



**UNIVERSIDAD DEL MAR
CAMPUS PUERTO ESCONDIDO**

**ELABORACIÓN DE CORPUS DE TEXTOS EN LENGUA
CHATINA**

T E S I S

**QUE PARA OPTENER EL TÍTULO DE
LICENCIADA EN INFORMÁTICA**

**PRESENTA
ARELI CHINCHILLA CORTÉS**

**DIRECTOR DE TESIS
M. EN C. MANUEL ALEJANDRO VALDÉS MARRERO**

PUERTO ESCONDIDO, OAXACA

2011

Dedicatoria

Este trabajo de tesis lo dedico con mucho cariño y amor:

*A **Dios** por darme la vida, salud y darme la fuerza necesaria para continuar luchando con todos los obstáculos que se han presentado a lo largo de la carrera.*

*A mis padres **Salvador Arnulfo Chinchilla Sandoval** y **Bertha Cortés López**, por su paciencia y motivación, a ti papá por el apoyo moral y económico que me brindaste a lo largo de la carrera, y aunque no estás cerca de mí, quiero que sepas que ocupas un lugar en mi corazón, pero muy en especial le dedico mi tesis a mi madre, por haberme apoyado en todo momento, por sus consejos, sacrificio, esfuerzo y por confiar en mí. Los quiero mucho.*

*A mis hermanos **Petris** y **Vladi** gracias por estar conmigo y apoyarme siempre, éste es un logro que quiero compartir con ustedes.*

*A mi sobrinito hermoso **Jafed Chávez Cortés** por haber llegado a mi familia y llenarnos de su alegría, sonrisas y gritos.*

*A mis abuelos **Gelacio Cortés Olivera** y **Odilona López Vásquez** (mis segundos papás), por guiarme en cada momento, por creer en mí y por todo lo que de ustedes he aprendido, de igual manera a **Salvador Arnulfo Chinchilla Morales** y **Juana Sandoval Sandoval** a quienes no he tenido la oportunidad de conocerlos pero que ocupan un lugar muy especial.*

*A mi tío **Gelacio Cortés López** (†) ... siempre te llevaré en mi corazón.*

Agradecimientos

Agradezco a la Universidad del Mar campus Puerto Escondido, por haberme brindado una formación profesional, por los conocimientos adquiridos durante estos años.

Al programa becanet-superior por la concesión de la beca de titulación.

Agradezco a los colaboradores del Proyecto Chatino, por facilitarme documentos e información necesaria para la realización de esta tesis.

A la M. en I. María de Jesús Estudillo Ayala por proporcionarme los archivos necesarios y a todas aquellas personas que muy amablemente cedieron las entrevistas.

Un total agradecimiento a mi director de tesis M. en C. Manuel Alejandro Valdés Marrero por haberme aceptado como su tesista, quiero manifestarle mi gratitud y admiración por su gran apoyo, gracias por su tiempo, por haber guiado el desarrollo de este trabajo y llegar a la culminación del mismo.

A mis revisores de tesis M.T.I. Remedios Fabián Velasco y M. en C. José Francisco Delgado Orta por sus observaciones y por el tiempo dedicado a la revisión de esta tesis desde el protocolo hasta su culminación.

A mis revisores de tesis Lic. Benjamín Gómez Ruiz y M en C. Margarito Mancilla Torres por sus observaciones y por el tiempo dedicado a la revisión de esta tesis.

A todos mis amigos, sin excluir a ninguno, gracias por su amistad, les agradezco el haber llegado a mi vida y el compartir momentos agradables y momentos tristes, gracias por ser mis amigos.

A todas aquellas personas que de alguna u otra forma, colaboraron en la realización del desarrollo del mismo.

Resumen

Este trabajo de tesis se denomina “Elaboración de corpus de textos en lengua chatina”. El objetivo principal consistió en la elaboración de un corpus de escritos en lengua chatina. El corpus se define a partir de los componentes léxicos y categorías sintácticas obtenidas a partir de las palabras comunes en documentos impresos de la lengua chatina, de las variantes de Tataltepec de Valdés, Santa Catarina Juquila y San Juan Quiahije.

Este trabajo se realizó sobre la lengua chatina ya que es importante preservar las lenguas indígenas, porque en ellas están inmersas las culturas y la forma de vivir de los pueblos nativos mexicanos.

El proceso de elaboración del corpus se realizó en varias etapas: recopilación de los textos, captura de los textos, implementación de la base de datos y desarrollo de una aplicación sencilla y amigable. La aplicación se desarrolló con la herramienta Microsoft Visual Basic 6.0.

La aplicación consta de cuatro módulos: El módulo “Corpus de textos” abre los archivos escritos en chatino, los muestra y también permite guardar los lexemas. El módulo “Ver Corpus” guarda los lexemas indicando su traducción, clasificación sintáctica y zona a la que pertenece y muestra el corpus de la base de datos. El módulo “Zona” agrega, modifica y elimina una zona geográfica que representa una variante del chatino. Por último, el módulo “Clasificación” agrega, modifica y elimina una clasificación sintáctica.

El corpus tiene tres zonas geográficas, correspondientes a las tres variantes antes mencionadas. En cuanto a la clasificación sintáctica, se tienen cuatro: adjetivo, pronombre, sustantivo o verbo.

El resultado de esta tesis es el corpus de texto en el que se encuentran almacenados 526 lexemas, cada uno con respectivas traducciones al idioma español, clasificación sintáctica y zona a la que pertenece.

Abstract

This thesis is called “Creation of a Chatino Language Text Corpus”. The main objective was the creation of a corpus, based on texts written in the Chatino language. The corpus is defined by lexical components and syntactic categories obtained from common words in chatino documents, specifically using the dialects from the Tataltepec de Valdés, Santa Catarina Juquila and San Juan Quiahije regions.

The preservation of native languages is important because in them are immersed the culture and way of life of Mexican natives, that is why this work was performed using the chatino language.

The process of creating the corpus was done in several stages: text compilation, text capturing, database implementation and development of a user-friendly simple software application. The application was developed in Microsoft Visual Basic 6.0.

The application has four modules: The “Text Corpus” module opens the files written in Chatino, shows them and allows the lexeme to be saved. The “Show Corpus” module saves the lexeme with its translation, syntactic classification and area where it belongs, also shows the database corpus. The “Area” module adds, modifies and deletes a geographical zone which represents a Chatino variant. Finally, the “Classification” module adds, modifies and deletes a syntactical classification.

The corpus has three geographical zones, each one matching a chatino variant. In reference to syntactical classifications, the corpus has four: adjective, pronoun, noun and verb.

The result of this thesis is a text corpus that stores 526 Chatino lexemes, each one with its own translation, syntactical classification and area of origin.

CONTENIDO

LISTADO DE FIGURAS.....	iii
LISTADO DE TABLAS.....	vi
GLOSARIO DE TÉRMINOS.....	vii
CAPÍTULO 1. INTRODUCCIÓN.....	1
CAPÍTULO 2. ANTECEDENTES.....	5
2.1. Estado del arte y trabajos relacionados.....	7
2.2. Justificación.....	12
2.3. Planteamiento del problema.....	14
2.4. Objetivos.....	17
2.5. Alcances y límites del estudio.....	17
CAPÍTULO 3. MARCO TEÓRICO.....	20
3.1. Lengua chatina.....	20
3.2. Procesamiento del Lenguaje Natural.....	26
3.3. Microsoft Visual Basic.....	33
3.4. MySQL.....	35
CAPÍTULO 4. DESARROLLO DEL TEMA.....	36
4.1. Recopilación de textos en chatino.....	36
4.2. Captura de textos en chatino.....	39
4.3. Diseño e implementación de la base de datos para el corpus de la lengua chatina.....	40

4.4. Desarrollo de la aplicación corpus de texto en chatino.....	43
4.5. Corpus de texto en chatino.....	48
CAPÍTULO 5. CONCLUSIONES Y TRABAJOS FUTUROS.....	51
ANEXO A. CHATINOS.....	55
ANEXO B. TEXTOS CAPTURADOS.....	69
ANEXO C. MANUAL DE USUARIO.....	72
ANEXO D. INSTALACIÓN Y RESTAURACIÓN DE LA BASE DE DATOS.....	84
ANEXO E. CONTENIDO DEL CD.....	102
REFERENCIAS.....	106

LISTADO DE FIGURAS

Figura 2.1.	Mapa de localización de las etnias en Oaxaca.....	13
Figura 2.2.	Densidad de hablantes de la lengua chatina (CIESAS 2008).....	15
Figura 3.1.	Distrito de Juquila y sus municipios.....	23
Figura 3.2.	Distrito de Sola de Vega y sus municipios.....	24
Figura 3.3.	Distrito de Jamiltepec y sus municipios.....	25
Figura 4.1.	Modelo entidad relación.....	42
Figura 4.2.	Pantalla PRINCIPAL.....	44
Figura 4.3.	Pantalla CORPUS DE TEXTOS.....	45
Figura 4.4.	Menús despegables de la pantalla CORPUS DE TEXTOS.	45
Figura 4.5.	Abrir archivos de textos.....	45
Figura 4.6.	Mostrar texto completo.....	46
Figura 4.7.	Listado de Lexemas.....	46
Figura 4.8.	Lexema guardado.....	46
Figura 4.9.	Lexemas guardados y repetidos.....	47
Figura 4.10.	Pantalla VER CORPUS.....	47
Figura 4.11.	Listado de lexemas, zonas y clasificaciones sintácticas existentes en la base de datos.....	47
Figura 4.12.	Lexema guardado.....	48
Figura 4.13.	Extracto del corpus de la base de datos.....	49
Figura A.1	Municipios donde se concentra la población chatina.....	57
Figura A.2.	Localización de Tataltepec de Valdés.....	61
Figura A.3.	Localización de Santa Catarina Juquila.....	63
Figura A.4.	Localización de San Juan Quiahije.....	64
Figura A.5.	Localización de Santos Reyes Nopala.....	65
Figura A.6.	Localización de Santiago Yaitepec.....	66
Figura A.7.	Localización de Santa Cruz Zenzontepec.....	67
Figura C.1.	Pantalla PRINCIPAL.....	72
Figura C.2.	Pantalla CORPUS DE TEXTOS.....	73

Figura C.3.	Mensajes de la opción Guardar Lexema de la pantalla CORPUS DE TEXTOS.....	74
Figura C.4.	Mensaje de la opción Guardar Todos de la pantalla CORPUS DE TEXTOS.....	74
Figura C.5.	Mensaje de error de la opción Borrar Lexema de la pantalla CORPUS DE TEXTOS.....	75
Figura C.6.	Menús despegables de la pantalla CORPUS DE TEXTOS.....	75
Figura C.7.	Abrir archivos de textos.....	75
Figura C.8.	Pantalla VER CORPUS.....	76
Figura C.9.	Listas despegables.....	76
Figura C.10.	Mensajes de la opción Guardar de la pantalla VER CORPUS.....	77
Figura C.11.	Pantalla ZONA.....	78
Figura C.12.	Mensajes de la opción Modificar de la pantalla ZONA.....	79
Figura C.13.	Mensajes de la opción Eliminar de la pantalla ZONA.	79
Figura C.14.	Mensajes de la opción Guardar de la pantalla ZONA.	80
Figura C.15.	Pantalla CLASIFICACIÓN.....	80
Figura C.16.	Mensajes de la opción Modificar de la pantalla CLASIFICACIÓN..	81
Figura C.17.	Mensajes de la opción Eliminar de la pantalla CLASIFICACIÓN....	82
Figura C.18.	Mensajes de la opción Guardar de la pantalla CLASIFICACIÓN.....	82
Figura D.1.	Pantalla de bienvenida del conector ODBC.....	85
Figura D.2.	Tipo de instalación del conector ODBC.....	85
Figura D.3.	Resumen del tipo de instalación del conector ODBC.....	86
Figura D.4.	Finalizar la instalación del conector ODBC.....	86
Figura D.5.	Pantalla de bienvenida del servidor MySQL.....	87
Figura D.6.	Tipo de instalación del servidor MySQL.....	87
Figura D.7.	Resumen del tipo de instalación del servidor MySQL.....	88
Figura D.8.	Instalación del servidor MySQL.....	88
Figura D.9.	Primera pantalla de publicidad del servidor MySQL.....	88
Figura D.10.	Segunda pantalla de publicidad del servidor MySQL.....	89
Figura D.11.	Fin de la instalación del servidor MySQL.....	89

Figura D.12.	Pantalla de bienvenida del servidor MySQL.....	91
Figura D.13.	Tipo de configuración del servidor MySQL.....	92
Figura D.14.	Primera pantalla del uso del servidor MySQL.....	92
Figura D.15.	Segunda pantalla del uso del servidor MySQL.....	92
Figura D.16.	Elegir unidad para guardar los archivos del servidor MySQL.....	93
Figura D.17.	Número de conexiones del servidor MySQL.....	93
Figura D.18.	Número de puerto del servidor MySQL.....	93
Figura D.19.	Elegir el juego de caracteres del servidor MySQL.....	94
Figura D.20.	Tipo de arranque del servidor MySQL.....	94
Figura D.21.	Ingresar contraseña para el servidor MySQL.....	94
Figura D.22.	Opción <code>Execute</code> del servidor MySQL.....	95
Figura D.23.	Finalizar la configuración del servidor MySQL.....	95
Figura D.24.	Pantalla de bienvenida de la instalación entorno gráfico de MySQL.....	96
Figura D.25.	Licencia del entorno gráfico de MySQL.....	97
Figura D.26.	Directorio de instalación del entorno gráfico de MySQL.....	97
Figura D.27.	Tipo de instalación del entorno gráfico de MySQL.....	97
Figura D.28.	Resumen del tipo de instalación del entorno gráfico de MySQL.....	98
Figura D.29.	Instalación del entorno gráfico de MySQL.....	98
Figura D.30.	Fin de la instalación del entorno gráfico de MySQL.....	98
Figura D.31.	Ingresar password para entrar a MySQL.....	99
Figura D.32.	Pantalla Principal.....	100
Figura D.33.	Pantalla opción <code>Restore</code>	100
Figura D.34.	Cuadro de diálogo abrir archivo sql.....	100
Figura D.35.	Restauración de la base de datos.	101
Figura E.1.	Contenido del CD.....	102
Figura E.2.	Archivos fuente de la aplicación.....	103
Figura E.3.	Documento de tesis.....	103
Figura E.4.	Archivos de instalación.....	103
Figura E.5.	Instalador MySQL.....	104
Figura E.6.	Respaldo de la base de datos.....	104
Figura E.7.	Textos escritos en chatino.....	105

LISTADO DE TABLAS

Tabla I.	Filtro de símbolos y números.....	18
Tabla II.	Comparación entre voces zapotecas y chatinas.....	21
Tabla III.	Elementos estructurales de los corpus.....	30
Tabla IV.	Diccionario de datos de la base de datos corpus.....	41
Tabla V.	Comparación de los corpus.....	52

GLOSARIO DE TÉRMINOS

4GL

Lenguaje de cuarta generación, esto significa que la mayoría de las tareas que se realizan con él se consiguen sin escribir código; son más fáciles de usar y suelen incluir interfaces gráficas (La revista informática 2006).

APLICACIÓN

Programa de computadora que se utiliza como herramienta para una operación o tarea específica (Definición ABC 2007).

DICCIONARIO DE DATOS

Define con precisión los datos que se manejan en un sistema y contiene las características lógicas de los datos a utilizar: nombre, tipo, tamaño y descripción de los campos. (Silberschatz, Korth & Sudarshan 2006).

FONOLÓGICO

Trata de cómo las palabras se relacionan con los sonidos que representan (Martínez-Mena 2007).

GUI

Graphical User Interface. Interfaz gráfica de usuario, es un medio de comunicación efectiva entre humano y una computadora (Pressman 2005).

LEXEMA

Unidad mínima o raíz de una palabra, con significado léxico (RAE 2010).

LINGÜÍSTICA	Estudio teórico del lenguaje que se ocupa de métodos de investigación y de cuestiones comunes a las diversas lenguas (RAE 2010).
MORFOLÓGICO	Trata de cómo las palabras se construyen a partir de unas unidades de significado más pequeñas llamadas morfemas (Martínez-Mena 2007).
ORIENTADO A EVENTOS (OE)	Son programas que esperan a que el usuario realice alguna acción, ya sea con el mouse o con el teclado para realizar alguna función (Ramírez-Ramírez 2001).
PRAGMÁTICO	Trata de cómo las oraciones se usan en distintas situaciones y de cómo el uso afecta al significado de las oraciones. (Martínez-Mena 2007).
RAD	<i>Rapid Application Development</i> . Desarrollo Rápido de Aplicaciones, permite construir sistemas utilizables en poco tiempo, es decir, permite al equipo de desarrollo crear un sistema completamente funcional dentro de periodos cortos de tiempo (Mena-Mendoza 2005).
SEMÁNTICO	Trata del significado de las palabras y de cómo dichos significados se unen para dar significado a una oración (Martínez-Mena 2007).
SGBD	<i>Data Base Management System</i> . Sistema de Gestión de Bases de Datos, es una colección de datos interrelacionados, su objetivo es proporcionar una forma de almacenar y recuperar la información de una

base de datos de manera que sea tanto práctica como eficiente (Silberschatz, Korth & Sudarshan 2006).

SINTÁCTICO

Trata de cómo las palabras pueden unirse para formar oraciones, fijando el papel estructural que cada palabra juega en la oración (Martínez-Mena 2007).

TONO

Cualidad de los sonidos, dependiente de su frecuencia, que permite ordenarlos de graves a agudos (RAE 2010).

CAPÍTULO 1. INTRODUCCIÓN

El presente trabajo de tesis consiste en la generación de un corpus de texto en lengua chatina en formato digital y la realización de una aplicación que permite, a partir de los archivos escritos en la lengua chatina, generar una base de datos (corpus) con las palabras extraídas de la información escrita recabada en chatino, a las cuáles se les puede ligar una o más traducciones, la zona a la que pertenece y su clasificación sintáctica (por ejemplo, adjetivo, pronombre, sustantivo o verbo).

En la actualidad la cultura chatina está desapareciendo, debido a que existe una decadencia de personas interesadas en interactuar y conocer la cultura de estas comunidades indígenas que aún conservan su idioma chatino y lo que se desea es ofrecer con orgullo la contribución del chatino a la cultura mundial ya que muchas personas en todo el mundo honran y respetan la contribución de las culturas indígenas, como lo demuestran los proyectos encontrados sobre este tema.

Dado que en este trabajo el tema principal es la elaboración del corpus, es necesario empezar definiendo qué es un corpus.

Un corpus es una colección de textos de un idioma ya sean de escritos o transcripciones del habla, elegidos para caracterizar un estado o variedad de un idioma de un dominio específico para un propósito en particular.

Un corpus es valioso porque los textos que lo componen permiten hacer afirmaciones o deducciones acerca de un idioma como un todo. El corpus puede ayudar en el estudio de un idioma o bien puede ayudar a realizar afirmaciones (Educ.ar 2006).

En esta tesis también se detallan las etapas que se llevaron a cabo para el desarrollo de este trabajo: para la recopilación de los textos en chatino se contó con el apoyo de las siguientes personas: el señor Erick, originario de Santa Catarina Juquila; la señorita Yolanda, de la comunidad de Santos Reyes Nopala, el joven Adán, originario de Panixtlahuaca; los integrantes del proyecto chatino (a los que se contactó vía electrónica) y la M. en I. María de Jesús Estudillo Ayala, quien proporcionó documentos impresos. También se descargaron archivos de Internet.

La captura de los textos en chatino se realizó en el software Microsoft® Bloc de notas, sin embargo al momento de la captura se presentaron algunos problemas técnicos, los cuales se detallarán en la sección 4.2.

Para el desarrollo de la aplicación se implementó una base de datos de acuerdo a los requerimientos del corpus de texto y se desarrolló una aplicación en Microsoft Visual Basic 6.0 para tener un mejor acceso a la base de datos y poder consultarla de manera sencilla, dándole un buen uso.

Por último se presenta el corpus generado con todos los textos en chatino recopilados, de las variantes de Tataltepec de Valdés, Santa Catarina Juquila y San Juan Quiahije.

Organización del documento de tesis

El presente documento de tesis está estructurado en cinco capítulos: introducción, antecedentes, marco teórico, desarrollo del tema y por último conclusiones, recomendaciones y trabajos futuros.

A continuación se describen de forma general los cuatro capítulos restantes y los anexos.

Capítulo 2. Antecedentes. En este capítulo se plantean los trabajos y proyectos relacionados a la lengua chatina disponibles hasta el momento, es decir los estudios que se han realizado. Todo esto con el objetivo de tener una idea básica de cómo abordar este trabajo.

De la misma forma se abordan temas como: Justificación, planteamiento del problema, objetivos, alcances y límites del estudio.

Capítulo 3. Marco teórico. Consiste en la teoría que existe acerca del tema “Elaboración de corpus de textos en lengua chatina”, iniciando con una descripción de la lengua chatina.

Capítulo 4. Desarrollo del tema. Se plantea un análisis detallado de la forma en que se llevó a cabo la recopilación de los textos en lengua chatina, la captura de los textos, de igual forma se describe el diseño de la base de datos, se describe la aplicación desarrollada en este documento y un ejemplo donde se ilustran los pasos que se deben realizar para el buen uso de la aplicación, desde el momento en que se abre un archivo de texto, enlistar los lexemas y guardarlos. Por último, se presenta el producto final que es el corpus de textos en chatino.

Capítulo 5. Conclusiones y trabajos futuros. En este último capítulo se mencionan las conclusiones a las que se llegó a lo largo de esta tesis y se mencionan posibles líneas de trabajos futuros que podrán extender los resultados obtenidos.

Anexo A. Chatinos. En este anexo se presenta una introducción de los chatinos, su cultura y origen, así como también una descripción de las zonas en donde se habla esta lengua, para conocer el objeto de estudio del presente trabajo.

Anexo B. Textos capturados. En este apartado se presenta una descripción de los textos recopilados de las variantes de chatino: Tataltepec de Valdés, Santa Catarina Juquila y San Juan Quiahije.

Anexo C. Manual de usuario. En este apartado se explica de forma detallada el funcionamiento de la aplicación.

Anexo D. Instalación y restauración de la base de datos. En este anexo se detalla el proceso de la instalación del manejador de base de datos MySQL 5.0, así como los pasos a seguir para el proceso de la restauración de la base de datos.

Anexo E. Contenido del CD. En este apartado del documento se da una breve descripción de los archivos que se encuentran en el CD-ROM, el código fuente, el documento de la tesis en archivo pdf., el instalador del proyecto, el instalador MySQL, respaldo de la base de datos y los textos en chatino.

Después haber dado una idea general acerca de este proyecto, se procederá a dar a conocer la información necesaria para comprender este documento a partir de los estudios previos que se han realizado.

CAPÍTULO 2. ANTECEDENTES

Actualmente se cuenta con pocos trabajos de investigación que abordan el tema elaboración de corpus de textos en lengua chatina. En este capítulo se plantean los trabajos y proyectos disponibles hasta el momento acerca de este tema de tesis, comenzando con una introducción a la lengua chatina.

