollas- como número de hablantes, multilingüismo, fenómenos de diglosia, uso de lenguas de acuerdo a edad y sexo, son elementos que sin lugar a dudas estarán presentes en el cuestionario sociolingüístico.

6.2. Técnicas de recolección de la información

La recopilación de este tipo de información es tarea bastante dispendiosa pues se trata de colegir la información de acuerdo a particularidades geográficas y culturales.

La mayor parte de la información lexical ha sido recogida mediante encuesta de terreno (guiada por el Cuestionario Lexical) y luego cada lingüista dedicó un buen espacio de trabajo de análisis para poder establecer la segmentación de los ítems que lo ameritan. Además se han tratado de establecer las relaciones de una palabra con otras del mismo cuestionario. Esta es una tarea dispendiosa y no ha sido lograda en su totalidad. Se espera que mediante la definición de un tratamiento computarizado de los datos obtenidos avanzar en esta parte del trabajo. Existe la conciencia de que el cuestionario permite avanzar rápidamente en una encuesta extensa, pero que se deberá tratar de verificar esta información en el uso diario de la lengua. Esta labor requiere mucho más tiempo del hasta ahora empleado por cada uno de los investigadores y por el momento rebasa con creces los límites de este proyecto.

Una vez se definan los ítems que serán grabados se precisa organizar trabajos de campo (o si es posible traslado de informantes a centros urbanos) para recoger esta información. La idea de las grabaciones es la de nutrir la Base de Datos con testimonios sonoros de los ítems del cuestionario. No serán por lo tanto grabaciones espontáneas de los hablantes.

La información del Cuestionario Gramatical (en las partes en que se divida) requiere que el investigador haya avanzado en el análisis gramatical de la lengua en cuestión, pues como ya se dijo se trata de dilucidar y poner en evidencia diferentes estructuras morfosintácticas.

La recopilación de información sociolingüística exige del investigador una sensibilidad a las diferentes manifestaciones *in situ* del habla cotidiana. Si se quiere obtener una información fidedigna de los actos de habla, es claro que el cuestionario sociolingüístico es un instrumento de ayuda valioso, pero no excluye la parte cualitativa de la indagación y la observación para afinar el conocimiento requerido.

6.3. Tratamiento de la información

En este trabajo se tendrán varios niveles: por un lado, el que cada investigador dé a sus datos; por otro lado, el tratamiento que en el marco general del proyecto se dé a todo el conjunto de datos. Nos referiremos específicamente a este último.

Ya se ha definido que los datos se presentarán en mapas, cuyo formato ya ha sido predeterminado. Pero para la búsqueda rápida de ítems se precisa desarrollar archivos en una aplicación que permita una organización autónoma de cada una de las lenguas, pero que al mismo tiempo se encuentren en relación. Esto se podrá realizar una vez avancemos en la entrada de datos en Filemaker. También se ha pensado la posibilidad de una interfase entre los programas Hypercard y Filemaker con el objeto de navegar de una aplicación a otra, o sea de navegar de la presentación en columnas o listas (Filemaker) a la presentación en mapas (Hypercard). En el numeral 2.3.2 Organización del Sistema de Gestión de Datos hemos hablado cómo se ha pensado organizar la información lexical en los mapas.

7. Referencias Bibliográficas

BLOOMFIELD, L.

1964 *Lenguaje*. Traducción del inglés por Alma Flor Ada de Zubizarreta. Universidad Nacional Mayor de San Marcos, Lima-Perú.

BOUQUIAUX, Luc et Jacqueline THOMAS

1976 *Enquête et description des langues à tradition orale*. Tomos I, II, III. SELAF, París.

COMRIE, Bernard & SMITH Norval

1977 *Lingua descriptive studies : Questionnaire*. Lingua 42. North-Holland Publishing Company. Netherlands.

GLEASON, H.A.

1961 *An Introduction to Descriptive Linguistics*. Holt, Rinehart and Winston. New York.

GECKELER, H.

- 1976 *Semántica estructural y teoría del campo léxico*. Strukturelle semantik und wortfeldtheorie, versión española de Marcos Martínez Hernández. Editorial Gredos, Madrid.

GIVON, T.

- 1984 *Syntax a functional-typological introduction*. Volume I. John Benjamins Company. Amsterdam/Philadelphia.

GREENBERG, J.

- 1961 *Universals of Language*. Report of a Conference Held at Dobbs Ferry, New York April 13-15. The M.I.T. Press.
- 1978 *Universals of Human Language*. Volume 4 Syntax. Stanford University Press. Stanford, California.

HAGEGE, C.

- 1982 *La Structure des Langues*. Presses Universitaires de France. París.

HAMMOND Michael, MORAVCSIK Edith and WIRTH Jessica

- 1988 *Studies in syntactic tipology*. John Benjamins Company. Amsterdam/Philadelphia.

HUDSON, R.A

- 1981 *La Sociolingüística*. Biblioteca de lingüística. Editorial Anagrama. Barcelona

KEY, Mary Ritchie (Edit.)

- 1979 *The grouping of South American Indian languages*. Gunter Narr Verlag Tübingen.
- 1991 *Language change in South American Indian languages*. University of Pennsylvania Press. Philadelphia.

ILSON, R.

1987 *A Spectrum of Lexicography*. John Benjamins publishing company.
Amsterdam/Philadelphia.

REY, A.

1980 *La Lexicologie*. Editions Klincksieck, Paris.

SHOPEN, T.

1985 *Language typology and syntactic description*. Volumen I, Volumen II,
Volumen III. Cambridge University Press.

TESNIERE, L.

1982 *Eléments de Syntaxe Structurale*. Editions Klincksieck, Paris.

Presentado por :
Tulio ROJAS CURIEUX
Universidad de los Andes -
Centro Colombiano de Estudios de Lenguas Aborígenes - CCELA
Carrera 1a. este N° 18A-10
Edificio P - segundo piso
Teléfonos: 2824066 Ext 2571 - 2572 / 2848073
Telefax: 2841890 / 2815771
Apartado Aéreo 4976
Santafé de Bogotá - Colombia
correo electrónico : trojas@cdcnet.uniandes.edu.co