Los chatinos son uno de los pueblos indígenas de México que viven al sureste del estado de Oaxaca. Su lengua es el chatino y forma parte del grupo de lenguas zapotecanas de la rama oriental de la familia lingüística otomague (Carrizosa s.f.).

Según Pride y Pride (2007), a principios del siglo XX, el chatino empezó a despertar el interés de lingüistas y antropólogos. Los primeros estudios los llevaron a cabo don Francisco Belmar en 1902, Mechling en 1912 y Boas en 1913.

Estos estudios hacían relación entre el chatino y otras lenguas; dichos investigadores establecían que el chatino estaba remotamente relacionado con el zapoteco y tenía tres variantes principales: la primera es el chatino de la zona alta, que es el que los

hablantes de otras variedades entienden más, la segunda es el chatino de Tataltepec; la tercera es el chatino de Zenzontepec que también se habla en Tlapanalquiahuit Tlacotepec.

Después de la Segunda Guerra Mundial, comenzaron los esfuerzos para que el chatino fuera lengua escrita. Este esfuerzo lo hicieron los misioneros del Instituto Lingüístico de Verano (ILV), que no sólo estudiaron la fonología y sintaxis del chatino, sino que también trataron de documentar sus variantes y el grado de inteligibilidad existente. En sus estudios sobre el chatino, los lingüistas del ILV han utilizado varias ortografías diferentes a través de los años, por ejemplo, una ortografía basada en el inglés que marca los tonos con números.

La misión del ILV en México, es servir a las comunidades lingüísticas del país principalmente a través del estudio científico de las lenguas, la traducción y la capacitación, apoyando el desarrollo de las comunidades indígenas. Su visión es que cada comunidad lingüística de México, valore su propia lengua y tenga la capacidad para lograr el desarrollo sustentable de su lengua en una variedad de ámbitos y medios para beneficio de sus hablantes.

Uno de los objetivos del ILV es apoyar los esfuerzos de las comunidades para elaborar y publicar materiales en los idiomas autóctonos de México. Otro objetivo importante del ILV en México es colaborar para aumentar la capacidad de las comunidades lingüísticas de México para lograr el desarrollo sustentable de sus propias lenguas. El tercer objetivo general del ILV en México es poner los resultados de sus investigaciones al alcance de aquellos que lo necesiten (Pride & Pride 1997).

En los tiempos actuales, las nuevas tecnologías han permitido un avance acelerado de las ciencias en general. A pesar de esto, en el campo de la lingüística su incorporación ha sido lenta, quizás por la misma complejidad de la materia de estudio: la lengua. Sin embargo, en los últimos años la utilización de programas computacionales para el análisis lingüístico en países más desarrollados se ha vuelto algo característico. Por lo anterior, los investigadores concluyeron que la recolección de un corpus digitalizado constituye una tarea necesaria y, además, el primer paso en la implementación de la tecnología computacional al análisis lingüístico en el país (Vergara-Heidke 2006).

2.1. Estado del arte y trabajos relacionados

La Universidad del Mar campus Puerto Escondido no cuenta con un proyecto similar al que se presenta, por lo que esta tesis es el inicio de nuevos proyectos de procesamiento de lenguaje natural e investigación sobre lenguas indígenas.

A continuación se mencionan algunos proyectos que se han realizado sobre la lengua chatina:

La traducción de la Biblia

Contiene todas las citas bíblicas y está escrita en el chatino de Tataltepec por el ILV. Esta traducción se realizó durante la gestión del entonces presidente, General Lázaro Cárdenas del Río, debido a que él creía que el protestantismo podría romper las barreras creadas por el catolicismo para la modernización de las comunidades indígenas (Pride & Pride 1997).

Proyecto de documentación y preservación del idioma Chatino (Proyecto Chatino 2008)

En este proyecto se unen esfuerzos comunitarios y académicos para documentar y apoyar a la preservación de todas las variantes del chatino en Oaxaca, México. Dicha investigación se basa en grabaciones del uso de la lengua por miembros de las comunidades, haciendo uso del audio y video, también se estudian los sonidos, la gramática, las palabras, así como los usos que se le dan a la lengua en las celebraciones religiosas y en la vida cotidiana.

Este proyecto inició en el verano de 2003, con la finalidad de preservar el idioma chatino, ya que la lengua materna de cada persona le permite llevar en su mente, la habilidad de aprender sobre el mundo, y de las cosas que hay en su alrededor, esto mismo le permite a la persona ser creativo en su lengua, por ejemplo, cómo crear poesía y literatura, etc.

Hilaria Cruz, colaboradora del Proyecto de documentación y preservación del idioma Chatino, del departamento de lingüística, de la Universidad de Texas en Austin en una publicación comentó que se ha dicho que los idiomas indígenas son inferiores a las lenguas del mundo, como español o inglés; sin embargo, los idiomas indígenas tienen

amplios vocabularios, poesía y reglas gramaticales muy complejas, como cualquier otra lengua.

También señala que el idioma chatino es un dialecto porque tiene poco tiempo que se empezó a escribir. Aunque esto tampoco es cierto. Se tiene el ejemplo del inglés y el español: Hace algunos siglos cuando la gente empezó a escribir inglés y español, a ellos también les decían que el latín era el único idioma, y que el inglés y el español eran dialectos. No haciendo caso a los comentarios de aquel tiempo, estas personas siguieron adelante con sus proyectos de escribir sus lenguas. De esta manera ellos desarrollaron alfabetos para sus lenguas, escribieron gramáticas y compilaron diccionarios. Con el paso del tiempo los primeros materiales que desarrollaron en estas lenguas se han ido mejorando, ahora existen obras completas y otros materiales de muy buena calidad, lo mismo se puede hacer con las lenguas indígenas de Oaxaca.

Se está estudiando y documentando las siguientes variantes: San Juan Quiahije, Santiago Yaitepec, San Marcos Zacatepec, Santa Lucía Teotepec, Tataltepec de Valdés y Santa Cruz Zenzontepec.

Los recursos digitales que se obtuvieron de este proyecto son: recursos académicos, recursos pedagógicos, gramática y grabaciones de la lengua chatina.

La información que está disponible en la página Proyecto Chatino (2008) es un material lingüístico y pedagógico de la lengua chatina que contiene:

- Recursos académicos, por ejemplo, comparativas y generales de las zonas donde se habla la lengua chatina, estos archivos se pueden descargar en formato pdf, en la página oficial de Proyecto Chatino (2008) y están escritos en español o inglés.
- Recursos pedagógicos comparativos y generales, ejemplo: consonantes en algunas variedades de chatino y una comparación de los tonos chatinos, entre otros, estos recursos pedagógicos son disponibles para las siguientes variantes del chatino: El chatino de Santa Cruz Zenzontepec, el chatino de Tataltepec de Valdés y el chatino oriental, dentro del chatino oriental incluyen al distrito de Juquila. Estos archivos se pueden descargar en archivo de tipo pdf y están escritos en español.

La información no disponible de forma libre de este proyecto son: gramática de la lengua chatina, grabaciones, guías lingüísticas, asesoría gramatical y escritura, asesoría de

revitalización de lengua, asesoría técnica sobre documentación de lengua, asesoría de inmersión de lengua, abogacía de lengua, referencias bibliográficas y asesoría sobre carrera de estudios, esta información es proporcionada contactando a los colaboradores del proyecto, por medio del correo electrónico o redes sociales.

El equipo de trabajo de este proyecto está integrado por las siguientes personas:

Anthony Woodbury, profesor de lingüística en la Universidad de Texas en Austin (UT) y director del proyecto chatino. Es originario de Nueva York, EE.UU, obtuvo su doctorado en lingüística en la Universidad de California, Berkeley. Ha trabajado desde 1978 con lenguas indígenas de Alaska, EE.UU, y es experto en la documentación de lenguas de escaso uso. En el año 2003 empezó a trabajar con Emiliana Cruz con el chatino de Quiahije.

Emiliana Cruz, hablante de la lengua chatina y originaria de Cieneguilla, San Juan Quiahije. Obtuvo su licenciatura en la Universidad de Evergreen, estado de Washington, en el año 2002. En el 2003 empezó sus estudios de doctorado. Emiliana Cruz y Anthony Woodbury fueron los fundadores del proyecto chatino. Emiliana se especializa en la historia de educación indígena, lexicografía y gramática del chatino.

Hilaria Cruz, hablante de la lengua chatina y originaria de Cieneguilla, San Juan Quiahije. Obtuvo su licenciatura en la Universidad de Evergreen, estado de Washington, en 1998.

En el 2004 empezó sus estudios de doctorado. Hilaria se especializa en el discurso político chatino, arte verbal, y gramática del chatino.

Eric Campbell, originario de Michigan, EE.UU, obtuvo su licenciatura en la Universidad de Michigan en 2005. En el año 2006 empezó sus estudios de doctorado. Actualmente se encuentra elaborando el primer diccionario de la variante de Zenzontepec. Se especializa en el desarrollo histórico de las lenguas chatinas y zapotecanas.

En el momento, está trabajando en la transcripción, traducción y análisis gramatical de grabaciones de la lengua.

Justin McIntosh, originario de Portland, Oregon, EE.UU, obtuvo su licenciatura en la Universidad de Evergreen, Estado de Washington. Empezó sus estudios de doctorado en el año 2006 y está trabajando en la comunidad de Santa Lucía Teotepec. Justin McIntosh se especializa en gramática del chatino, el discurso, la educación, y la economía política de la región chatina.

Jeff Rasch, originario de Texas, EE.UU, obtuvo su doctorado en lingüística en la universidad de Rice, Houston, Texas con una tesis “The basic morpho-syntax of Yaitepec Chatino.” Jeff Rasch se especializa en la gramática chatina y está trabajando con el chatino de Santiago Yaitepec.

John Ryan Sullivant, originario de Louisiana, EE.UU, obtuvo su licenciatura en la Universidad de Tulane, Nueva Orleans, Estado de Louisiana. En el año 2007 empezó sus estudios de doctorado y se especializará en el chatino de Tataltepec de Valdés.

Stéphanie Villard, originaria de Francia, en el año 2004 obtuvo su licenciatura en la UT. Empezó sus estudios de doctorado en el año 2005 y se especializa en el chatino de San Marcos Zacatepec.

Proyecto Chatino de la Universidad de Texas, Austin (Sherzer s.f.)

Este proyecto se realizó en abril de 2008, es un escrito que contiene una lista de conjunción de verbos con diferentes pronombres y tiempos y un listado de palabras en orden alfabético (números, animales, colores, algunos sustantivos y verbos). La parte central de la colección está construida por grabaciones de discurso natural en un amplio rango de géneros, incluyendo narrativas, ceremonias, oratorias, conversaciones, entrevistas, gramáticas, lecciones gramaticales, listas de palabras y canciones. Muchas de estas grabaciones vienen con transcripciones y traducciones en español, inglés, o portugués. Estos trabajos contienen una gran riqueza de información sobre las culturas indígenas latinoamericanas, así como también conocimientos sobre el medio ambiente natural en el cual viven sus hablantes.

Esta información está publicada en archivo digital de grabaciones y textos en, y sobre los idiomas indígenas de Latinoamérica. El acceso a los recursos del archivo es gratis.

La mayor parte de los recursos en la base de datos está disponible al público, pero algunos tienen restricciones especiales de acceso.

AILLA (Archivo de los Idiomas Indígenas de Latinoamérica) maneja cuatro niveles de acceso que van desde el libre acceso público a un acceso más restringido, los cuatro niveles de acceso son: acceso libre al público, acceso mediante una contraseña, el depositante controla el acceso al recurso y por último una persona indígena controla el acceso al recurso.

La información que no está disponible de forma libre en este proyecto son: conversaciones, entrevistas, juego de datos, narrativas, oratorias, lecciones gramaticales, listas de palabras y canciones. Esta información es proporcionada, solicitando una cuenta de usuario, la cual se puede crear desde la página principal de AILLA.

AILLA acepta todo tipo de materiales en, y sobre los idiomas indígenas de Latinoamérica ya sea literatura (escritos, grabaciones, videos, poesía, ficción, ensayos), materiales educativos (libros de texto, impresos, libros de trabajo) y trabajos académicos (grabaciones de todo tipo, transcripciones, traducciones, gramática, diccionarios).

El proyecto AILLA fue fundado en el año 2000, por los profesores, Joel Sherzer, Anthony Woodbury y Mark McFarland, del Departamento de Antropología, Departamento de Lingüística y Director de la División de Servicios Digitales Bibliotecarios de las Bibliotecas Generales de la UT, respectivamente. Desde hace muchos años, el profesor Joel Sherzer, ha realizado investigaciones sobre el idioma Kuna de Panamá y ha hecho una recolección de tamaño considerable de grabaciones de discursos Kuna que incluye ejecuciones de gran importancia para este pueblo indígena panameño.

A través de los años también se han coleccionado variedad de grabaciones y textos por investigadores en los Estados Unidos, Latinoamérica y Europa, Joel Sherzer se ha preocupado por la preservación de estos materiales de arte oral indígena, si estos materiales fueran recolectados, constituirían un recurso valioso para la documentación y el análisis de los idiomas indígenas de Latinoamérica y creación de materiales educativos para proyectos de recuperación de idiomas.

Con el desarrollo acelerado de tecnología a un precio razonable, AILLA ha aprovechado los medios digitales para la colección, digitalización y preservación de dichas colecciones en formatos estándares. Estos formatos pueden ser accesados por cualquier

persona que use herramientas de software ampliamente disponibles. Estos recursos se ofrecen a un público global, por medio de Internet, pero lo más importante es que se incluye a los hablantes de los idiomas que se encuentran en el archivo.

Actualmente, AILLA maneja financiamiento otorgado por el National Endowment for the Humanities y el National Science Foundation.

Conclusión sobre el estado del arte

Después de leer los proyectos se concluyó que en la actualidad existen trabajos relacionados al chatino principalmente en su forma escrita, y en su forma hablada se están creando repositorios de audio y video y no existe como tal un corpus en formato electrónico de la lengua chatina; sólo existen algunos escritos en dicha lengua ya que es muy difícil de conseguirlos.

El Proyecto Chatino se tomó como base para las clasificaciones sintácticas que se utilizan en esta tesis (adjetivo, pronombre, sustantivo y verbo).

En este proyecto se está tratando de conjuntar las zonas ya que normalmente se encuentran en diferentes fuentes.

2.2. Justificación

Este trabajo ayudará al estudio de la lengua chatina para así preservar su cultura, ya que se está perdiendo gran parte de la misma. Esto se debe al desconocimiento y mal manejo de los términos de esta lengua, debido a que la lengua chatina es muy difícil de estudiar y muchas de las personas que la hablan, desconocen su escritura, además, este trabajo les permitirá a los pueblos indígenas contar con una base documental que puedan consultar.

Con la realización de este trabajo se pretende poner la base para favorecer el aprendizaje o enriquecer los conocimientos sobre la cultura, escritura y documentación del chatino, para que posteriormente otros estudios contribuyan a comprender y entender su forma de pensar y vivir de las culturas que manejan esta lengua.

Esta investigación es base para las tareas del procesamiento de lenguaje natural (PLN), por ejemplo: la traducción automática, la respuesta a preguntas, la recuperación y

extracción de la información, reconocimiento de voz, reconocimiento de escritura, síntesis de voz y minería de textos.

Para el PLN, primeramente se necesita el corpus del lenguaje, el cual es un conjunto de textos almacenados en formato electrónico, agrupados con el fin de estudiar una lengua o una determinada variedad lingüística. Su objetivo es constituirse en elementos de referencia para el estudio de una frase concreta o un cierto aspecto de una lengua (Montero-Martínez 2008).

Actualmente no existe un corpus de la lengua chatina en formato digital, sin embargo es importante la realización de un corpus ya que sirve para tener un concentrado de palabras en chatino, que se pueden utilizar en diferentes áreas de estudio.

Con la realización de este trabajo se podrá extraer información que servirá de ayuda a la investigación lingüística, ya sea a nivel fonológico, morfológico, sintáctico, semántico o pragmático.

Para facilitar lo anterior, se considera necesario diseñar e implementar una aplicación con una interfaz amigable y sencilla que permite ingresar palabras en chatino y almacenarlas en una base de datos para posteriormente realizar consultas a ésta, en forma automática, que sea independiente del dominio para las regiones donde se habla el chatino y que en un futuro sea la base para poder implementar otros dialectos además del chatino, por ejemplo de los grupos etnolingüísticos que existen en Oaxaca: amuzgos, chinanteco, chocho, chontal, cuicateco, huave, ixcatéco, mazateco, mixe, mixteco, náhuatl, triqui, zapoteco, zoque y el popoloca, este último en peligro de extinción (Berumen-Barbosa 2003). (Fig. 2.1).

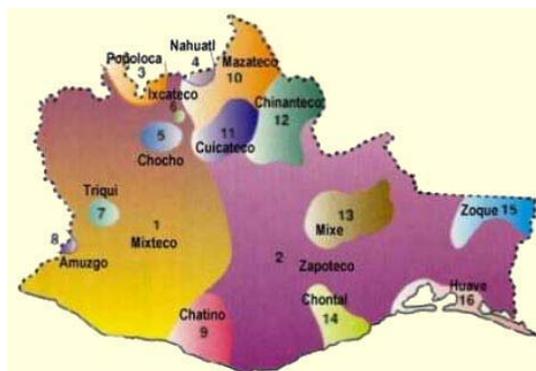


Figura 2.1. Mapa de localización de las etnias en Oaxaca.

2.3. Planteamiento del problema

Con el paso del tiempo el monolingüismo en chatino ha ido disminuyendo, en mayor parte debido a los medios masivos de comunicación como la televisión y la radio, la construcción de mejores vías de comunicación, además de que se fundaron escuelas y los profesores que imparten clases hablan español, razón por la cual los alumnos de los pueblos chatinos aprenden el español en la escuela, además de su lengua indígena.

El chatino empezó a atraer el interés de lingüistas y antropólogos, se realizaron estudios que se referían a las relaciones entre el chatino y otras lenguas. Los pioneros en llevar a cabo estos estudios fueron: Belmar (1902), Boas (1913) y Mechling (1912), pero a estos investigadores no les preocupaba que el chatino fuera lengua con escritura. Por lo que sólo se enfocaron hacia estudios fonológicos de la lengua en chatino, y no de su escritura (Pride & Pride 1997).

Es importante preservar y difundir un idioma con raíces históricas, para que ésta no disminuya, ya que si desaparece se perdería la esencia e identidad de los pueblos nativos mexicanos. Cabe recalcar que la presencia de los grupos étnicos, le permite al estado de Oaxaca contar con una riqueza cultural, costumbres y tradiciones vastas, conservadas hasta la fecha, sin embargo casi en su totalidad, persiste el rezago, la marginación y la pobreza en sus 4,031 localidades de más de 30% de hablantes de lengua indígena, ubicadas en 379 municipios, con una población total de 1'339,742 personas (Berumen-Barbosa 2003).

En la actualidad esta cultura está desapareciendo, debido a que las comunidades indígenas que hablan la lengua chatina son pocas (Fig. 2.2). Además, existe una decadencia de personas interesadas en interactuar y conocer la cultura de estas comunidades indígenas que aún conservan su idioma chatino. Otras de las razones es que existen otras lenguas que están acaparando a estas comunidades.

El lenguaje chatino es muy difícil de estudiar ya que las zonas donde se habla son de difícil acceso, debido a que se encuentran en comunidades lejanas, de ahí la dificultad para recabar esa información.

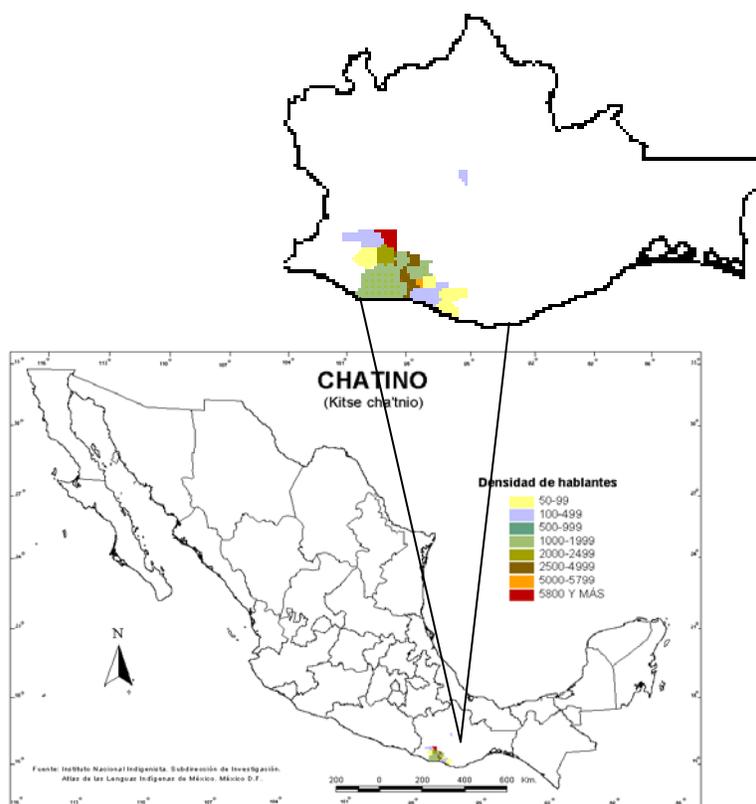


Figura 2.2. Densidad de hablantes de la lengua chatina (CIESAS 2008).

Existen documentos escritos en esta lengua, de los cuales podemos mencionar: la traducción de la Biblia, obras lingüísticas de H. McKaughan y B. McKaughan (1951); K. Pride (1961, 1965); L. Pride (1963); K. Pride y L. Pride (1970); J. Upson y R. Longacre (1965). Muchos de estos escritos no están en formato electrónico, y si se encuentran en formato electrónico, no están preparados para ser procesados y poder hacer investigaciones y herramientas de PLN con dicha lengua (Pride & Pride 1997).

Se requiere preparar los documentos para hacer diferentes investigaciones tanto en el ámbito lingüístico como en el cultural, socioeconómico e informático también se requieren hacer diferentes herramientas y tecnologías que incorporen al lenguaje chatino.

A partir de la información recabada y contenida en los documentos, se realizó una recopilación de textos categorizados (poesía, cuentos, leyendas y diálogos, etc.) escritos en chatino, y la aplicación para generar un corpus digital con términos de la lengua chatino. Se generó una base de datos con las palabras extraídas de la información recabada escrita en chatino, a las cuáles se les liga una o más traducciones en español, la zona a la que

pertenece y su clasificación sintáctica (sustantivo, verbo, etc.). Con esto se cuenta con una base para que el lenguaje chatino sea estudiado de una forma informática, y sirva como plataforma para herramientas y tecnologías nuevas.

Para mostrar los resultados o para tener acceso a la base de datos del corpus se diseñó una aplicación con una interfaz amigable y sencilla, para que cualquier persona con conocimientos mínimos de computación pueda hacer uso de la misma.

Metodología de solución

Con la finalidad de cumplir los objetivos propuestos y garantizar su cumplimiento, la elaboración del corpus siguió la siguiente metodología:

- Se entrevistó a personas de las comunidades de: Santa Catarina Juquila, Santos Reyes Nopala y Tataltepec de Valdés, que hablan la lengua chatina, haciendo un análisis posterior de cuáles son los hogares donde hay facilidad para encontrar documentos escritos en chatino o personas que lo sepan escribir.
- Una vez ubicados los hogares de habla chatina, se hizo la recopilación de los textos, tratando de obtener adicionalmente textos de internet, documentos capturados y libros en bibliotecas, entrevistando a personas que hablan esta lengua para verificar y corroborar los escritos existentes.
- Se capturó de forma manual dichos documentos en un editor de texto, para posteriormente generar automáticamente un corpus de texto.
- Se diseñó la base de datos de acuerdo a los requerimientos del corpus de texto, en este paso se identificaron los atributos y entidades para realizar el modelo entidad - relación.
- Se desarrolló la aplicación en el lenguaje de programación Microsoft Visual Basic 6.0, ya que es un lenguaje fácil y amigable.
- Se realizaron pruebas y correcciones de errores.
- Se generó el corpus, introduciendo las traducciones de los términos, zona geográfica y clasificación de los términos.
- Se elaboró un manual de usuario.

2.4. Objetivos

Objetivo general

Construir un corpus digital y desarrollar una interfaz para consultar las unidades léxicas, definidas a partir de documentos en chatino.

Objetivos específicos

1. Recopilar textos escritos en la lengua chatina de los municipios de San Juan Quiahije, Santa Catarina Juquila y Tataltepec de Valdés, Oaxaca, México.
2. Digitalizar dichos textos en archivos electrónicos.
3. Desarrollar una aplicación que facilite el procesamiento de las palabras de los textos en chatino.
4. Construir el corpus, el cual deberá tener la siguiente estructura: la palabra extraída de la información recabada escrita en chatino, la traducción, la zona a la que pertenece y su clasificación sintáctica (sustantivo, verbo, etc.).

2.5. Alcances y límites del estudio

Alcances

El objetivo principal del desarrollo de esta tesis consiste en la generación de un corpus digital y desarrollar una aplicación para consultar las unidades léxicas, definidas a partir de documentos en chatino.

- La aplicación cuenta con cuatro módulos: “Corpus de Textos”, “Ver Corpus”, “Zona” y “Clasificación”.
- En el módulo “Corpus de Textos”, se abren archivos con extensión txt.
- Al abrir un archivo que contenga alguno de estos caracteres que se muestran en la tabla I, automáticamente los reemplaza por un espacio en blanco, excepto el apóstrofe (', `) y el guión (-).
- El módulo “Corpus de Textos”, contiene la opción de separar lexemas, visualizando un listado de los lexemas del archivo completo, sin mostrar repeticiones.
- No guarda los lexemas que ya se encuentran almacenados en la base de datos.

- Muestra en pantalla, el número de los lexemas guardados y los repetidos.

Tabla I. Filtro de símbolos y números.

!	"	#	\$	%	&
()	*	+	,	-
.	/	:	;	<	=
>	?	@	[\]
^	_	{		}	~
»	Ç	æ	Æ	ç	£
¥	f	ª	ı	ı	«
1	2	3	4	5	6
7	8	9	0		

- En el módulo “Ver Corpus”, se puede escribir la traducción a los lexemas, con la posibilidad de enlazarles más de una traducción.
- Muestra todos los lexemas en orden alfabético, que ya cuentan con su traducción.
- En el módulo “Zona”, se puede modificar alguna zona, agregar nuevas zonas y eliminar alguna zona, siempre y cuando ésta no esté siendo utilizada.
- En el módulo “Clasificación”, se puede modificar alguna clasificación sintáctica, agregar nuevas clasificaciones y eliminar alguna clasificación, siempre y cuando ésta no esté siendo utilizada.
- El corpus contiene 526 lexemas diferentes, con su traducción y clasificación.

Límites

- Los textos que serán procesados en la aplicación deberán cumplir con los siguientes parámetros:
 - Los textos deberán ser archivos de texto con extensión txt.
 - Si el lexema cuenta con dos palabras, al momento de capturarlo se debe separar con un guión (-).
- El módulo “Ver Corpus” no cuenta con la opción Modificar, en caso de que se quiera cambiar algún dato.

- El módulo “Ver Corpus” no cuenta con la opción Eliminar, en caso de que se quiera eliminar algún lexema.
- El corpus no cuenta con ejemplos del uso de cada lexema.
- El corpus no cuenta con videos.
- El corpus no cuenta con audio.
- El corpus no cuenta con tonos de pronunciación.
- De los 3458 lexemas, sólo se alcanzó a capturar 526 traducciones, recalando que no se cuenta con la traducción de todos los textos recopilados.

Una vez establecido los objetivos, alcances y límites de la elaboración de corpus de textos en lengua chatina, se procederá a detallar los conceptos teóricos acerca de este documento.

CAPÍTULO 3. MARCO TEÓRICO

En este capítulo se presentan las bases teóricas necesarias para el buen funcionamiento de los conceptos relacionados en la elaboración de este documento de tesis.

3.1. Lengua chatina

Los chatinos son uno de los pueblos indígenas de México que viven al sureste del estado de Oaxaca. Su lengua es el chatino y forma parte del grupo de lenguas zapotecanas de la rama oriental de la familia lingüística otomague (Carrizosa s.f.).

El Chatino pertenece a la rama de las lenguas zapotecas del grupo lingüístico otomangue. Se separó del Zapoteco de forma definitiva hace 24 siglos.

El origen de esta lengua es muy controvertido ya que para algunos historiadores el chatino es un dialecto del zapoteco, mientras que para otros su origen es diferente. El historiador José María Bradomín niega completamente que el chatino sea un lenguaje

derivado del zapoteca y dice que ni morfológica ni fonéticamente ambas lenguas tengan afinidad alguna, además de que el arribo de cada uno de estos pueblos se realizó por rumbos opuestos y tampoco sostuvieron relaciones de ninguna especie; así mismo, presenta una comparación de palabras zapotecas y chatinas en donde efectivamente no existe parecido entre ellas salvo en dos palabras.

El idioma chatino fue ampliamente estudiado por don Francisco Belmar (1902), distinguido lingüista oaxaqueño que con entusiasmo y dedicación se propuso averiguar el secreto del origen chatino (Pérez-Sánchez 2004).

Así demostró de manera evidente que el chatino tiene el mismo origen que el zapoteca, es decir, que pertenece a la familia lingüista del mixteco y zapoteco, en la tabla II se muestra esta breve comparación entre voces zapotecas y chatinas.

Tabla II. Comparación entre voces zapotecas y chatinas.

Zapoteco	Castellano	Chatino
<i>Rguiu</i>	Hombre	<i>Quihiu</i>
<i>Guná</i>	Mujer	<i>Guná</i>
<i>Shiñé</i>	Mi hijo	<i>Siñé</i>
<i>Iqué</i>	Cabeza	<i>Iquié</i>
<i>Quicha</i>	Pelo	<i>Quitza</i>
<i>Loocuá</i>	Frente	<i>Ticaá</i>
<i>Diagá</i>	Oreja	<i>Ñagaá</i>
<i>Shi</i>	Nariz	<i>Shiñ</i>
<i>Ruhúa</i>	Boca	<i>Tuhúa</i>
<i>Ludyé</i>	Lengua	<i>Luczé</i>
<i>Luxú</i>	Barba	<i>Luzu</i>
<i>Náa</i>	Mano	<i>Yáa</i>
<i>Guichi</i>	Pueblo	<i>Quichi</i>
<i>Rene</i>	Sangre	<i>Tene</i>
<i>Gobicha</i>	Sol	<i>Cuichá</i>
<i>Loo</i>	Cara	<i>Tuloó</i>

Como se advierte la semejanza es muy clara, aún más si se toma en cuenta de que no existen reglas para uniformar la escritura de nuestros idiomas nativos y cada quien escribe según le parece.

Algunos otros lingüistas también comparten la idea que el chatino y el zapoteco no son lenguas diferentes sino que el primero proviene del segundo y que se separaron o diferenciaron hacia el año 400 a.C.

La voz chatino viene de dos dicciones: *cha*, que significa idioma o palabra y *kiña*: que quiere decir trabajo; puede entonces traducirse como “idioma difícil o trabajoso”, lo que hace suponer que la designación de ese nombre fue hecha por una raza distinta a la que lo habla.

En cuanto al origen del nombre los chatinos se llamaron a sí mismos *Kitse cha'tnio*. Los términos *cha'tnio*, *ch'tña* y *tasa'jnya* son variantes usadas en Nopala y Yaitepec, en Tataltepec y en Zezontepec, respectivamente, para designar la propia lengua y significa “palabra trabajosa”, “palabra difícil”, o bien “trabajo de las palabras”.

En la actualidad se distinguen cuatro variantes del idioma: la de Santiago Yaitepec, que se habla en siete municipios, la de Zezontepec, norte de la región Chatina y la de Tataltepec de Valdés y Nopala, ambas difundidas en la zona de la región de la costa, para más detalle consultar el anexo A.

Para ubicar las zonas en donde se habla chatino, se presenta el mapa de Oaxaca con sus diferentes distritos y municipios que se mencionarán a lo largo de este trabajo.

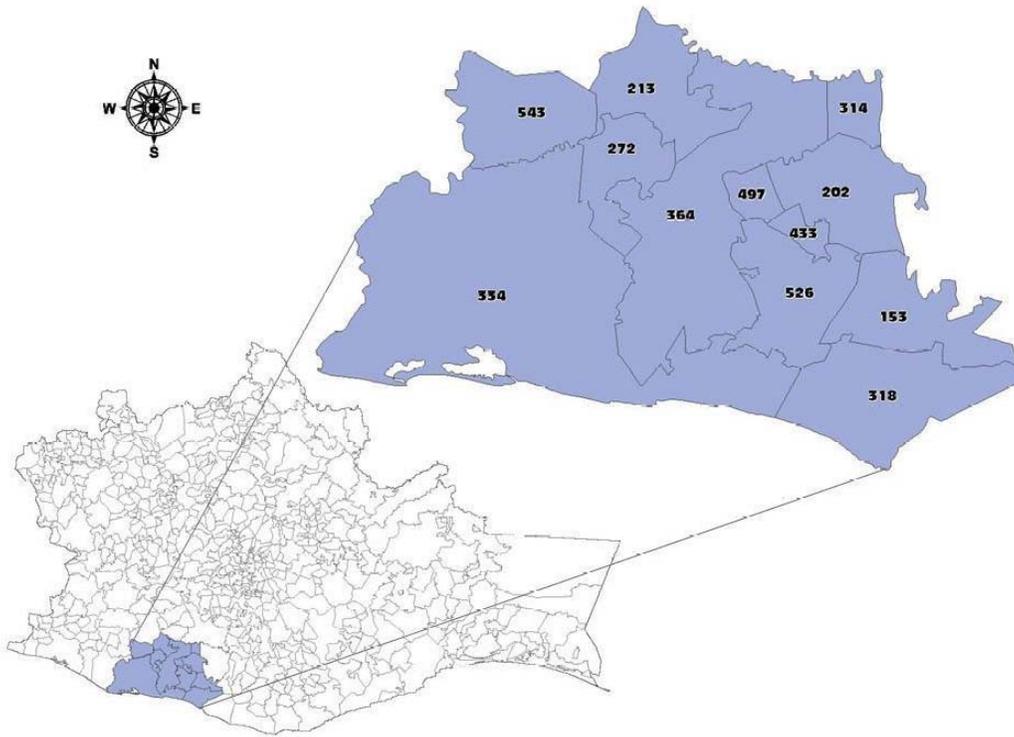
La figura 3.1 muestra el distrito de Juquila, donde se encuentran entre otros los municipios de San Juan Quiahije, Santa Catarina Juquila, Santa María Temaxcaltepec, Santiago Yaitepec, Santos Reyes Nopala y Tataltepec de Valdés.

La figura 3.2 muestra el distrito de Sola de Vega, donde se encuentran entre otros el municipio de Santa Cruz Zenzontepec.

La figura 3.3 muestra el distrito de Jamiltepec.

En el anexo A se describen los municipios donde se habla este idioma, donde se incluye: significado de su nombre, localización, municipios con los que colinda y su principal flora y fauna.

Distrito: Juquila

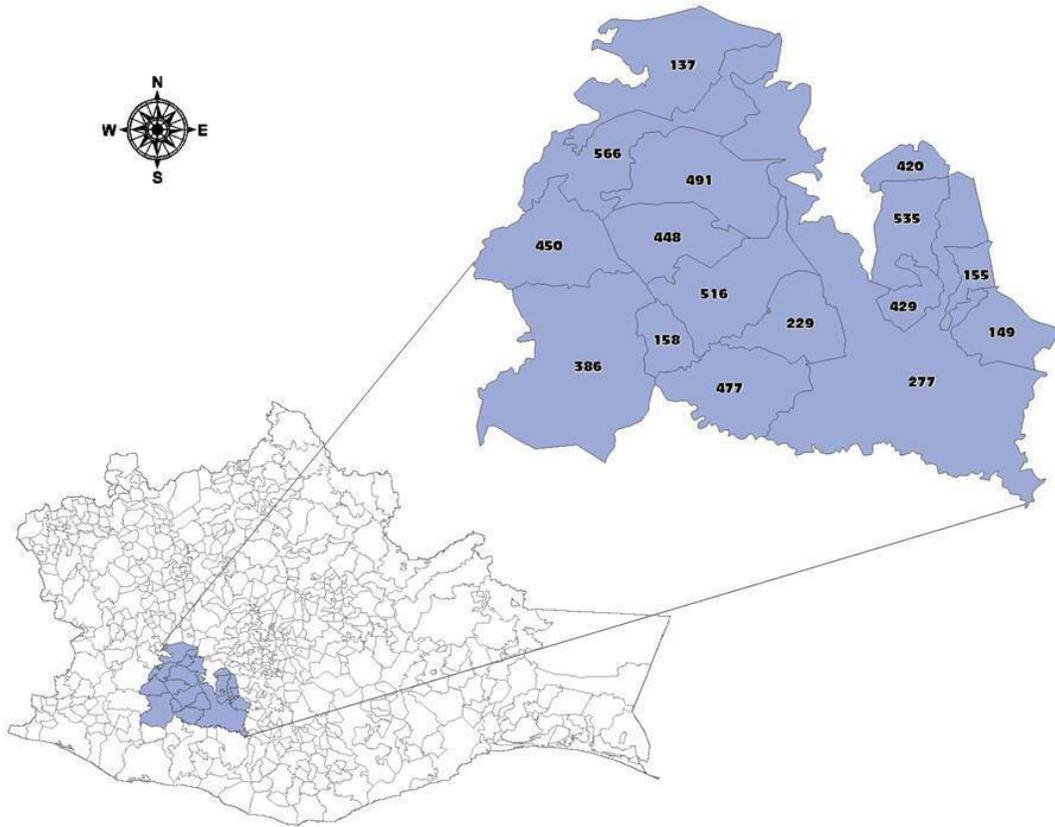


Municipios

153	San Gabriel Mixtepec	334	Villa de Tututepec de Melchor Ocampo
202	San Juan Lachao	364	Santa Catarina Juquila
213	San Juan Quiahije	433	Santa María Temaxcaltepec
272	San Miguel Panixtlahuaca	497	Santiago Yaitepec
314	San Pedro Juchatengo	526	Santos Reyes Nopala
318	San Pedro Mixtepec	543	Tataltepec de Valdés

Figura 3.1. Distrito de Juquila y sus municipios.

Distrito: Sola de Vega

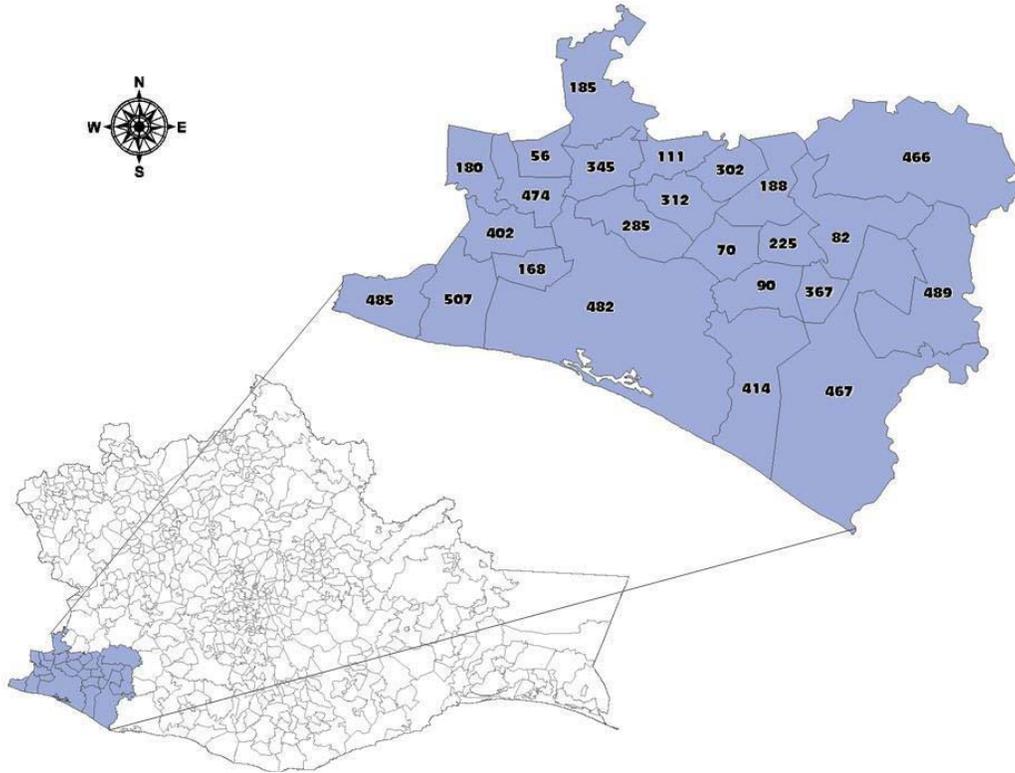


Municipios

137	San Francisco Cahuacua	429	Santa María Sola
149	San Francisco Sola	448	Santa María Zaniza
155	San Ildefonso Sola	450	Santiago Amoltepec
158	San Jacinto Tlacotepec	477	Santiago Minas
229	San Lorenzo Texmelucan	491	Santiago Textitlan
277	Villa Sola de Vega	516	Santo Domingo Teojomulco
386	Santa Cruz Zenzontepec	535	San Vicente Lachixio
420	Santa María Lachixio	566	Zapotitlán del Río

Figura 3.2. Distrito de Sola de Vega y sus municipios.

Distrito: Jamiltepec



Municipios

056	Mártires de Tacubaya	312	San Pedro Jicayan
070	Pinotepa de Don Luis	345	San Sebastián Ixcapac
082	San Agustín Chayuco	367	Santa Catarina Mechoacán
090	San Andrés Huaxpaltepec	402	Santa María Cortijo
111	San Antonio Tepetlapa	414	Santa María Huazolotitlan
168	San José Estancia Grande	466	Santiago Ixtayutla
180	San Juan Bautista lo de Soto	467	Santiago Jamiltepec
185	San Juan Cacahuatepec	474	Santiago Llano Grande
188	San Juan Colorado	482	Santiago Pinotepa Nacional
225	San Lorenzo	485	Santiago Tepextla
285	San Miguel Tlacamama	489	Santiago Tetepec
302	San Pedro Atoyac	507	Santo Domingo Armenta

Figura 3.3. Distrito de Jamiltepec y sus municipios.

Es necesario hacer mención que el chatino, al igual que todos los idiomas humanos, es clasificado como un lenguaje natural. A diferencia de los lenguajes creados para un fin específico, como los lenguajes de programación, los lenguajes naturales no tienen un conjunto finito de reglas sintácticas y semánticas inviolables. Por tanto, no se puede hacer el procesamiento de los lenguajes naturales de forma manual, por las variantes, la sintaxis, las ambigüedades, etc. Ésta sería una de las desventajas de estudiar un lenguaje manualmente, por lo que es necesaria una herramienta o sistema informático que automatice este proceso y que facilite la comunicación hombre-computadora por medio del lenguaje humano, ya sea la voz o el texto. Para ello se necesita estudiar una de las áreas de la inteligencia artificial que es el procesamiento del lenguaje natural, ya que es la disciplina encargada de producir estos sistemas informáticos que posibilitan dicha comunicación.

3.2. Procesamiento del Lenguaje Natural

Para comprender mejor que es el Procesamiento del Lenguaje Natural (PLN), a continuación se da la definición de lenguaje y lenguaje natural. Un lenguaje es la función que expresa pensamientos y permite la comunicación entre los humanos. Esta función es llevada a cabo por medio de: gestos, señales, voz y texto y el lenguaje natural, es la herramienta que utilizan las personas para expresarse (Vallez & Pedraza-Jiménez 2007).

Muñoz-Fernández (2007) define al PLN como el encargado de producir sistemas informáticos que posibiliten la comunicación hombre-hombre u hombre-máquina mediante el lenguaje natural; busca poder crear programas eficaces que puedan analizar, entender y generar lenguajes que los humanos utilizan habitualmente, de manera que el usuario pueda llegar a comunicarse con la computadora de la misma forma que lo haría con un humano.

Vallez y Pedraza (2007) definen al PLN como el reconocimiento y utilización de la información expresada en lenguaje humano (lenguaje natural) a través del uso de sistemas informáticos.

Una vez definidos los conceptos anteriores se puede afirmar que el procesamiento del lenguaje natural (PLN) es una subdisciplina de la Inteligencia Artificial y la Lingüística, cuyo objeto de estudio son los problemas derivados de la generación y comprensión automática del lenguaje natural (Vallez & Pedraza-Jiménez 2007).

El procesamiento del lenguaje natural se estructura normalmente en los siguientes cuatro niveles de análisis (Sosa 1997):

Análisis morfológico. Su función consiste en detectar la relación que se establece entre las unidades mínimas que forman una palabra, como puede ser el reconocimiento de sufijos o prefijos.

Este nivel de análisis mantiene una estrecha relación con el léxico. El léxico es el conjunto de información sobre cada palabra que el sistema utiliza para el procesamiento. Las palabras que forman parte del diccionario están representadas por una entrada léxica, y en caso de que ésta tenga más de un significado o diferentes categorías gramaticales, tendrá asignada diferentes entradas.

En el léxico se incluye la información morfológica, la categoría gramatical, irregularidades sintácticas y representación del significado. Normalmente el léxico sólo contiene la raíz de las palabras con formas regulares, siendo el analizador morfológico el que se encarga de determinar si el género, número o flexión que componen el resto de la palabra son adecuados.

Análisis sintáctico. Este análisis tiene como función etiquetar cada uno de los componentes sintácticos que aparecen en la oración y analizar cómo las palabras se combinan para formar construcciones gramaticalmente correctas.

Análisis semántico. Consiste en extraer la información semántica, es decir, el significado, de la frase. En esta fase se deben resolver ambigüedades léxicas y estructurales.

Análisis pragmático. Su finalidad es incorporar al análisis semántico la aportación significativa que pueden hacer los participantes, la evolución del discurso o información presupuesta, es decir, de cómo las oraciones se usan en distintas situaciones y de cómo el uso afecta al significado de las oraciones.

Para hacer la automatización, primero se debe entender cómo funciona el lenguaje natural, por tanto se inicia este estudio con la lingüística.

Lingüística

Es una disciplina cuyo objeto de estudio es el lenguaje natural, y se especializa en la construcción de una teoría general de la estructura de la lengua y su descripción. Se trata de una ciencia teórica, dado que formula explicaciones diseñadas para justificar los fenómenos

del lenguaje, esto es, el diseño de teorías sobre algunos aspectos del lenguaje y una teoría general del mismo.

Su función no es ni la de evitar el desperfecto de la lengua ni mucho menos procurar una mejoría (Idoneos 1999).

Análisis gramatical (Delgado-Orta 2005)

La gramática es la rama de la lingüística que tiene por objeto el estudio de la forma y composición de las palabras (morfofonología), así como de su interrelación dentro de la oración o de la frase (sintaxis). El estudio de la gramática muestra el funcionamiento de las palabras en una lengua.

El análisis gramatical está compuesto básicamente por tres partes: análisis léxico, análisis sintáctico y análisis semántico.

Análisis léxico. El análisis léxico es un análisis a nivel de caracteres, su misión es a partir de éstos y por medio de patrones reconocer los lexemas enviando al analizador sintáctico el componente léxico. El analizador léxico lee los caracteres de la oración de entrada y genera una cadena de lexemas.

Las funciones principales del análisis léxico son: leer los caracteres de entrada, generar una cadena de componentes léxicos (tokens) y eliminar comentarios y delimitadores como espacios en blanco, signos de puntuación y fines de línea.

Análisis sintáctico. La función principal del análisis sintáctico es analizar sintácticamente una cadena de tokens y encontrar para ella una representación de su estructura mediante un árbol sintáctico o de derivación que tiene como raíz el axioma de la gramática, y como nodos terminales la sucesión ordenada de símbolos que componen la cadena analizada.

Esta estructura se genera partiendo del análisis de los lexemas de acuerdo a la función que tiene dentro de la oración y de acuerdo al orden de aparición dentro de la cadena según las reglas gramaticales del lenguaje.

Análisis semántico. A partir de la estructura generada por el analizador sintáctico, se encarga de extraer el significado o sentido de la oración y genera una estructura lógica.

Una de las tareas clave de la interpretación semántica consiste en considerar qué combinaciones de significados de palabras individuales son posibles a la hora de crear un

significado coherente de la oración, lo que puede reducir el número de posibles significados para cada palabra de una oración determinada.

Antes de realizar cualquier análisis gramatical, se necesita el listado de lexemas o términos definidos para el lenguaje objeto de estudio. Por ello se requiere la construcción de un corpus.

Corpus

Un corpus lingüístico es un conjunto de textos o fragmentos de textos recopilados para realizar análisis lingüísticos.

Antes de la década de 1950 ya existía un amplio uso de corpus. Los objetivos entonces eran muy variados: adquisición del lenguaje, convenciones ortográficas, gramática comparada, sintaxis y semántica (Encarta 2009).

El inglés fue la primera lengua en ser recopilada en grandes corpus. A continuación se mencionan algunos trabajos realizados de esta lengua: Randolph Quirk en 1959 comenzó la construcción del Survey of English Usage, que contiene un millón de palabras de inglés británico hablado y escrito; Nelson Francis y Henry Kučera empezaron en 1961 la recopilación del corpus Brown compuesto por un millón de palabras en inglés americano escrito, fue el primer corpus general legibles por computadoras de los textos preparados para la investigación lingüística en Inglés moderno (VARIENG 2010.). En 1975 Stig Johansson comenzó el Lancaster - Oslo/Bergen corpus (LOB) compuesto por un millón de palabras en inglés británico escrito, paralelo al corpus Brown (Johansson 1978).

En la tabla III se presentan los elementos importantes que contienen los corpus mencionados anteriormente.

Tabla III. Elementos estructurales de los corpus.

Nombre del corpus	Elementos estructurales	Tamaño	Fuentes	Lengua	Vigencia (años)
Brown	Genero grupo, categoría, contenido de la categoría, número de textos*	Un millón de palabras	Texto de prosa.	Inglés Americano	47
LOB	Categoría y número de texto*	Alrededor de un millón de palabras	Distribuidos en 15 categorías de texto, 9 informativos y 6 imaginativos.	Inglés Británico	33
BNC	Colección de ejemplos*	100 millones de palabras	El 90% incluye, extractos de los periódicos y diarios regionales y nacionales, letras publicadas e inéditas de libros académicos, ensayos de los memorándums de la escuela y de la universidad, etc. El 10% incluye transcripciones ortográficas de las conversaciones.	Inglés británico	20
ICE	Colección de ejemplos*	1.061.264 palabras	**	Inglés	23
CREA	Palabra, significado y contexto*	160 millones de palabras	El 90 % incluye textos escritos y el 10 % textos orales. Contiene textos publicados en libros, revistas, periódicos, transcripción oral, ciencia y tecnología, política, vida cotidiana, economía arte, salud, ficción, etc.	Español	36

*La descripción de los elementos estructurales de cada corpus no se detallan, dado que no se encuentra disponible la estructura interna de la base de datos, sólo se puede consultar mediante la interfaz gráfica.

**No se encontró este dato.

Tabla III. Elementos estructurales de los corpus (continuación).

Nombre del corpus	Elementos estructurales	Tamaño	Fuentes	Lengua	Vigencia (años)
CORDE	Palabra, significado y gramática*	250 millones de palabras	Ficción: compuesto por textos de verso y prosa (lírico, épico, dramático). No ficción: prosa estructurada en didáctica, científica, de sociedad, de prensa y publicidad, religiosa, histórico-documental y jurídica.	Español	17

*La descripción de los elementos estructurales de cada corpus no se detallan, dado que no se encuentra disponible la estructura interna de la base de datos, sólo se puede consultar mediante la interfaz gráfica.

Gracias a los avances en tecnología informática, en las décadas de 1980 y 1990, se empiezan a recopilar corpus de tamaños considerables, 100 millones de palabras o más. Para el inglés destacan los siguientes: el Cobuild Bank of English, el British National Corpus (BNC) (Oxford 2009) y el International Corpus of English (ICE).

Aparte del inglés, se empiezan a recopilar corpus para otras muchas lenguas, por ejemplo para el español el Corpus de referencia del español actual (CREA) y el Corpus diacrónico del español (CORDE) (RAE s.f.).

Desde mucho tiempo antes de la aparición de las computadoras, se utilizaban corpus como base para los estudios de lingüística descriptiva, normalmente basados en encuestas o cuestionarios. Pero, en la actualidad, se entiende por corpus lingüístico un conjunto de textos en soporte informático representativo de alguna variedad lingüística. Sus características y tamaños son muy variados. Algunos corpus pretenden ser representativos de la lengua en su totalidad y toman extractos de muchos tipos diferentes de textos. Otros son mucho más selectivos y se centran en material que trata solamente de un determinado aspecto lingüístico.

Normalmente, los textos de un corpus están seleccionados de modo que sean lo más representativos posible de la variedad de lengua que se pretende analizar, aunque, debido a lo costoso de su recopilación, a veces se trabaja con conjuntos de textos que no fueron diseñados en principio para los fines con que se utilizan. Es costosa la recopilación de un corpus debido a la gran variedad de textos que existen, el tiempo invertido, las variantes de

los idiomas, el lenguaje y sus reglas no son fijas sino que siguen aumentándose y modificándose, y al regionalismo de cada lengua.

El tamaño de un corpus suele depender de aspectos prácticos, como el tiempo disponible para recopilar, procesar y almacenar los datos.

Hay que tener presente que todos los corpus, independientemente de su tamaño, son limitados, pues no pueden dar cuenta de todos los usos de la lengua. Deben tenerse en cuenta tanto el tamaño como la representatividad del corpus a la hora de valorar el alcance de las conclusiones derivadas de él.

Los datos de los corpus tienen varias ventajas frente a datos basados en la intuición o competencia de los lingüistas: son observables, y por tanto, verificables por otros investigadores.

Además de ser observables, los datos que se pueden extraer de los corpus pueden someterse muy fácilmente a estudios estadísticos, proporcionando una base para un estudio mucho más sistemático y objetivo de la lengua. Hoy en día, existen numerosos programas estadísticos o que muestran contextos de uso, que facilitan las tareas de los lingüistas.

Otra de las características de muchos corpus lingüísticos actuales es la anotación, es decir, el enriquecimiento automático o manual de los textos con información de diverso tipo: estructural, morfológica, sintáctica, semántica, etcétera.

La utilización de corpus puede ser empleada por lingüistas que trabajan en muy diferentes disciplinas: lexicografía, lexicología, sociolingüística, sintaxis, traducción automática, etcétera (Encarta 2009).

Para este trabajo se está realizando el corpus sin indicar los tonos debido a que no va dirigido a lingüistas, pero algunos textos recopilados los incluyen. El tono más frecuente se indica con el número ²³, y el menos frecuente con el ¹², la mayor parte de las palabras españolas que ya están incorporadas al chatino como *préstamos* han agregado el tono indicado con el número ⁴⁵.

Se pueden distinguir ocho tonos distintos en la última sílaba de la palabra, dos de nivel, tres deslizamientos ascendentes y tres deslizamientos descendentes, estos se indican con números de la siguiente manera: ² nivel alto, ⁴ nivel bajo, ²¹ deslizamiento ascendente hasta lo más alto, ³² deslizamiento ascendente hasta el nivel alto, ⁴³ deslizamiento ascendente del nivel bajo, ¹² deslizamiento descendente desde el nivel más alto, ²³

deslizamiento descendente del nivel alto,⁴⁵ deslizamiento descendente del nivel bajo (Pride & Pride 1997).

El proceso de recopilar un corpus es largo y tedioso, por lo que se recomienda crear software que permita automatizar parte de esta recopilación usando como fuente diferentes textos de la lengua específica del corpus.

Dicho software debe ser programado utilizando un lenguaje de programación. Dado que los usuarios del software de un corpus normalmente no serán expertos informáticos, se necesita crear software amigable e intuitivo, y existen lenguajes como Microsoft Visual Basic que permiten hacer dicho software con poco esfuerzo.

3.3. Microsoft Visual Basic

Algunos ubican a Microsoft Visual Basic (VB) como un lenguaje de programación de propósito general, otros le atribuyen al carácter de Desarrollo Rápido de Aplicaciones (RAD), e incluso hay quienes dicen que es un 4GL. Lo más cercano a la verdad, dentro de estas definiciones, es que Visual Basic es un lenguaje de programación de propósito general, pero esa apreciación es sólo una parte de lo que el lenguaje ofrece; en términos generales, Microsoft VB es una plataforma de desarrollo de aplicaciones.

Características

Se seleccionó Microsoft VB 6.0 como plataforma de desarrollo, ya que proporciona elementos gráficos de diseño de aplicaciones, compilador, editor de programas con capacidad de depuración al instante y asistencia basada en colores, ayuda contextual, poderosos asistentes (*wizards*) que automatizan tareas de uso común, tan diferentes en su objeto como en su complejidad, éstas son sólo algunas de sus características.

Microsoft VB 6.0 lleva en su nombre los dos elementos generales que lo conforman. La parte “Visual” indica que es una interfaz gráfica de usuario (GUI), que en lugar de requerir el uso de comandos escritos para realizar tareas, solo requiere el uso de herramientas gráficas, de tal forma que desarrollar se asemeja mucho a estar “dibujando” el trabajo.

La parte “Basic” indica que aquellas tareas que no se pueden realizar a través de las herramientas gráficas, es decir, que no pueden ser dibujadas, son posibles a través de un lenguaje de programación basado en el lenguaje de propósito general llamado Código de Instrucciones Simbólicas Multipropósito para Principiantes (BASIC) con el que prácticamente se puede desarrollar cualquier cosa.

La parte “Visual” está basada en tecnología orientada a objetos (OO), la parte “Basic” es de tipo procedural, y la unión de estas dos tecnologías da lugar a un producto híbrido, orientado a eventos (EO) (Ramírez - Ramírez 2001).

Ventajas

Visual Basic es uno de los lenguajes de programación que más entusiasmo despiertan entre los programadores de PCs, tanto expertos como novatos. En el caso de los programadores expertos por la facilidad con la que desarrollan aplicaciones complejas en muy poco tiempo, en comparación con lo que cuesta programar en Microsoft Visual C++, por ejemplo. En el caso de los programadores novatos por el hecho de ver de lo que son capaces de hacer a los pocos minutos de empezar su aprendizaje.

El precio que hay que pagar por utilizar Microsoft VB 6.0 es una menor velocidad o eficiencia en las aplicaciones.

Visual Basic es un lenguaje de programación visual, también llamado lenguaje de 4ª generación. Esto quiere decir que un gran número de tareas se realizan sin escribir código, simplemente con operaciones gráficas realizadas con el ratón sobre la pantalla.

Visual Basic es también un programa basado en objetos, aunque no orientado a objetos como C++ o Java. La diferencia está en que Microsoft VB 6.0 utiliza objetos con propiedades y métodos, pero carece de los mecanismos de herencia y polimorfismo propios de los verdaderos lenguajes orientados a objetos como Java y C++ (Microsoft 2010.).

Para almacenar los datos en forma permanente, Visual Basic soporta el manejo de archivos y bases de datos. Dado el volumen de datos de un corpus y la cantidad de consultas requeridas, es recomendable utilizar bases de datos.

Visual Basic soporta diversos manejadores de bases de datos que soporten el estándar de acceso a base de datos ODBC. Uno de los manejadores de base de datos más utilizados en la actualidad es MySQL

3.4. MySQL

Mysql es un sistema manejador de base de datos (SGBD). Técnicamente Mysql es una aplicación que permite administrar archivos llamados bases de datos, una base de datos no es más que una colección de datos interrelacionados, ya sean estos textos, números o archivos binarios, los cuales se almacenan y se mantienen debidamente organizados por medio del SGBD (Ullman 2003).

Características

Algunas de las características que proporciona este manejador de base de datos son: alto rendimiento, alta confiabilidad y facilidad de empleo.

Además de ser un sistema de gestión de bases de datos multiusuario y multiplataforma.

Ventajas

Mysql es un manejador de base de datos open source o código abierto, lo que significa que se puede utilizar gratuitamente, su característica de código abierto permite que se desarrollen nuevas aplicaciones o se adquieran existentes (Ullman 2003).

Si el usuario es nuevo a la tecnología de base de datos, MySQL ofrece una gama comprensiva de servicios de las herramientas, de la ayuda, del entrenamiento y de asesoramiento de la base de datos (Oracle 2010).

Además de ser una base de datos robusta y veloz al momento de realizar operaciones, la ventaja de este manejador de base de datos es que tiene alto rendimiento en cuanto al volumen de datos que maneja.

Una vez que se ha dado a conocer los conceptos teóricos importantes de esta tesis, a continuación se plantea en el capítulo cuatro, denominado desarrollo del tema, un análisis detallado del proceso que se llevo a cabo para la realización del corpus.

CAPÍTULO 4. DESARROLLO DEL TEMA

En este capítulo se detalla la forma en que se llevó a cabo la recopilación de los textos en lengua chatina, así como los problemas técnicos que se presentaron al momento de su captura, de igual forma se presenta una descripción del diseño de la base de datos, se describe brevemente el funcionamiento de la aplicación que se realizó y por último se presenta el corpus de texto en chatino desarrollado en trabajo.

4.1. Recopilación de textos en chatino

Para la recopilación de los textos en la lengua chatina en un principio se propusieron tres municipios: Santos Reyes Nopala, Santa Catarina Juquila y Tataltepec de Valdés, Oaxaca, México. Se eligieron estos municipios por dos razones: la primera porque se conoce a personas que habitan en dichas comunidades y proporcionaron documentos escritos en la

lengua chatina, la segunda razón porque son comunidades que no están muy alejadas y se tiene acceso a ellas.

A continuación se explica brevemente como se llevó el proceso de la recopilación de los textos, las personas que se entrevistaron, los lugares que se visitaron y los textos que se consiguieron de Internet.

En la comunidad de San Pedro Mixtepec, se entrevistó al señor Erick originario de Santa Catarina Juquila, que puede hablar la lengua chatina pero desconoce totalmente su escritura, por lo que le interesa la escritura de la misma y está investigando de forma personal manuscritos de la lengua ya que quiere “rescatar” su cultura.

El señor Erick comentó que algunas personas que son Testigos de Jehová le han ido a preguntar acerca del chatino, por lo que se entrevistó a estas personas, que a su vez explicaron los motivos por los que a ellos les interesa la escritura de la misma, y que desafortunadamente no la pueden escribir pero que están trabajando sobre la escritura.

De la comunidad de Santa Catarina Juquila se recopilaron del libro “Chatino de la Zona Alta, Oaxaca” una narración, un diálogo, números, pronombres, un listado de palabras y frases.

En la comunidad de Santos Reyes Nopala, se entrevistó a la Srita. Yolanda encargada de la biblioteca. Ella comentó que en la biblioteca sólo cuenta con un libro referente a este tema, llamado “Buscando el origen de los chatinos de Nopala” escrito por Javier G. Pérez Sánchez en el año 2004, en su segunda edición (en 2010 salió la tercera edición). Este libro sólo contiene un poco de antecedentes y versiones históricas de esta comunidad, de dónde viene su origen, cuáles son sus costumbres, etc.

También mencionó que esta comunidad está perdiendo su cultura en cuanto a la lengua y a las nuevas generaciones ya no les llama la atención hablarla, algunos porque se “avergüenzan” de la misma, porque no existen escuelas bilingües o porque sus padres ya no les inculcan sus culturas, en fin, muchos fueron los problemas mencionados por el entrevistado.

La comunidad de Santos Reyes Nopala se descartó de esta tesis, por las razones citadas en la entrevista y dado que no se contó con apoyo de alguien que conociera la lengua.

Se contactó vía electrónica a los integrantes del proyecto chatino (Proyecto Chatino 2008), solicitando la colaboración de personas que pudieran proporcionar textos en la lengua chatina. Al día siguiente de la petición respondió la lingüista Emiliana Cruz, hablante de la lengua chatina y originaria de Cieneguilla, San Juan Quiahije, quien actualmente reside en Austin, Texas, diciendo lo siguiente:

Desafortunadamente, no tenemos textos escritos en la lengua chatina. Los resultados de los estudios lingüísticos que hemos realizado en el idioma Chatino se encuentran en esta página <http://lenguaChatino.com>, de hecho, es difícil que consigas textos en Chatino, ya que hasta el momento todavía no se ha avanzado con la escritura.

La variante que se ha trabajado más es la variante de San Juan Quiahije, es posible que podamos apoyarte con un texto de San Juan Quiahije.

Otra opinión, a lo mejor el idioma Chatino no es la lengua adecuada para tu tesis, en Oaxaca hay varias lenguas que ya se escriben, por ejemplo el zapoteco, eso podrá ser más fácil para ti, ya que el Chatino no tiene mucho material. El material que nosotros tenemos es más lingüístico (E Cruz, correo electrónico, 03 de marzo de 2011).

Dado que se contó con este apoyo, se decidió utilizar la región de San Juan Quiahije y recopilar más textos de esta zona.

A través de la lingüista Emiliana Cruz se contactó con su hermana Hilaria Cruz, también hablante de la lengua chatina y originaria de Cieneguilla, San Juan Quiahije, quien actualmente reside en Austin, Texas. La lingüista Hilaria Cruz está realizando un proyecto sobre esta lengua de San Juan Quiahije, en conjunto con el joven Luis Adán, quien es originario de Panixtlahuaca y que se encuentra viviendo en Puerto Escondido.

La lingüista Hilaria Cruz apoyó en la recopilación de textos en la lengua de San Juan Quiahije, enviándolos vía electrónica o a través del joven Luis Adán.

La lingüista Hilaria Cruz también proporcionó la dirección de una red social (Cruz 2011) para obtener conversaciones y frases.

De la comunidad de Tataltepec de Valdés se recopilaron algunos enunciados del libro llamado “Libro del Adulto”, además de documentos impresos proporcionados por la M. en I. María de Jesús Estudillo Ayala. Estos documentos contienen un listado de palabras en chatino (números, animales, colores, algunos sustantivos y verbos).

De Internet se descargó la Biblia escrita en chatino de la variante de Tataltepec de Valdés en un archivo pdf., de 2.89 MB.

Dificultades

En resumen, las dificultades encontradas fueron: la mayoría las personas que hablan la lengua chatina, desafortunadamente desconocen su escritura, además de que no existen muchos escritos en chatino.

La recopilación no fue nada sencilla, con base en los comentarios de las personas entrevistadas se observó que mientras algunas personas que no hablan el chatino están interesadas en preservar esta lengua, por lo general quienes la hablan no les interesa en lo más mínimo que se pierda su lengua, ya que ha sido absorbida por el español.

Una vez realizado el proceso de la recopilación de los textos escritos en chatino, se procedió a la captura de los mismos en un archivo de texto, para almacenarlo en un formato digital y posteriormente generar un corpus de texto.

4.2. Captura de textos en chatino

La captura de los textos en chatino se realizó en Microsoft® Bloc de notas, ya que es un editor de texto sencillo, fácil de usar y muy básico para crear o editar archivos en modo de texto, a diferencia de los procesadores de texto que usan archivos binarios.

Se eligió este editor de texto ya que para este documento no se requiere darle formato al texto, efectos, etc. Además de que Microsoft® Bloc de notas es una herramienta que viene incluida en todas las versiones del sistema operativo Microsoft Windows y los archivos de texto no suelen ocupar demasiado espacio.

Al momento de la captura se presentaron algunos problemas técnicos: los documentos recopilados contenían textos con caracteres y al momento de la captura el

editor de texto no permitió escribir esos caracteres por lo que los archivos almacenados son de texto plano.

Para almacenar el corpus generado a partir de estos archivos de texto en forma permanente y dado el volumen de datos de un corpus, se diseñó una base de datos.

4.3. Diseño e implementación de la base de datos para el corpus de la lengua chatina

Después de llevar a cabo la captura de los textos escritos en chatino en un formato digital, se procedió al diseño e implementación de la base de datos para el corpus desarrollado en este documento.

En el diseño, se comienza con la descripción del diccionario de datos correspondiente a este corpus, ya que es necesario determinar los rubros que la componen y de esa forma ajustar el diseño de la base de datos, y conseguir un almacenamiento óptimo con la menor utilización de espacio, así mismo se presenta el modelo Entidad-Relación, que muestra las relaciones que existen entre las diferentes tablas.

Diccionario de datos

Para darle un buen uso a la base de datos y para que cualquier persona que desee tener acceso a ella la puedan consultar, es necesario realizar un diccionario de datos.

A continuación se menciona la descripción de los rubros que componen el diccionario de datos:

- Campo. Contiene el nombre del campo.
- Tipo. Contiene el tipo de dato del campo (numérico, carácter, booleano etc.).
- Tamaño. Contiene el tamaño del campo (número de letras, números de dígitos o decimales).
- Descripción. Significado de cada campo, es decir lo que llevará en su contenido.

Una vez dada la descripción de los rubros que componen el diccionario de datos, en la tabla V se presenta el diccionario de datos correspondiente a la base de datos realizada en este proyecto.

Tabla IV. Diccionario de datos de la base de datos corpus.

Nombre de la tabla: corpus			
Descripción: Contiene los términos en la lengua chatina			
Campo	Tipo	Tamaño	Descripción
idtermino	Numérico		Llave primaria
termino	Caracter	25	Lexema en chatino

Nombre de la tabla: termino			
Descripción: Contiene la traducción, zona y clasificación			
Campo	Tipo	Tamaño	Descripción
idtermino	Numérico		Identificador del término en chatino, llave foránea, de la tabla corpus
traduccion	Caracter	25	Traducción en español que le corresponde al término en chatino
idzona	Caracter	25	Identificador de la zona a la que pertenece el término, llave foránea de la tabla zona
idclasificacion	Caracter	25	Identificador de la clasificación que le corresponde al término, llave foránea de la tabla clasificación

Nombre de la tabla: zona			
Descripción: Contiene las zonas a las que pertenece el termino en chatino			
Campo	Tipo	Tamaño	Descripción
idzona	Numérico		Llave primaria
nombrezona	Caracter	25	Nombre de la zona a la que pertenece el término en chatino (Tataltepec de Valdés, Juquila, San Juan Quiahije)

Nombre de la tabla: clasificación			
Descripción: Contiene la clasificación que le corresponde al término en chatino			
Campo	Tipo	Tamaño	Descripción
idclasificacion	Numérico		Llave primaria
nombrecclasificacion	Caracter	25	Nombre de la clasificación (sujeto, predicado, verbo, pronombre, sustantivo)

Una vez definidos los campos de las base de datos y tener la documentación de cómo se ha diseñado, es necesario definir las relaciones que existen entre las tablas. Para mostrar dichas relaciones, se utiliza el modelo Entidad-Relación.

Modelo Entidad – Relación

La figura 4.1 muestra el modelo Entidad-Relación de la base de datos del corpus.

Se tomó la opción de crear más de una tabla ya que en un futuro, además de estar añadiendo textos, se podrá extender esta base de datos agregando videos y grabaciones de voz.

Las relaciones que se observan en la figura 4.1 están determinadas de esta manera: un término en chatino puede tener más de una traducción, pertenecer a más de una zona y tener más de un tipo de clasificación.

La implementación de esta base de datos se llevó a cabo en el manejador de base de datos MySQL, ya que como se mencionó en la sección 3.4, además de ser uno de los manejadores de base de datos más utilizado en la actualidad, tiene alto rendimiento en cuanto al volumen de datos que maneja.

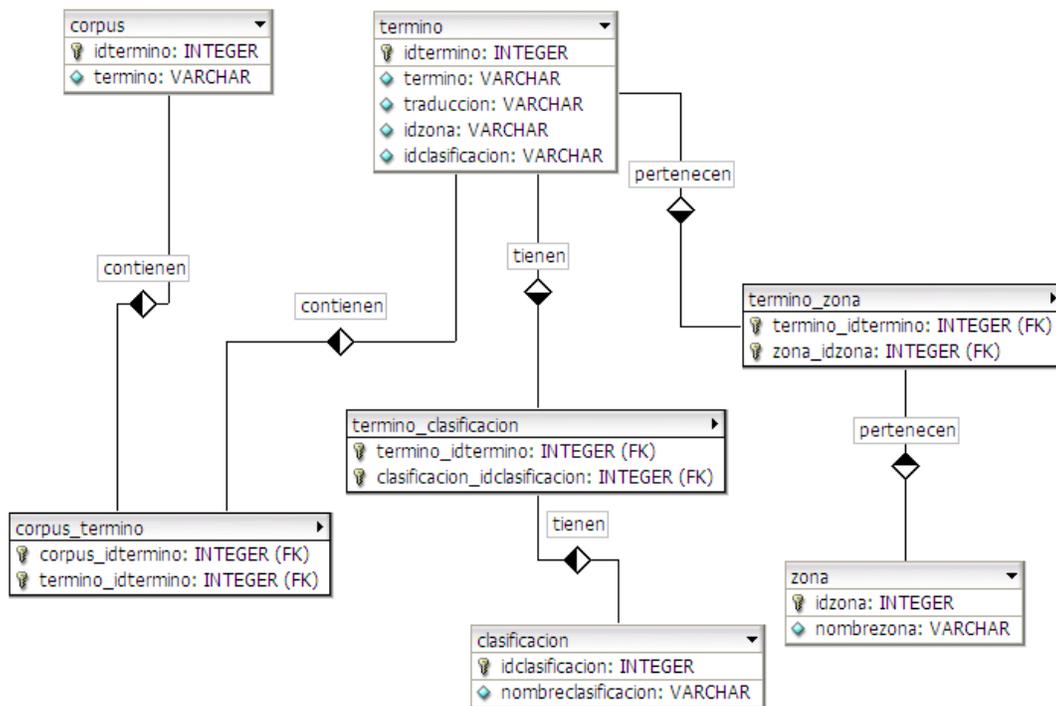


Figura 4.1. Modelo Entidad Relación.

4.4. Desarrollo de la aplicación corpus de texto en chatino

Después de llevar el proceso de la recolección de los textos escritos en la lengua chatina y la captura de los mismos, se implementó una base de datos de acuerdo a los requerimientos del corpus de texto. Para tener un mejor acceso a la base de datos, se desarrolló una aplicación en Microsoft Visual Basic 6.0 que facilita la separación de las palabras de los textos en chatino.

Funcionamiento básico

El funcionamiento básico de la aplicación se muestra a continuación, tomando un ejemplo práctico. Para información más detallada se realizó un manual de usuario, que se encuentra en el anexo C.

La aplicación se basa en un entorno gráfico como el que se muestra en la figura 4.2. Con el botón `Entrar` se entra al sistema de corpus de textos en chatino mostrando la pantalla `CORPUS DE TEXTOS` (Fig. 4.3).

Esta aplicación necesita archivos de texto plano, es decir, archivos sin formato, para mayor detalle consultar la tabla I de la sección 2.5

Al elegir la opción `Abrir` del menú `Archivo` (Fig. 4.4a), automáticamente se abrirá una nueva pantalla con archivos de textos (Fig. 4.5), en donde se elige un archivo de texto en chatino y se mostrará el texto completo en la lista de la pantalla `CORPUS DE TEXTOS` (Fig. 4.6), para este ejemplo en particular se elige el archivo “`enunciados.txt`” perteneciente a la variante de Tataltepec de Valdés. Para fines del ejemplo, este archivo no ha sido previamente procesado por la aplicación, es decir, las palabras contenidas en el archivo no estaban en el corpus.

La opción `Separar Lexema` (Fig. 4.4b) enlista los lexemas del archivo de texto previamente abierto, sin mostrar repeticiones de palabras, como se muestra en la figura 4.7.

El botón `Guardar Lexema`, guarda los lexemas uno por uno. Para este ejemplo se elige de la lista el lexema *kyi* y se presiona el botón `Guardar Lexema`. Como el lexema no se encontraba en el corpus, la aplicación muestra el mensaje de la figura 4.8.

Ahora se desea guardar el listado de los lexemas todos juntos, por tanto se presiona el botón `Guardar Todos`.

Este botón guardará todos los lexemas mostrados en la pantalla y visualizará un mensaje mostrando el número de lexemas guardados y lexemas repetidos. Continuando con el ejemplo, se están guardando los lexemas del mismo archivo mencionado anteriormente, “enunciados.txt”, este archivo cuenta con 12 lexemas; como ya se guardó el lexema *kyi*, se muestra el mensaje de la figura 4.9.

La opción *Ver Corpus* del menú *Corpus* (Fig.4.4b) automáticamente muestra la pantalla *VER CORPUS* (Fig. 4.10).

En esta pantalla el usuario tiene la opción de elegir el lexema, la zona y la clasificación, y escribir la traducción correspondiente al lexema que haya elegido. Las listas desplegables contienen los lexemas, zonas y clasificaciones previamente agregados (Fig. 4.11).

Siguiendo con el ejemplo, lo que se hará es elegir de la lista desplegable el lexema *kyi*, escribir la traducción que en este caso *kyi* significa “papel”, elegir de las listas desplegables la zona a la que pertenece, variante de Tataltepec de Valdés, y la clasificación que le corresponda que sería un sustantivo; se procede a guardarlos en la base de datos dando clic en el botón *Guardar* (Fig. 4.12).

Una vez guardados los datos en el corpus de la base de datos, si se elige el botón *Mostrar Corpus* de la pantalla *VER CORPUS*, se mostrarán en la tabla los lexemas que ya cuenten con su respectiva traducción, zona a la que pertenece y clasificación sintáctica que le corresponde (adjetivo, pronombre, sustantivo o verbo) dentro del corpus. En la sección 4.5 se muestra parte del corpus generado con todos los textos en chatino recopilados.



Figura 4.2. Pantalla PRINCIPAL.

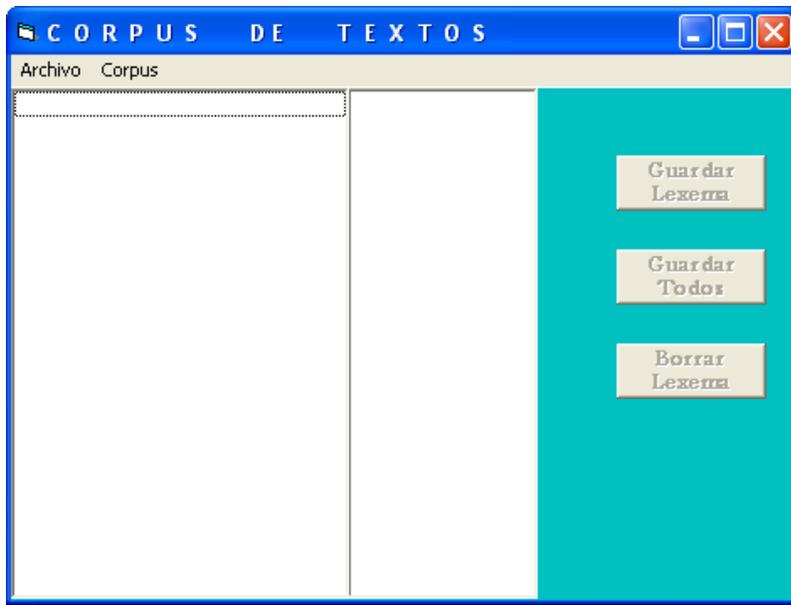


Figura 4.3. Pantalla CORPUS DE TEXTOS.



Figura 4.4. Menús despegables de la pantalla CORPUS DE TEXTOS. (a) Menú Archivo. (b) Menú Corpus.

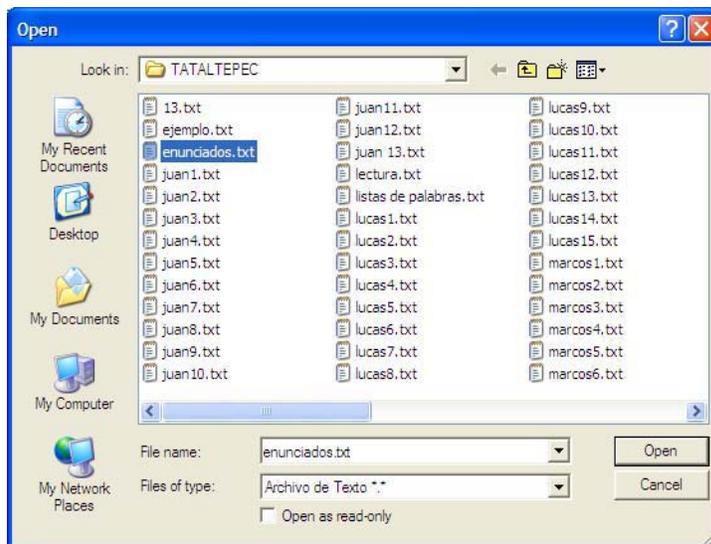


Figura 4.5. Abrir archivos de textos.

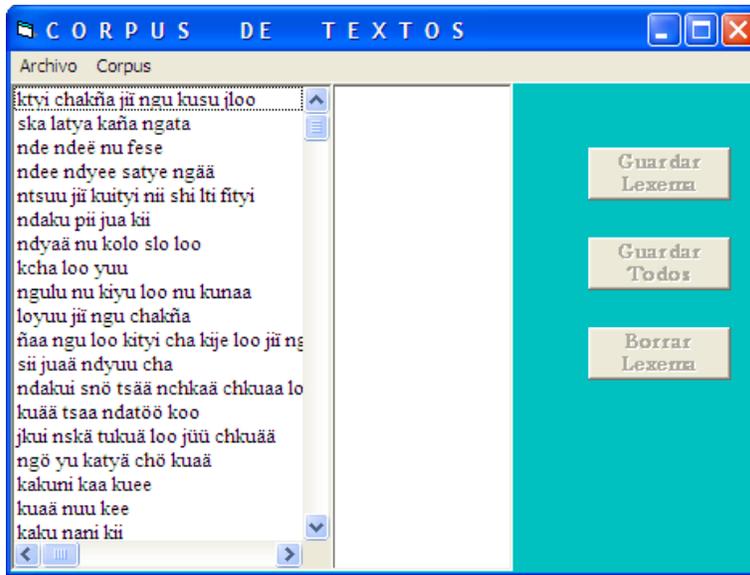


Figura 4.6. Mostrar texto completo.

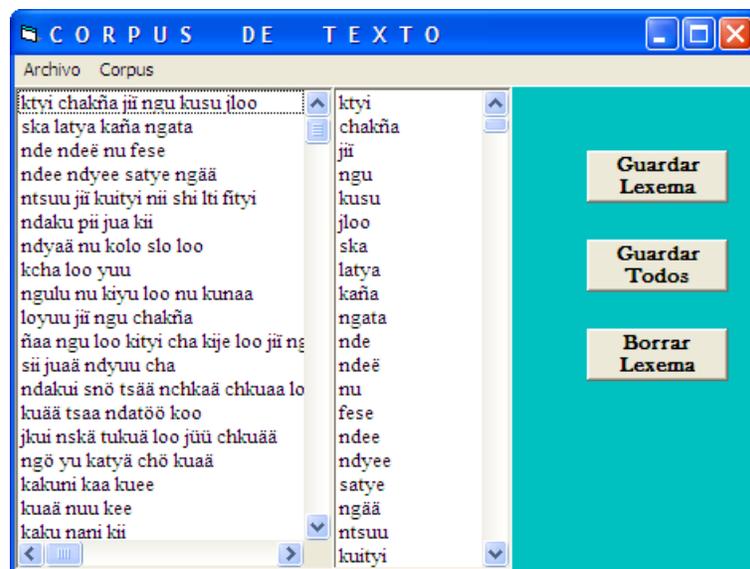


Figura 4.7. Listado de Lexemas.

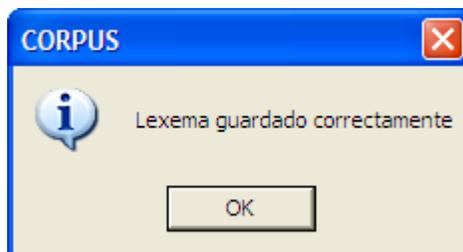


Figura 4.8. Lexema guardado.

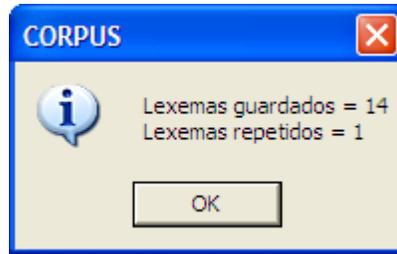


Figura 4.9. Lexemas guardados y repetidos.

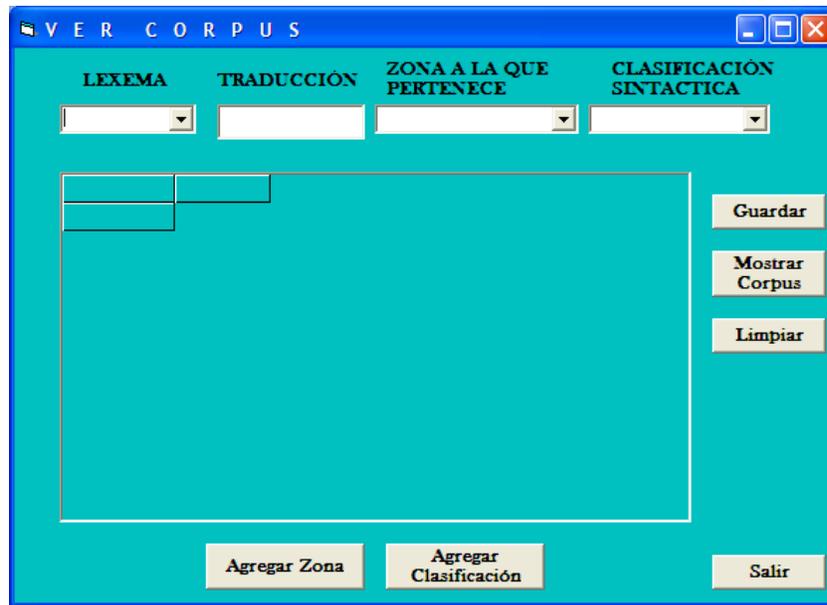


Figura 4.10. Pantalla Ver Corpus.

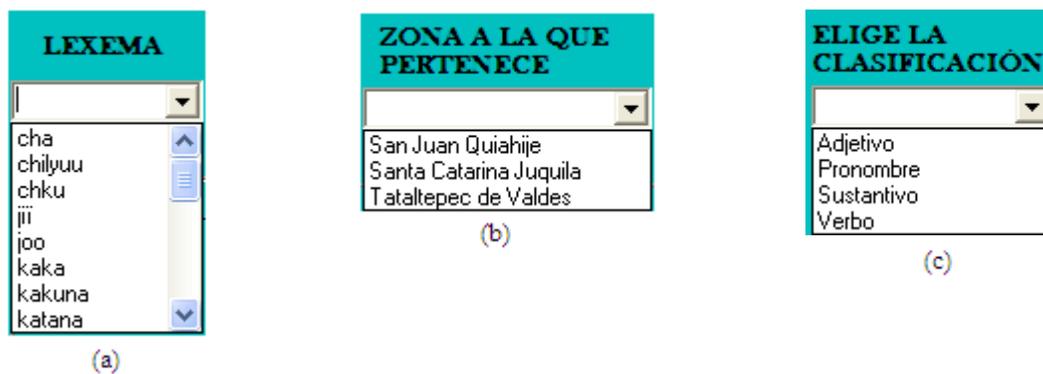


Figura 4.11. Listado de lexemas, zonas y clasificaciones sintácticas existentes en la base de datos.

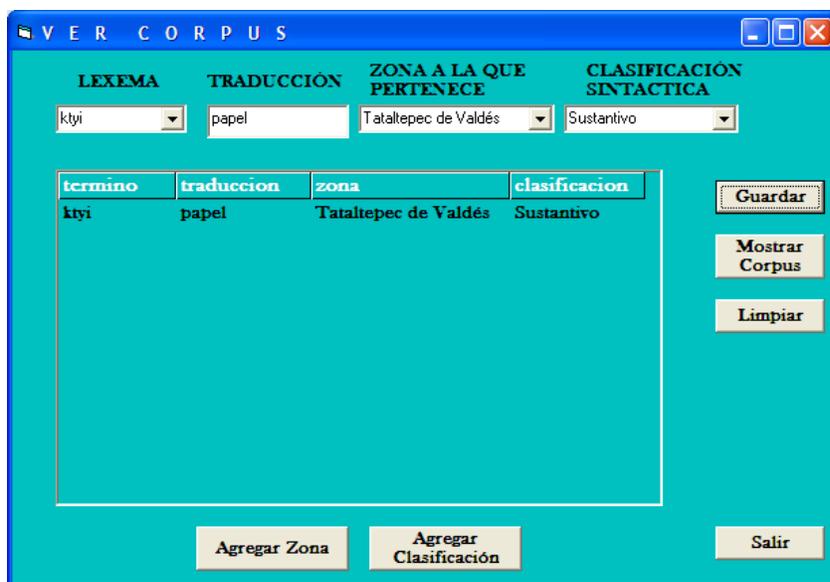


Figura 4.12. Lexema guardado.

4.5. Corpus de texto en chatino

Los textos recopilados; de la variante de Tataltepec de Valdés se capturó la Biblia “Nuevo Testamento”, los primeros 100 números, los pronombres personales, nombres de animales, verbos, sustantivos y enunciados; de la variante de San Juan Quiahije se capturaron los primeros 100 números, los pronombres personales, frases y conversaciones; y de la variante de Santa Catarina Juquila se capturaron los cien primeros números, los pronombres personales, una narración de cómo se celebraban las bodas, un diálogo y un listado de palabras.

Cada texto recopilado tiene características diferentes, algunos tienen traducción, algunos tonos, algunos clasificación, etc. Para información más detallada sobre los textos recopilados, consultar el Anexo B.

El corpus cuenta con aproximadamente 3500 lexemas capturados, de los cuales 526 lexemas cuentan con su respectiva traducción, zona y clasificación sintáctica. En la figura 4.13 se presenta parte del corpus que se muestra al pulsar el botón `Mostrar Corpus` de la pantalla `VER CORPUS`.

No fue posible capturar la traducción de los 3500 lexemas, debido a que no se contó con un hablante de chatino y la mayoría de los textos conseguidos estaban completamente en chatino.

termino	traduccion	zona	clasificacion
itya	agua	Tataltepec de Váldez	Sustantivo
ka	vaca	San Juan Quiahije	Sustantivo
ke	cocinar	Santa Catarina Juquila	Verbo
ku	comer	Santa Catarina Juquila	Verbo
ku'ma	ustedes	Tataltepec de Váldez	Sustantivo
kutsi	amarillo	Santa Catarina Juquila	Sustantivo
la	iglesia	Santa Catarina Juquila	Sustantivo
na	yo	San Juan Quiahije	Pronombre
na'	yo	Santa Catarina Juquila	Pronombre
naä'	yo	Tataltepec de Váldez	Pronombre
ne	ahora	San Juan Quiahije	Sustantivo
nga'a	rojo	Tataltepec de Váldez	Adjetivo
ngata	negro	Santa Catarina Juquila	Adjetivo
ngata	negro	Tataltepec de Váldez	Adjetivo
ngatsi	amarillo	Tataltepec de Váldez	Sustantivo
ntqen	estar	San Juan Quiahije	Verbo
ntyja	dormir	San Juan Quiahije	Verbo
ska	número-uno	Santa Catarina Juquila	Sustantivo
ska	número-uno	San Juan Quiahije	Sustantivo
ska-centu	número-cien	San Juan Quiahije	Sustantivo
ska-siye-tu	número-cien	Santa Catarina Juquila	Sustantivo
ska-syentu	número-cien	Tataltepec de Váldez	Sustantivo
sno'	número-ocho	Tataltepec de Váldez	Sustantivo
snu'	número-ocho	Santa Catarina Juquila	Sustantivo
ti	número-diez	Santa Catarina Juquila	Sustantivo
tii	número-diez	Tataltepec de Váldez	Sustantivo
tsa	día	Santa Catarina Juquila	Sustantivo
tsaka	número-uno	Tataltepec de Váldez	Sustantivo
tsan	día	San Juan Quiahije	Sustantivo
tsää	día	Tataltepec de Váldez	Sustantivo
yaka	árbol	Tataltepec de Váldez	Sustantivo
yka	árbol	Santa Catarina Juquila	Sustantivo

Figura 4.13. Extracto del corpus de la base de datos.

Después de detallar el proceso del desarrollo de esta tesis, en el siguiente capítulo denominado conclusiones y trabajos futuros, se presentan las conclusiones finales y posibles líneas de trabajos futuros.

CAPÍTULO 5. CONCLUSIONES Y TRABAJOS FUTUROS

En este capítulo se mencionan las conclusiones a las que se llegaron a lo largo de este documento y se proponen algunas mejoras que se le pueden implementar a este trabajo.

Conclusiones

En esta tesis se ha cumplido el objetivo principal propuesto, generar un corpus digital y desarrollar una interfaz para consultar las unidades léxicas, definidas a partir de documentos en chatino.

Los textos en lengua chatina fueron recopilados en los municipios de San Juan Quiahije, Santa Catarina Juquila y Tataltepec de Valdés, Oaxaca, México. El municipio de Santos Reyes Nopala, debido a las dificultades enfrentadas al momento de llevar a cabo la recopilación, se sustituyó por el municipio de San Juan Quiahije, en el “Capítulo 4. Desarrollo del Tema” se explican las razones de esta decisión.

Los documentos que existen sobre esta lengua no se encontraban en formato digital, por lo tanto, se capturaron los escritos recopilados en archivos de texto en formato electrónico.

Los textos que se recopilaron se encuentran en el disco compacto CD anexo a este trabajo, dentro de la carpeta *Textos en Chatino*. Una descripción de los mismos se encuentra en el Anexo B.

Para facilitar la separación de las palabras de los textos en chatino y para poder consultar la base de datos de manera sencilla, se desarrolló una aplicación en Microsoft Visual Basic 6.0, la cual cuenta con cuatro módulos: “Corpus de Textos”, “Ver Corpus”, “Zona” y “Clasificación”.

Por último, se generó el corpus digital, el cual contiene 526 lexemas extraídos de los escritos en chatino que se recopilaron cada uno con la traducción que le corresponde, la zona a la que pertenece y su clasificación sintáctica (adjetivo, pronombre, sustantivo o verbo).

En la tabla IV se puede observar que el corpus desarrollado en este trabajo contiene elementos estructurales que ninguno de los corpus mencionados presenta, por lo que en el Corpus Chatino se pueden realizar más tareas sobre el procesamiento del lenguaje natural.

Tabla V. Comparación de los corpus.

Nombre del corpus	Elementos estructurales	Tamaño	Fuentes	Lengua	Vigencia (años)
Corpus Chatino	Palabra, traducción, clasificación sintáctica y zona geográfica	3,458 palabras	Contiene textos de la Biblia, lista de palabras y libros.	Chatino	-
Brown	Genero, grupo, categoría, contenido de la categoría, número de textos*	Un millón de palabras	Texto de prosa.	Inglés Americano	47

*La descripción de los elementos estructurales de cada corpus no se detallan, dado que no se encuentra disponible la estructura interna de la base de datos, sólo se puede consultar mediante la interfaz gráfica.

Tabla V. Comparación de los corpus (continuación).

Nombre del corpus	Elementos estructurales	Tamaño	Fuentes	Lengua	Vigencia (años)
LOB	Categoría y número de texto*	Alrededor de un millón de palabras	Distribuidos en 15 categorías de texto, 9 informativos y 6 imaginativos.	Inglés Británico	33
BNC	Colección de ejemplos*	100 millones de palabras	El 90% incluye, extractos de los periódicos y diarios regionales y nacionales, letras publicadas e inéditas de libros académicos, ensayos de los memorándums de la escuela y de la universidad, etc. El 10% incluye transcripciones ortográficas de las conversaciones.	Inglés británico	20
ICE	Colección de ejemplos*	1.061.264 palabras	**	Inglés	23
CREA	Palabra, significado y contexto*	160 millones de palabras	El 90 % incluye textos escritos y el 10 % textos orales. Contiene textos publicados en libros, revistas, periódicos, transcripción oral, ciencia y tecnología, política, vida cotidiana, economía arte, salud, ficción, etc.	Español	36
CORDE	Palabra, significado y gramática*	250 millones de palabras	Ficción: compuesto por textos de verso y prosa (lírico, épico, dramático). No ficción: prosa estructurada en didáctica, científica, de sociedad, de prensa y publicidad, religiosa, histórico-documental y jurídica.	Español	17

*La descripción de los elementos estructurales de cada corpus no se detallan, dado que no se encuentra disponible la estructura interna de la base de datos, sólo se puede consultar mediante la interfaz gráfica.

**No se encontró este dato.

Trabajos futuros

Es importante mencionar que este proyecto es el primer trabajo realizado en esta institución enfocada al procesamiento de lenguaje natural e investigación sobre lenguas indígenas, y se espera que este trabajo sea la base para la realización de nuevos proyectos relacionados a estas áreas.

A continuación se proponen posibles mejoras que se pueden implementar a este trabajo.

- Implementar al módulo “Ver Corpus” las opciones Modificar y Eliminar.
- El corpus no cuenta con ejemplos del uso de cada lexema, ya que en esta investigación no se contó con un hablante de la lengua chatina, pero si se contara por lo menos con una persona hablante de cada variante de chatino, se le podrían adicionar ejemplos.
- Agregar tonos de pronunciación a cada lexema. Para ello se necesitan lingüistas.
- En este trabajo sólo se abordaron las variantes de Tataltepec de Valdés, Santa Catarina Juquila y San Juan Quiahije, con el tiempo se podrían agregar textos de todas las variantes que existen.
- Crear corpus de otras lenguas además del chatino, por ejemplo; mixteco, náhuatl, zapoteco, etc.
- Este trabajo podría ser la base para crear diccionarios, traductores y todos los demás productos del procesamiento de lenguaje natural.

ANEXO A. CHATINOS

En este anexo se da una introducción de los chatinos, asimismo una descripción de los municipios donde se habla este idioma.

Microrregión chatina (Pérez-Sánchez 2004)

Dentro de la región de la Costa Oaxaqueña, se localiza la microrregión Chatina en el sureste del Estado de Oaxaca ocupando los Distritos de Juquila y Sola de Vega en un área aproximada de 7677 km². Sus límites etnográficos son los siguientes: al norte colinda con los zapotecos al igual que por el este; al norte y al oeste limita con los mixtecos y al sur con los negroides de la costa. Geográficamente, abarca desde las altas montañas de la Sierra Madre del Sur que alcanzan alturas de más de dos mil quinientos metros sobre el nivel del mar, hasta las angostas llanuras costeras del Océano Pacífico, que pertenecen estas últimas al distrito de Juquila; su territorio forma una cuchilla que inicia en el litoral del Pacífico y se va angostando a medida que penetra en el estado.

El relieve es montañoso como la mayor parte de entidad, destacan algunas elevaciones de importancia como el cerro de la Virgen en el distrito de Juquila y el cerro de la Cumbre que pertenecen a Sola de Vega; aparte de las planicies costeras existen algunos pequeños valles en medio de las montañas.

Debido a lo accidentado de su orografía, se tienen climas variados como calurosos húmedos de la costa, caluroso seco en los lugares bajos y que se localizan a espaldas de las montañas como San Pedro Juchatengo, Juquila; templado tropical en los pequeños valles y frío húmedo en las altas montañas.

Las lluvias se presentan en el verano y parte del otoño, siendo la Chatina una de las zonas más lluviosas del estado.

Todos sus ríos pertenecen a la vertiente del Pacífico por lo que guardan las mismas características, sus aguas son veloces, sus cauces angostos en la mayor parte de su longitud y sólo se vuelven anchurosos al acercarse al Océano Pacífico. Los principales ríos son: El Atoyac-Juchatengo o Verde, el de Santiago Amoltepec, el del Maíz o de Manialtepec (que pasa por Nopala), el Ranas, y llevan agua todos los días del año.

En sus bosques, tanto tropicales como mixtos, se encuentran maderas como pino, cedro, caoba, huanacaxtle, encino, macuil o palo de rosa. Su fauna comprende una gran variedad de mamíferos muchos de ellos en peligro de extinción: venado, tigrillo, tejón, oso hormiguero, marta, onza, león, jabalí, danta (tapir), ocelote, perro de agua (nutria) y muchos reptiles.

La microrregión Chatina está comunicada con la capital del estado por vía terrestre por una carretera que por mucho tiempo fue terracería en más de ciento cincuenta kilómetros y que a partir de 1991, afortunadamente, el entonces Jefe del Ejecutivo Federal, Lic. Carlos Salinas de Gortari, y el Mandatario Estatal, Lic. Heladio Ramírez López, se percataron del potencial que en materia agrícola, ganadera, forestal, comercial y turística existen en los pueblos chatinos y ordenaron el inicio de los trabajos de pavimentación de la carretera Sola de Vega - Puerto Escondido así como el ramal Juquila - Cerro del Vidrio, ambas en su primera etapa, y se terminó en 1996 la última etapa que abarca el tramo Ojo de Agua - San Gabriel Mixtepec, con lo que quedó concluida la carretera Oaxaca - Puerto Escondido.

Sin duda alguna, con esta obra la zona chatina se integra al desarrollo de la entidad y seguramente aportará sus riquezas al pueblo oaxaqueño, además de acercar aun más a la capital del estado con la zona turística de Puerto Escondido y Bahías de Huatulco al igual que con el Santuario de la Virgen de Juquila.

Dentro de la región chatina se localizan poblaciones muy importantes, por ejemplo: San Miguel Sola de Vega, San Pedro Juchatengo, Santa Catarina Juquila, Santiago Yaitepec, Panixtlahuaca, Tataltepec de Valdés, San Juan Lachao Pueblo Nuevo, San Gabriel Mixtepec, San Pedro Mixtepec, Santiago Cuixtla y Santos Reyes Nopala.

La población chatina se concentra principalmente en los siguientes municipios: Santos Reyes Nopala, San Juan Quiahije, San Miguel Panixtlahuaca, Santiago Yaitepec, Santa Cruz, Zezontepec, San Juan Lachao, Santa María Temaxcaltepec, Santa Catarina Juquila, Tataltepec de Valdés y la Villa de Tututepec, casi todos pertenecientes al Distrito de Juquila en asentamientos dispersos. Los vestigios de la cultura chatina en la costa se remontan alrededor del año 400 a.C. aproximadamente, aunque las estelas en Nopala datan del 600 a 800 d. C (Fig. A.1).



Figura A.1. Municipios donde se concentra la población chatina.

Cultura chatina

Igual que otros pueblos indígenas, los chatinos tienen sus mitos de creación tomados de figuras naturales sobresalientes, transmitidas por medio de la tradición oral y forman parte de su visión del mundo que también es concebida como una manera de actuar. Del panteón divino sobresalen la Santa Abuela, el Santo Padre Dios, la Santa Madre Tierra, la Santa

Madre Luna, los Dioses del Agua, del Viento, de la Lluvia, de las Montañas, las Santas Ciénegas y la Santa Lumbre o Santo Fuego.

Una parte de la cultura Chatina se encuentra establecida en la comunidad de Río Grande, perteneciente al Municipio de Villa de Tututepec, específicamente dentro del área que componen las colonias Las Flores, Benito Juárez, Loma Bonita y La Torre, donde hasta hoy en día han predominado sus costumbres y tradiciones.

Una de éstas y que poco se ha hablado en algunas investigaciones realizadas sobre la cultura chatina, es la del Santo Trapo o *cuu joo steé*, esta es una de las tradiciones que aún se conservan en la cultura chatina de Río Grande. Consiste en que a los 7 días después del nacimiento de un bebé, la familia de éste le organizan un pequeño convivio, donde junto al lugar donde se encuentra la madre y el niño se colocan dos piedras en forma de pirámide, cubriéndose un lado y dejando al descubierto el otro que es el lugar por donde se coloca la leña para atraer el Santo Fuego, mismo que quemará o consumirá todo lo que ahí se deposite en su honor. Esto se realiza antes de que salga el sol, posteriormente las mujeres de la familia, se preparan para hacer el atole dulce, un mole de pollo, memelas rellenas de frijol, y las tortillas de maíz hechas a mano. Cuando se ha terminado de preparar toda la comida, se reúne la familia del recién nacido alrededor de las piedras colocadas anteriormente, donde comienzan a depositar un poco de cada uno de los productos preparados para que sea la Madre Tierra y el Santo Fuego quienes inicien la convivencia. Los participantes en el ritual piden en esos momentos a Dios por la vida y la salud del nuevo ser que se ha integrado a la familia.

Posteriormente, se procede a servir la comida, disfrutando todos del atole, el mole de pollo y las memelas rellenas de frijol. Cabe destacar que los abuelos maternos y paternos, tíos, hermanos, primos, tendrán que permanecer en el domicilio hasta que el Padre Sol se haya ocultado para poder retirarse a sus respectivos hogares (Alberto-Robledo 1994).

Hay alrededor de 40 fiestas patronales en la región y cada comunidad celebra, además, rituales referidos a las labores agrícolas, velaciones, fiestas de barrio o de comunidad. La única fiesta regional importante es la del 8 de diciembre en honor a la virgen de Juquila. En Zezontepec, las celebraciones se acompañan de teatro, música, danza, juegos pirotécnicos, misas, comidas rituales, el inventario del Santo, y la llevada de sus

bienes a la casa del nuevo mayordomo, quien es la persona encargada de ciertas cofradías religiosas (Carrizosa s.f.).

Origen de los chatinos (Pérez-Sánchez 2004)

El origen del pueblo chatino es un misterio; éste no se ha podido precisar por la falta de datos sobre su historia, los que existieron, tal vez como en otras culturas, fueron destruidos por los conquistadores españoles a su llegada a estas tierras, por considerarlos obra del demonio.

Según sus tradiciones, los chatinos se creen originarios del mar. Algunas otras consejas hay de que los chatinos descendieron de grandes embarcaciones o canoas a la playa, procedentes de un país desconocido, ubicado en una isla en medio del océano. Pero su estrecha relación lingüística con el pueblo zapoteca descarta la hipótesis de que no sean autóctonos de las tierras que hoy habitan y hace suponer que su pasado prehistórico está ligado con el de aquella cultura.

En una de las versiones históricas sobre el origen de los chatinos se dice que algunos navegantes, arrebatados por alguna tempestad desde las islas de la Australia y costeano desde el lugar de su procedencia en la misma América, pudieron desembarcar en la laguna de Chacahua, en los márgenes del Río Grande o en alguna barra inmediata, y determinados a permanecer en el lugar, formaron las poblaciones de la Parroquia de Juquila, extendiéndose después hasta Teojomulco y Tezacoalco.

En otra de las versiones, se dice que fue su arribo por el Océano Pacífico en desembarco desastroso, directamente al sur de la ciudad de Oaxaca, frente a la desembocadura del Río Grande y en los acantilados del cerro de la Encomienda, en tales acantilados azotó la embarcación arrojada por un huracán y en pedazos se fue al fondo, pero dejando en tierra náufragos suficientes para dar origen a un pueblo en donde desde entonces se llama La Encomienda.

Este fue el origen de la cultura chatinteca, de donde comienza a extenderse, primero en la feracidad hirviente de los bajos endémicos y luego en las montañas cercanas que brindaban mayor seguridad en la salud y en el peligro de otras fuerzas extrañas. Y sus vestigios afloran en todo donde se remueve la tierra como en Piedra Parada, El Sapo Viejo, donde puede ser que transitaron en su fase nómada, extendida desde Chacahua hasta Puerto

Escondido, para ir asentándose después hasta el pie de montaña Mixteca y los lindes de la Serranía Zapoteca, como pueden atestiguarlo otros vestigios monolíticos y cerámicos más perfeccionados en Yaitepec, Cerro Peine, y contornos de Nopala y Teojomulco.

El reino Chatino mixteco permaneció cerca de 300 años, hasta la llegada de los Mexicas. Sin embargo por varias razones siempre conservaron alguna parte de su región independiente de los dominadores.

Zona

A continuación se describen las siete zonas o municipios donde se habla la lengua chatina.

Tataltepec de Valdés (Inafed 2009)

Significa "En el cerro del abuelo". Y se compone de *Tata* que quiere decir "abuelo", *L* de ligadura eufónica, *Tepetl* que significa "cerro" y *C* que quiere decir "en", esta población se fundó aproximadamente entre los años 400 y 300 a.C.

Se localiza en la región de la costa del Estado, en las coordenadas 97°33' longitud oeste, 16°18' latitud norte y a una altura de 370 metros sobre el nivel del mar.

Limita al norte con Santa Cruz Zenzontepec (municipio perteneciente al Distrito de Sola de Vega); al sur con Villa de Tututepec de Melchor Ocampo; al oriente con San Juan Quiahije y San Miguel Panixtlahuaca; al poniente con Santiago Tetepec y Santiago Jamiltepec ambos municipios pertenecientes al Distrito de Jamiltepec (Fig. A.2).

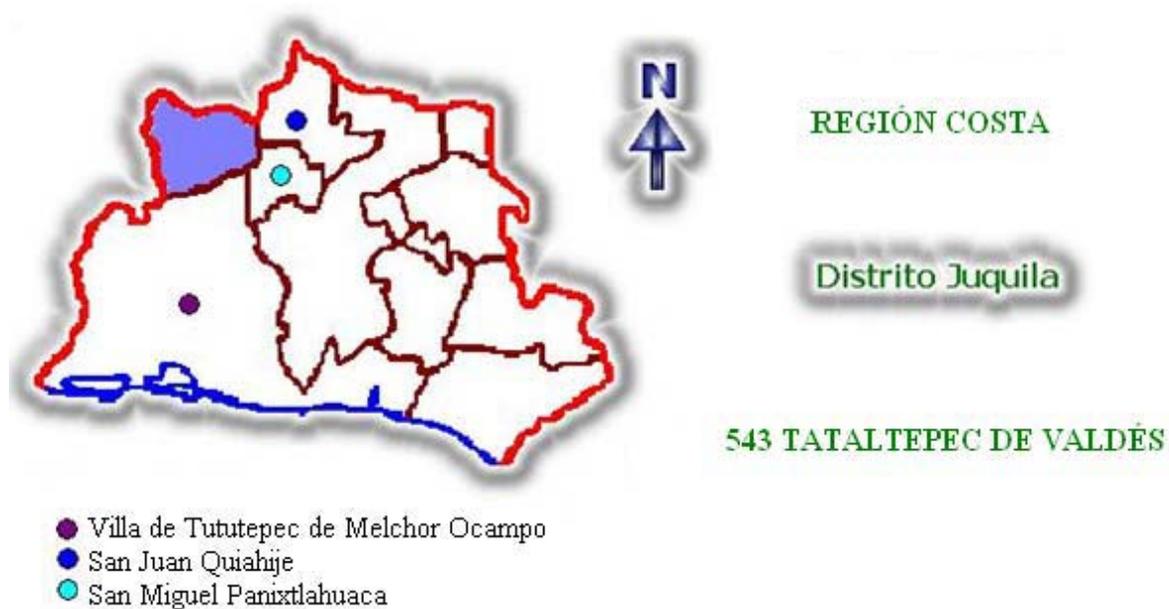


Figura A.2. Localización de Tataltepec de Valdés.

Entre su principal flora y fauna se puede encontrar:

Flora: Pinos de ocote, gretados, encinos, cedro, macuil, zopilotes, macahuites y frutillos, bugambilias, rosales de diferentes tipos: rosa de castilla, rosa durazno, labio de mujer, tulipanes.

Fauna: Leones, tigrillos, venados, tejones, mapaches, iguanas, tucanes, armadillos, coyotes, zorros, tlacuaches, faisán, chachalacas, zopilotes, garzas, gavilanes, águila, palomas, culebras, conejos, vacas, caballos, burros, jabalís, ardillas, pericos, cotorras, loros, guacamayas, tecolotes, golondrinas, cenizotes, calandrias, zanates, peces y camarones.

Santa Catarina Juquila

El nombre de Juquila se compone de las palabras: *Xiuhquilla* que significa “lugar de la legumbre hermosa”, se compone de *Xiuh*, “cosa hermosa”; *Quilitl*, “legumbre”; la variante *Tla*, “Lugar de”. En chatino lleva el nombre de *Sku'we*, que quiere decir: “Quelite tintóreo” (Guzmán-Ventura 2009).

La comunidad de Santa Catarina Juquila, se fundó en el año de 1272. Según textos antiguos, los primeros habitantes vinieron de la región de la mixteca, quienes se establecieron a las orillas de un caudaloso río que les proporcionaba las condiciones apropiadas para un buen desarrollo empezando a poblar lo que en la actualidad es Juquila.

En 1725 por su importancia religioso–turística se le dio la categoría de cabecera municipal y distrito político (Inafed 2009).

El municipio de Santa Catarina Juquila se ubica en la porción sur del Estado de Oaxaca, en las coordenadas de latitud norte 16° 14' y 97° 17' de longitud oeste, a una altura de 1,460 metros sobre el nivel del mar (Fig. A.3) (Inafed 2009).

Colinda con los siguientes municipios: al norte con Villa Sola de Vega y Santiago Minas (ambos del Distrito de Sola de Vega), al sur con Villa de Tututepec de Melchor Ocampo, al este con Santiago Yaitepec, San Juan Lachao, San Pedro Juchatengo y Santos Reyes Nopala, al oeste con Villa de Tututepec de Melchor Ocampo, San Miguel Panixtlahuaca y San Juan Quiahije (Guzmán-Ventura 2009).

Entre su principal flora y fauna se puede encontrar:

Flora: Bosques maderables: pinos, de las variedades chino y lacio; maderas finas como el cedro rojo, roble, macuil y zopilote; encino en sus tipos: de piedra, de tierra, de mosco, de hoja ancha, de cucaracha, de espina y prieto; palo de cuachepil, palmito y fresno. Plantas curativas: el cuajilote (contra la tos y los parásitos), el suace, la hierba de zorrillo (contra la epilepsia), mirto, ruda, itamorrial o camote de venado (para aliviar los golpes internos), florifuncio blanco (para el mal de ojo o fiebre), el potonchihuite para atenuar la fiebre, el bejuco (para fracturas de huesos), y la hoja de anona. Flora comestibles: bejuco de agua, la palma de corozo, chilacayotas, las granadas, el plátano de Costa Rica, de la India, de roatán, perón, manzano, enano y guineo. Se producen también las anonas, limas, naranjas y cañas. Otras plantas comestibles: calabacitas, cuatecos, chepiles, berros, verdolagas, hierbamora, hierba santa y camote de platanar. La flor del palmito y de cuachepil. Los nanacates: de corral, de cuachepil, amarillos, de sal, de flor, de orejita, de lucero. También frijol temprano, cuarenteño, ancho, delgado, bayo y blanco, chile tusta, piquín, canario y paradito.

Fauna: Animales silvestres: venado, el armadillo, danta (tapir), jaquimia, cuaño (ardilla), tejón, zorrillo, tlacuache, tigrillo, onza, marta, nutria (perro de agua). Reptiles: víbora de cascabel, culebra voladora de maguey, ratoneras, coralillos, chuchupasles, tilcuates, la culebra de orejita y la de agua. Aves: tingüilros, perdices, guacamayas, pericos y zanates. Insectos: chinches, nigua, hormigas rojas, arrieras, garrapatas, conchudas,

pinolillos (larvas de garrapatas), toro pinto, chicatanas, tábano, avispa negra o amarilla y cuquitos que producen miel y cera negra (Guzmán-Ventura 2009).



Figura A.3. Localización de Santa Catarina Juquila.

San Juan Quiahije (Inafed 2009)

Quiahije significa en zapoteco "Piedra Montes", se compone de *quia* que quiere decir "piedra" y *guixe* que significa "manto", en Chatino lleva el nombre de *Quihuixí*, que significa: "Zacate de anís" de *qui*, zacate, y *huixí*, anís.

Se localiza en la región de la costa del Estado, en las coordenadas 97° 19' longitud oeste, 16° 18' latitud norte y a una altura de 1,960 metros sobre el nivel del mar. Limita al norte con Santiago Minas y Santa Cruz Zenzontepec (ambos del distrito de Sola de Vega); al sur con San Miguel Panixtlahuaca y Santa Catarina Juquila; al oriente con Santa Catarina Juquila; al poniente con Tataltepec de Valdés (Fig. A.4)

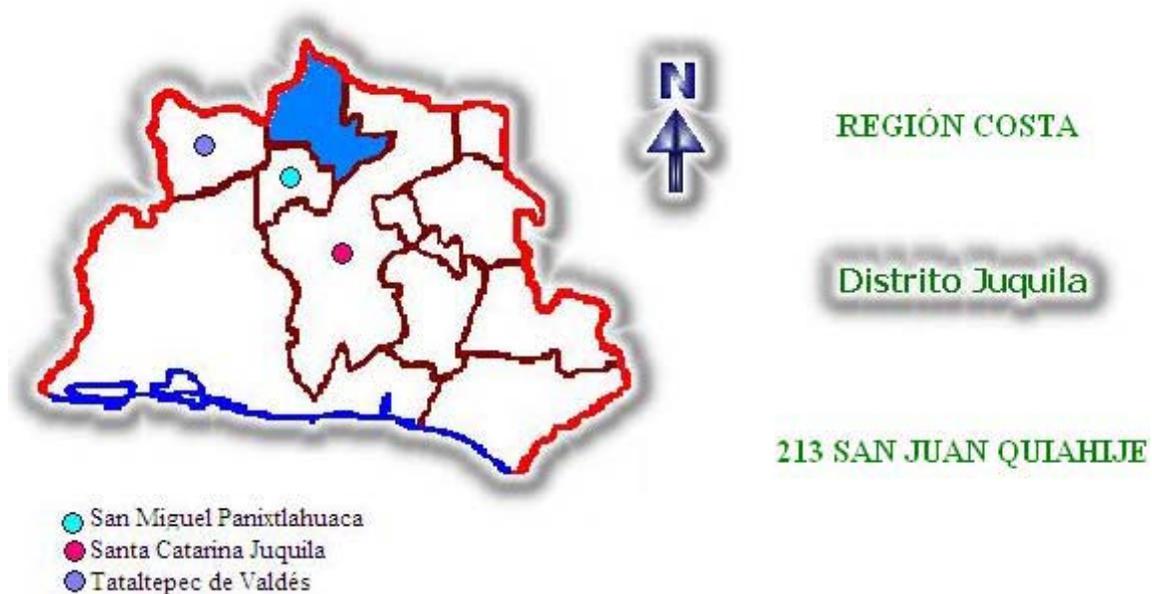


Figura A.4. Localización de San Juan Quiahije.

Entre su principal flora y fauna se puede encontrar:

Flora: Pinos, encinos.

Fauna: Venados, conejos, iguanas, armadillos, ardillas, tejón, mapaches, jabalí, zorrillo, tigrillos, zorros y coyotes.

Santos Reyes Nopala

El vocablo de Nopala, deriva de *Nopatlan*, que significa “llano de nopales” en náhuatl, el nombre castellano de Santos Reyes, se relaciona con la evangelización cristiana.

El nombre chatino de Nopala es *Bya* que significa “llano de nopales” (Inafed 2009).

El municipio de Santos Reyes Nopala se ubica al sureste del Estado de Oaxaca ocupando un área de 196.48 km², en el sureste de la zona chatina.

Su territorio comienza desde las llanuras costeras del océano Pacífico entre Puerto Escondido y San José Manialtepec y penetra hasta las montañas de la Sierra Madre del Sur en donde se establece la población chatina de Santa Lucía Teotepec, Agencia Municipal de este Ayuntamiento (Pérez-Sánchez 2010).

Santos Reyes Nopala se localiza a los 16°8' de latitud norte y a los 93° 18' 30'' de longitud oeste, limita al norte con los municipios de Santa María Temaxcaltepec, Santiago Yaitepec y San Juan Lachao, al sur con el Municipio de San Pedro Mixtepec y Villa de Tututepec de Melchor Ocampo, al este colinda con el Municipio de San Gabriel Mixtepec y al oeste con la cabecera distrital de Santa Catarina Juquila (Fig. A.5) (Inafed 2009).



Figura A.5. Localización de Santos Reyes Nopala.

Entre su principal flora y fauna se puede encontrar:

Flora: Es variada y típica de la región cálida, se encuentran palmeras así como variedad de plantas.

Fauna: Venado, tejón, iguana y armadillo.

Santiago Yaitepec (Inafed 2009)

Significa "En los tres cerros", se compone de *Yai* que significa "tres", *tepetl* que quiere decir "cerro" y *c* que significa "en".

Se localiza en la región de la costa del Estado, en las coordenadas 97° 16' longitud oeste, 16° 13' latitud norte y a una altura de 1,840 metros sobre el nivel del mar (Fig. A.6).

Limita al norte, oeste y sur con Santa Catarina Juquila, al norte y este con San Juan Lachao, al sur con Santos Reyes Nopala y al este con Santa María Temaxcaltepec.



Figura A.6. Localización de Santiago Yaitepec.

Entre su principal flora y fauna se puede encontrar:

Flora: Pino, oyamel, ocote, madroño, moral, encino, zacatón, cucharita y aile.

Fauna: Tlacuache, zorrillo, armadillo, liebre, conejo, culebra y en extinción el venado.

Santa Cruz Zenzontepec (Inafed 2009)

Zenzontepec significa “en la infinidad de cerros” o “en los muchos cerros”, literalmente significa “cuatrocientos cerros”. Proviene de *Cenzontli* que significa “cuatrocientos” y *Tepetl* que significa “cerros”. Actualmente, se le conoce como “el trinar de pájaros Zenzontle entre cuatrocientos cerros”.

Se localiza en la región de la Sierra Sur, pertenece al Distrito de Sola de Vega a una altura de 950 metros sobre el nivel del mar en coordenadas 16°32' latitud norte y 97°30' longitud oeste (Fig. A.7).

Limita al norte con Santiago Amoltepec y Santa María Zaniza, al este con Santo Domingo Teojomulco, San Jacinto Tlacotepec y Santiago Minas, al oeste con Santiago

Ixtayutla y Santiago Tetepec (ambos del Distrito de Jamiltepec) y al sur con Tataltepec de Valdés y San Juan Quiahije (ambos del Distrito de Juquila).



Figura A.7. Localización de Santa Cruz Zenzontepec.

Entre su principal flora y fauna se puede encontrar:

Flora: Flores: Flor del río, jazmín, girasol, flor de muerto, flor de chiche, bugambilia, uña de tigre, flor de nanche, tronadora, laurel, florifundio, flor de leche, flor de gallo, flor de malacate, flor de cacahuanano, flor de paraíso, flor acahual, azucena, flor de jícara, flor de guayaba, flor de piragüita, flor pegajosa. Plantas comestibles: quintonil, chepil, guachepil, chepiche, pitona, nopales, hierbabuena, pie de gallo, hierbasanta, epazote, cogollo de huichicata, flor de calabaza, guía de calabaza, papaloquelite. Árboles: mangal, cuapinol, cuachepil, huanacasle, pino, cedro, cacalotillo, cuajiniquil, huaje, albercoque, venenillo, pochote, encino negro, encino amarillo, zapote negro, cangrejillo, cacahuanano, ciruelo, mamey, árbol de nanche, guanabo, árbol de lima, naranja, toronja, macahite, sauce, fresno, nogal, garoble, tepehuaje, aguacate. Frutos: mango, chirimoya, mamey, lima, limón, zapote negro, nanche, ciruela, cocos, aguacate, limón, papaya, piña.

Fauna: Es representada por las siguientes especies: león, tigre, jabalí, venado, tejón, tlacuache, zorro, armadillo, ardilla, mapache, zorrillo, murciélagos.

De todas las variantes de la lengua chatina habladas en los anteriores municipios, para este documento de tesis, se trabajó con las variantes de Tataltepec de Valdés, Santa Catarina Juquila y San Juan Quiahije.

ANEXO B. TEXTOS CAPTURADOS

En este anexo se da una breve descripción de los textos recopilados de las variantes de: Tataltepec de Valdés, Santa Catarina Juquila y San Juan Quiahije, así mismo se menciona cuáles textos se capturaron.

De la variante de Tataltepec de Valdés, se capturaron en archivos de texto los siguientes documentos:

- La Biblia, este documento se descargó de Internet en un archivo pdf., de 2.89 MB (Sherzer s.f.).

Este documento contiene el “Nuevo Testamento”. Todo el documento está escrito en chatino y los libros que se encuentran son:

Evangelios: San Mateo, San Marcos, San Lucas y San Juan.

Hechos de los Apóstoles, Carta a los Romanos, 1^a Carta a los Corintios, 2^a Carta a los Corintios, Carta a los Gálatas, Carta a los Efesios, Carta a los Filipenses, Carta a los Colosenses, 1^a Carta a los Tesalonicenses, 2^a Carta a los Tesalonicenses, 1^a

Carta a los Timoteo, 2^a Carta a los Timoteo, Carta a Tito, Carta a Filemón, Carta a los Hebreos, Carta a Santiago, 1^a Carta de Pedro, 2^a Carta de Pedro, 1^a Carta de Juan, 2^a Carta de Juan, 3^a Carta de Juan, Carta de Judas y Apocalipsis.

- Lista de Palabras (Proyecto Chatino 2008), proporcionado por la M. en I. María de Jesús Estudillo Ayala.

Este documento contiene un listado de palabras en orden alfabético (números, animales, colores, algunos sustantivos y verbos) mostrando la palabra en chatino (seguido de una letra que representa el tono), en seguida la clasificación sintáctica y por último sus posibles traducciones en español. Ejemplo: *Ktyi*^(M) (Sustantivo) papel.

- Los pronombres personales y verbos (Sherzer s.f.), proporcionados por la M. en I. María de Jesús Estudillo Ayala.

Este documento muestra la conjugación del verbo en chatino, con diferentes pronombres y tiempos (tiempo pasado, futuro, progresivo y habitual), mostrando el pronombre (seguido de una letra que representa el tono), en seguida el verbo (seguido de una letra que representa el tono) y por último la traducción en español. A continuación se presenta un ejemplo del verbo “Ver”, con el pronombre “Yo” en tiempo pasado: *naä*^(M), *na’ ä*^(AM), yo ví.

- Enunciados, del libro llamado “Libro del Adulto” (INEA 1992), proporcionados por la M. en I. María de Jesús Estudillo Ayala.

Este libro contiene ejemplos, enunciados y palabras, todo está escrito en chatino.

De los documentos mencionados anteriormente y que están agregados al corpus pertenecientes a la variante de Tataltepec de Valdés, se tradujeron en el corpus los 100 primeros números, los pronombres personales, nombres de animales, colores, algunos verbos y sustantivos.

De la variante de San Juan Quiahije, se capturaron en archivos de texto los siguientes documentos:

- Los 100 primeros números y los pronombres personales, proporcionados vía electrónica por la lingüista Hilaria Cruz, mostrando el número y pronombre en español seguido de su traducción en chatino.
- Frases y conversaciones, obtenidas a través de una red social de la página Chatino Language Documentation (Cruz 2011).

De los documentos mencionados anteriormente y que están agregados al corpus pertenecientes a la variante de San Juan Quiahije, se tradujeron en el corpus, los 100 primeros números, los pronombres personales y frases.

De la variante de Santa Catarina Juquila, se capturaron en archivos de texto los siguientes documentos:

- Los 100 primeros números, los pronombres personales, listado de palabras, frases, narración (cómo se celebraban las bodas) y un diálogo (Pride & Pride 1997).

En el caso de las narraciones y diálogos, el texto se muestra en chatino (seguido de un número que representa el tono) y su traducción en chatino. Ejemplo;

A: Ndi'i²³ wa⁴ mlyi⁴³ ¿Rsu⁴⁵ nda⁴³ wa² ni⁴?

B: Su'we²³ 'a²

A: Buenas tardes comadre. ¿Cómo están ustedes?

B: Muy bien.

Y el resto del libro muestra la palabra en español y su traducción en chatino (seguido de un número que representa el tono), ejemplo: árbol, yka'².

De los documentos mencionados anteriormente y que están agregados al corpus pertenecientes a la variante de Santa Catarina Juquila, se tradujeron en el corpus, los 100 primeros números, los pronombres personales y algunos verbos.

En el siguiente anexo se describe detalladamente el funcionamiento de la aplicación que se desarrolló en este trabajo.

ANEXO C. MANUAL DE USUARIO

En este anexo se da una breve descripción de la aplicación creada en esta tesis. La aplicación requiere haber sido instalada previamente, tener instalado el manejador de base de datos MySQL y haber recuperado la base de datos (Anexo D).

En la figura C.1, se observa la pantalla PRINCIPAL de la aplicación.



Figura C.1. Pantalla PRINCIPAL.

Esta pantalla cuenta con los botones *Entrar* y *Salir*, además del nombre de la institución, los logotipos correspondientes a la institución y a la carrera y el nombre del tema correspondiente a este documento de tesis.

El botón *Entrar*, permite acceder al sistema de Corpus de Textos en chatino mostrando la pantalla CORPUS DE TEXTOS (Fig. C.2). El botón *Salir* automáticamente cierra la aplicación.

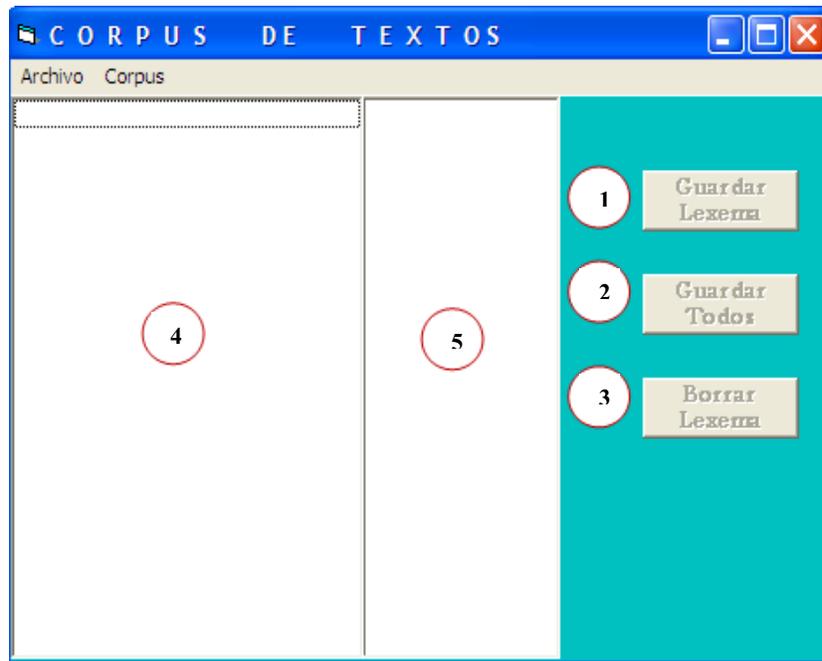


Figura C.2. Pantalla CORPUS DE TEXTOS.

La pantalla CORPUS DE TEXTOS (Fig. C.2), cuenta con los siguientes componentes: tres botones *Guardar Lexema* (componente 1), *Guardar Todos* (componente 2) y *Borrar Lexema* (componente 3); dos listas, en la primera básicamente se visualiza el texto completo tal y como está en el archivo de texto (componente 4) y en la otra lista se visualiza un listado de los lexemas del archivo de texto, sin mostrar repeticiones (componente 5).

El botón *Guardar Lexema*, almacena los lexemas uno por uno y muestra el mensaje de la figura C.3a, se debe seleccionar el lexema que se desee guardar antes de dar clic en el botón *Guardar Lexema*, en caso contrario muestra el mensaje de la figura C.3b. Si el lexema ya existe en el corpus muestra el mensaje de la figura C.3c.

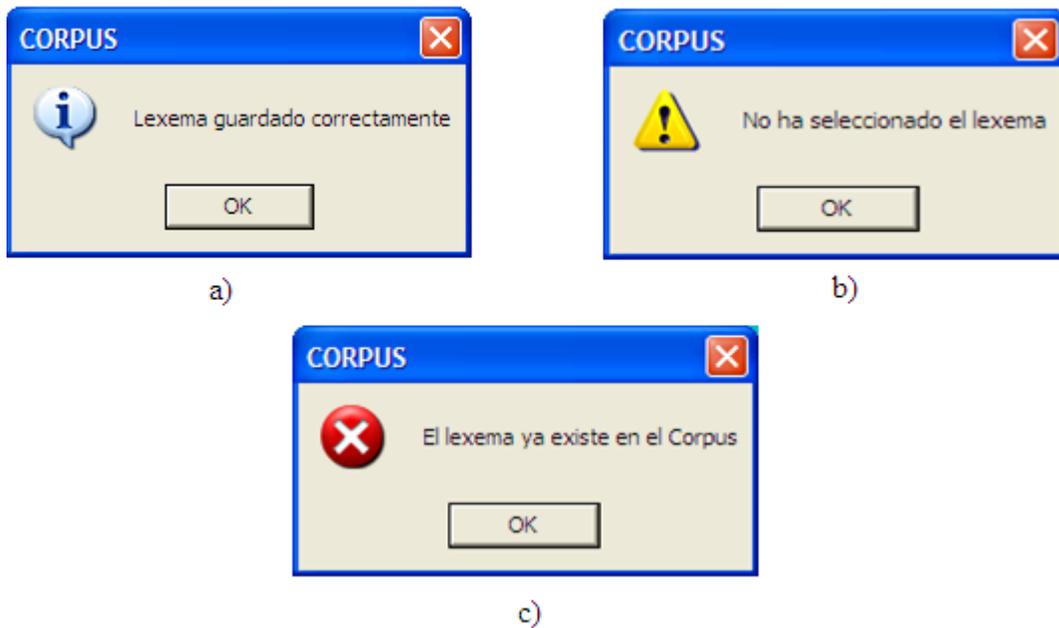


Figura C.3. Mensajes de la opción Guardar Lexema de la pantalla CORPUS DE TEXTOS. (a) No ha seleccionado lexema. (b) Lexema guardado. (c) Ya existe el lexema.

El botón Guardar Todos, almacena todo el listado de lexemas, mostrando un mensaje de cuántos lexemas han sido guardados y cuántos están repetidos (Fig. C.4), el numero 0 indica que no se guardó ningún lexema o que no hay lexemas repetidos.

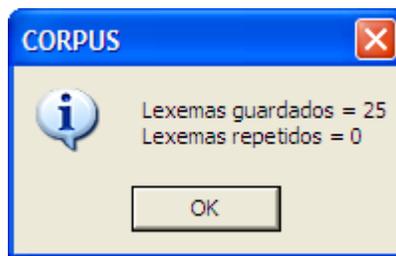


Figura C.4. Mensaje de la opción Guardar Todos de la pantalla CORPUS DE TEXTOS.

El botón Borrar lexema elimina el lexema seleccionado de la lista de lexemas. Si no selecciona el lexema antes de dar clic en el botón Borrar Lexema, muestra el mensaje de la figura C.5.

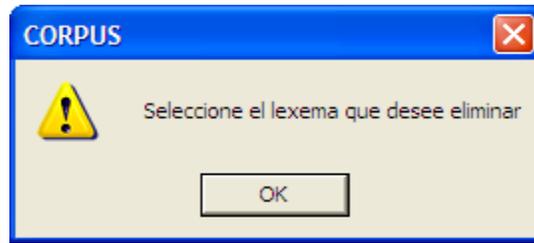


Figura C.5. Mensaje de error de la opción **Borrar Lexema** de la pantalla **CORPUS DE TEXTOS**.

La pantalla de la figura C.2, también cuenta con dos menús despegables, **Archivo** (Fig. C.6a) y **Corpus** (Fig. C.6b), cada uno con sus respectivas opciones.



Figura C.6. Menús despegables de la pantalla **CORPUS DE TEXTOS**. (a) Menú **Archivo**. (b) Menú **Corpus**.

La opción **Abrir** del menú **Archivo** muestra un cuadro de diálogo con archivos de texto (Fig. C.7).

La opción **Salir** del menú **Archivo** regresa a la pantalla **PRINCIPAL** (Fig. C.1).

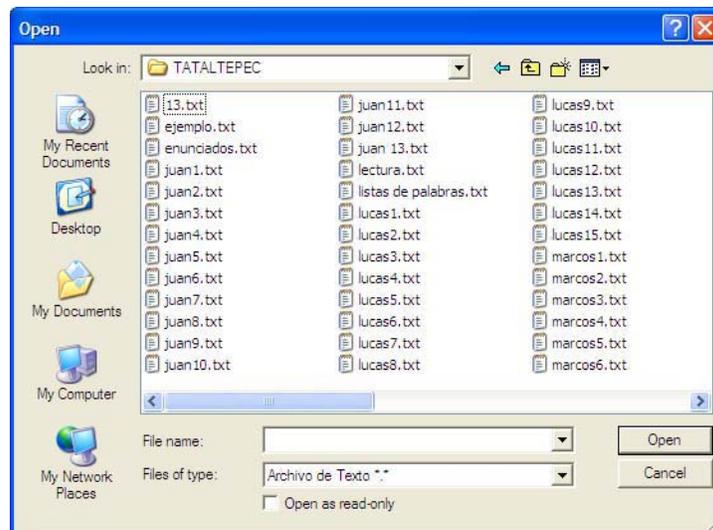


Figura C.7. **Abrir** archivos de textos.

La opción *Separar Lexema del menú Corpus* (C.6b) enlista los lexemas del texto en chatino, llenando el componente 5 de la figura C.2.

La opción *VER CORPUS* (Fig. C.6b) muestra la pantalla *VER CORPUS* (Fig. C.8).

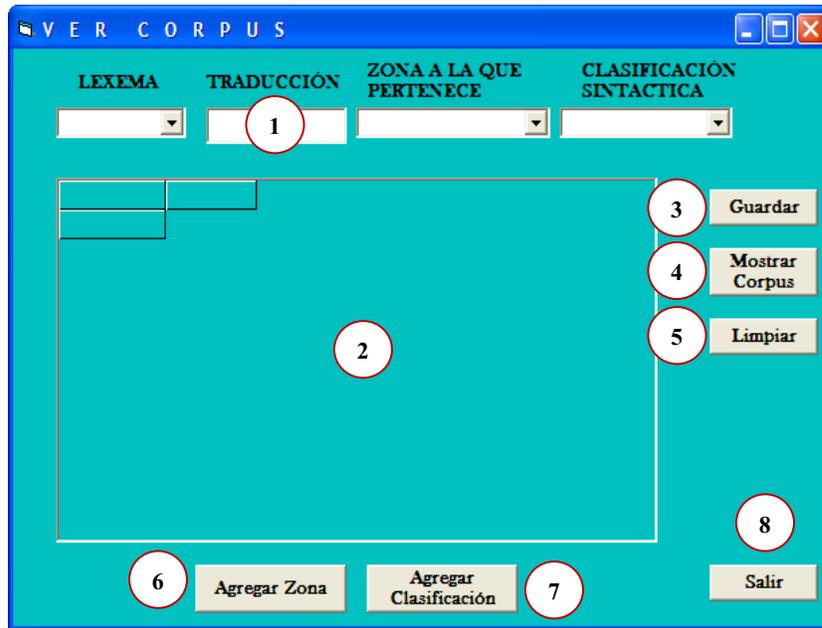


Figura C.8. Pantalla *VER CORPUS*.

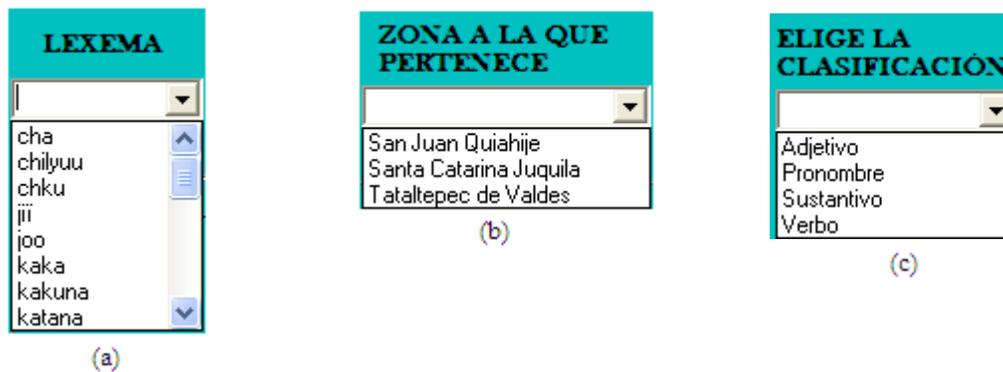


Figura C.9. Listas despegables. (a) Lexemas. (b) Zonas. (c) Clasificación Sintáctica.

La pantalla *VER CORPUS* (Fig. C.8) cuenta con los siguientes componentes: tres listas despegables, en donde el usuario tiene la opción de elegir lexema (Fig. C.9a), zona a la que pertenece el lexema (Fig. C.9b) y clasificación sintáctica que le corresponde (Fig. C.9c), también cuenta con un cuadro de texto (componente 1) en donde se debe escribir la traducción correspondiente del lexema, además de una tabla (componente 2) en la cual se

mostrarán los lexemas que ya cuenten con su respectiva traducción, zona y clasificación sintáctica dentro del corpus. Por último cuenta con seis botones: Guardar (componente 3), Mostrar Corpus (componente 4), Limpiar (componente 5), Agregar Zona (componente 6), Agregar Clasificación (componente 7) y Salir (componente 8).

El botón Guardar almacena en el corpus el lexema, traducción, zona y clasificación mostrando un mensaje de éxito (Fig. C.10a). En caso de que el usuario diera clic en el botón Guardar sin antes haber llenado los campos correspondientes, muestra el mensaje de la figura C.10b.

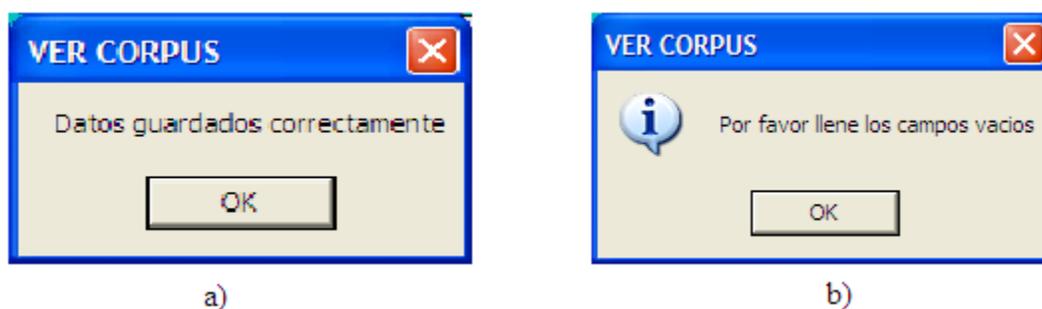


Figura C.10. Mensajes de la opción Guardar de la pantalla VER CORPUS. (a) Datos guardados. (b) Campos vacios.

El botón Mostrar Corpus despliega, en una tabla, una lista de lexemas con su traducción, zona y clasificación correspondiente en orden alfabético.

El botón Limpiar, restablece la pantalla VER CORPUS a su estado inicial.

El botón Agregar Zona abre la pantalla ZONA (Fig. C.11), en donde se puede agregar, modificar y eliminar las zonas existentes.

El botón Agregar Clasificación abre la pantalla CLASIFICACIÓN (Fig. C.15), en donde se puede agregar, modificar y eliminar las clasificaciones existentes.

El botón Salir regresa a la pantalla CORPUS DE TEXTOS.



Figura C.11. Pantalla ZONA.

La pantalla ZONA de la figura C.11, cuenta con los siguientes componentes: una lista desplegable (componente 1) que muestra las zonas dadas de alta en la base de datos, dos campos, uno para ingresar el nombre de la zona que se va a modificar o eliminar (componente 2) y la otra para ingresar el nombre de la zona que se va a agregar (componente 3), cuenta también con cinco botones Modificar (componente 4), Agregar Nueva Zona (componente 5), Eliminar (componente 6), Guardar (componente 7) y Salir (componente 8).

El botón Modificar cambia el nombre de la zona mostrando un mensaje de éxito (Fig. C.12a), siempre y cuando haya seleccionado la zona que desee modificar, en caso contrario muestra un mensaje de error (Fig. C.12b). Antes de presionar el botón, se debieron hacer los cambios en el componente 2.

El botón Agregar Nueva Zona activa el campo (componente 3) para ingresar el nombre de la zona que se va a agregar.

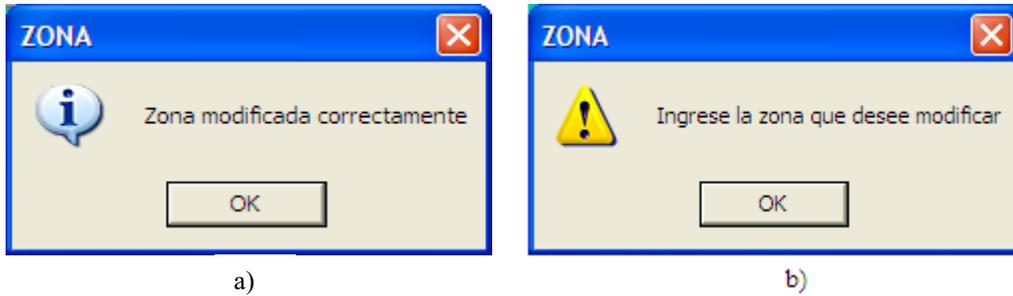


Figura C.12. Mensajes de la opción Modificar de la pantalla ZONA. (a) Mensaje de éxito. (b) Mensaje de error.

El botón Eliminar borra la zona que haya elegido, mostrando el mensaje de la figura C.13a. En caso de que el usuario diera clic en el botón Eliminar sin antes haber llenado los campos correspondientes, muestra el mensaje de la figura C.13b. Si la zona tiene datos relacionados muestra el mensaje de la figura C.13c.

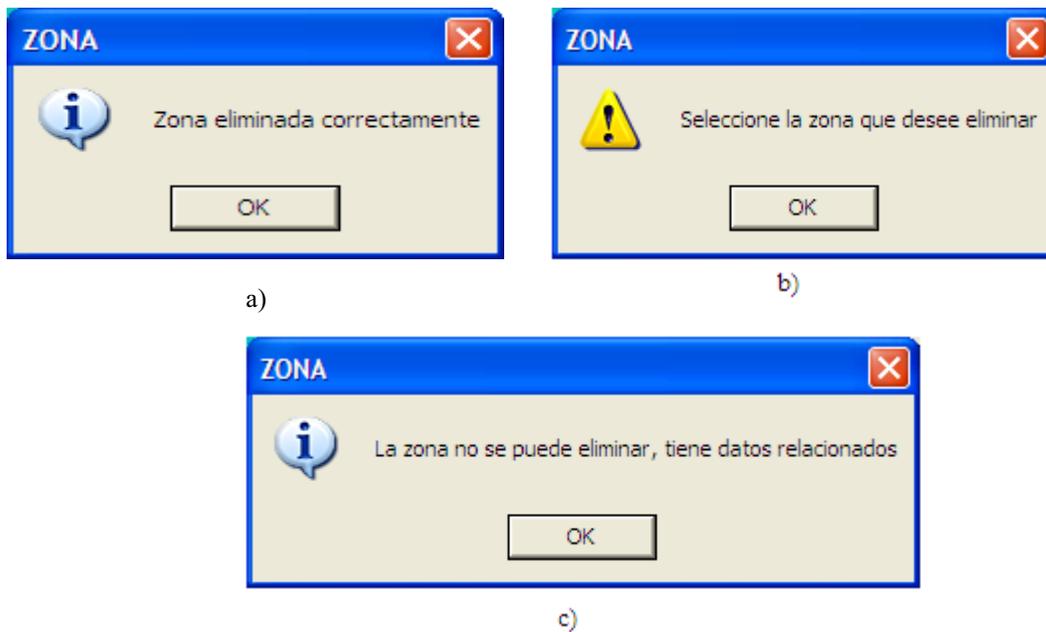


Figura C.13. Mensajes de la opción Eliminar de la pantalla ZONA. (a) Zona eliminada. (b) Seleccione la zona. (c) No se puede eliminar.

El botón Guardar almacena la zona que desee agregar mostrando un mensaje como el que se observa en la figura C.14a. En caso de que el usuario diera clic en el botón

Guardar sin antes haber escrito el nombre de la zona, muestra el mensaje de la figura C.14b. Si la zona ya existe muestra el mensaje de la figura C.14c.

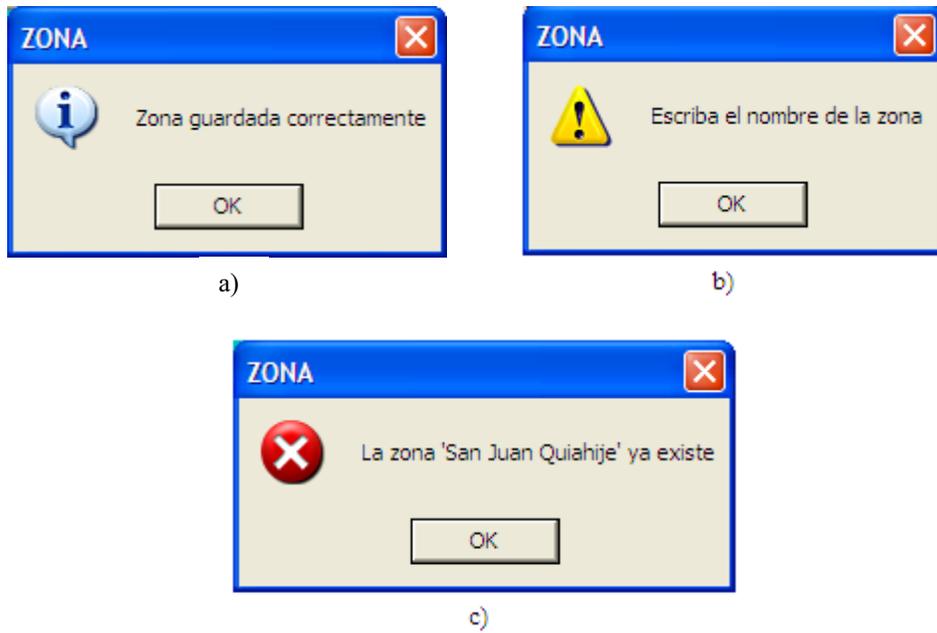


Figura C.14. Mensajes de la opción Guardar de la pantalla ZONA. (a) Zona guardada. (b) Mensaje de error. (c) Zona existente.

El botón Salir regresa a la pantalla VER CORPUS .

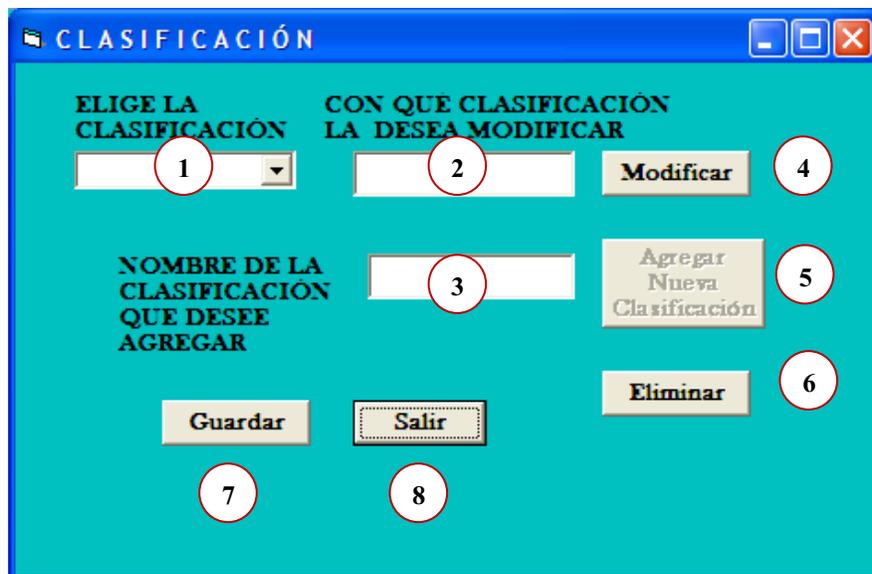


Figura C.15. Pantalla CLASIFICACIÓN .

La pantalla CLASIFICACIÓN de la figura C.15, cuenta con los siguientes componentes: una lista despegable (componente 1) que muestra las clasificaciones dadas de alta en la base de datos, dos campos, uno para ingresar el nombre de la clasificación que se va a modificar o eliminar (componente 2) y la otra para ingresar el nombre de la clasificación que se va a agregar (componente 3), cuenta también con cinco botones Modificar (componente 4), Agregar Nueva Clasificación (componente 5), Eliminar (componente 6), Guardar (componente 7) y Salir (componente 8).

El botón Modificar cambia el nombre de la clasificación mostrando un mensaje de éxito (Fig. C.16a), siempre y cuando haya seleccionado la clasificación que desee modificar, en caso contrario muestra un mensaje de error (Fig. C.16b). Antes de presionar el botón, se debieron hacer los cambios en el componente 2.

El botón Agregar Nueva Clasificación activa el campo (componente 3) para ingresar el nombre de la clasificación que se va a agregar.

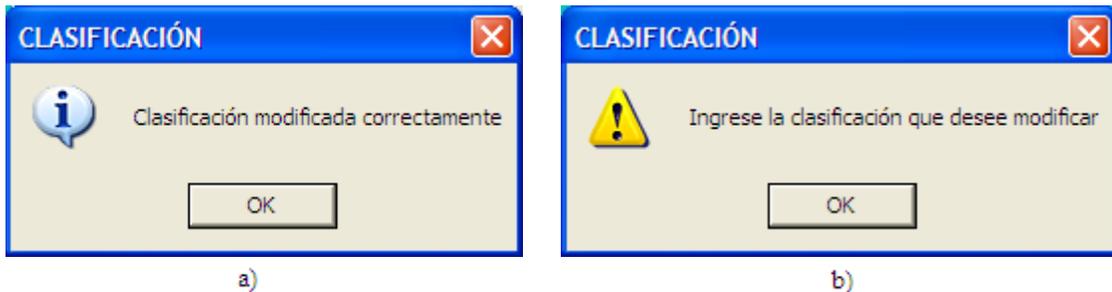


Figura C.16. Mensajes de la opción Modificar de la pantalla CLASIFICACIÓN. (a) Mensaje de éxito. (b) Mensaje de error.

El botón Eliminar borra la clasificación que haya elegido, mostrando el mensaje de la figura C.17a. En caso de que el usuario diera clic en el botón Eliminar sin antes haber llenado los campos correspondientes, muestra el mensaje de la figura C.17b. Si la clasificación tiene datos relacionados muestra el mensaje de la figura C.17c.

El botón Guardar almacena la clasificación que desee agregar mostrando un mensaje como el que se observa en la figura C.18a. En caso de que el usuario diera clic en el botón Guardar sin antes haber escrito el nombre de la clasificación, muestra el mensaje de la figura C.18b. Si la clasificación ya existe muestra el mensaje de la figura C.18c.

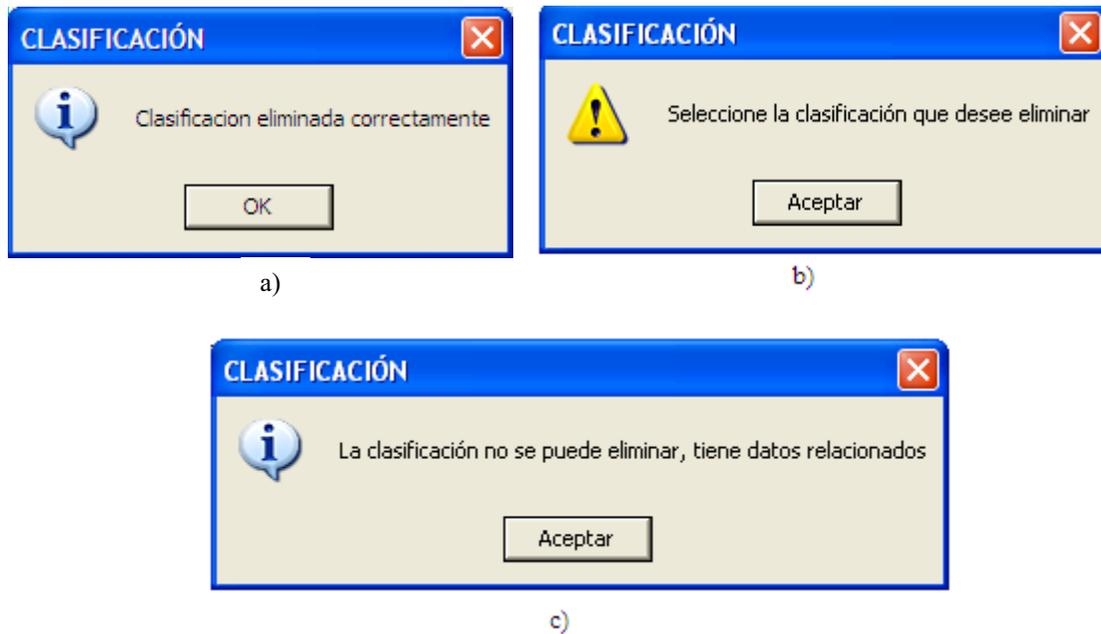


Figura C.17. Mensajes de la opción Eliminar de la pantalla CLASIFICACIÓN. (a) Clasificación eliminada. (b) Seleccione la clasificación. (c) No se puede eliminar.

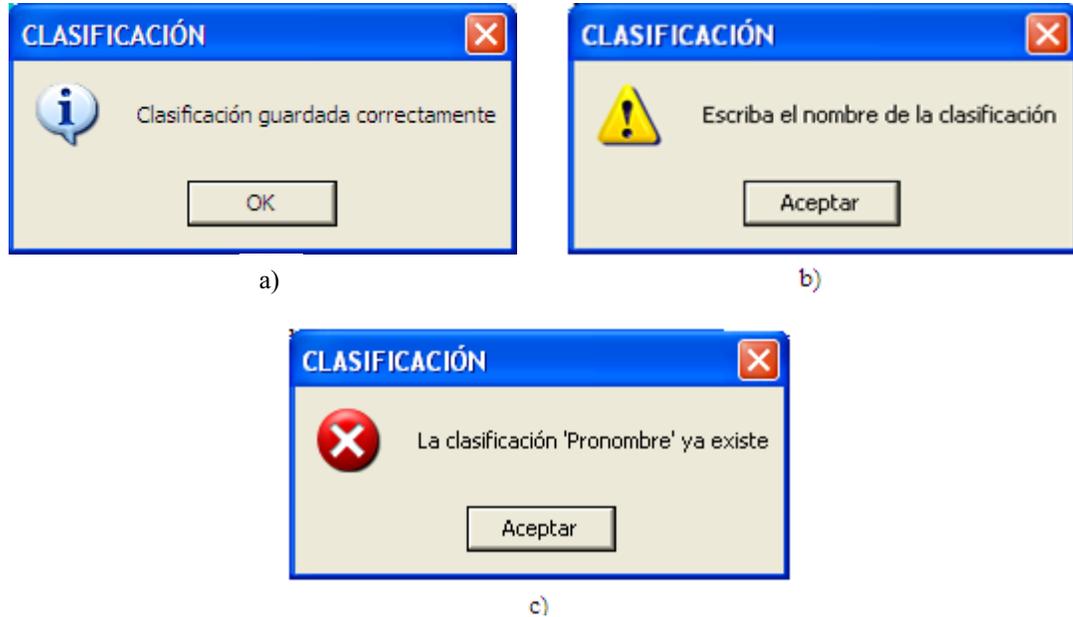


Figura C.18. Mensajes de la opción Guardar de la pantalla CLASIFICACIÓN. (a) Clasificación guardada. (b) Mensaje de error. (c) Clasificación existente.

El botón Salir regresa a la pantalla VER CORPUS.

Después de presentar una descripción sobre el funcionamiento de la aplicación desarrollada en este documento, en el siguiente anexo se detalla el proceso de la instalación y restauración de la base de datos.

ANEXO D. INSTALACIÓN Y RESTAURACIÓN DE LA BASE DE DATOS

En este anexo se detalla el proceso de instalación del sistema gestor de base de datos MySQL 5.0 en un entorno Windows, así como los pasos a seguir para el proceso de la restauración de la base de datos.

Instalación de la base de datos

La instalación de MySQL es muy intuitiva, del tipo “siguiente”.

Primero se comienza instalando el conector ODBC para MySQL, ejecutando el archivo “mysql-connector-odbc-5.1.5-win32” que se muestra en la figura E.5. Se ejecuta el archivo y aparece el cuadro de diálogo de bienvenida que se muestra en la figura D.1, y se pulsa el botón *Next*.

Aparecerá una pantalla donde se debe seleccionar el tipo de instalación, se puede elegir entre una instalación típica (*Typical*), completa (*Complete*) o a la medida

(Custom). Para este caso, ya que no se necesitan características especiales, se selecciona Typical y se da clic en el botón Next (Fig. D.2).

Antes de iniciarse la instalación mostrará una pantalla (Fig. D.3) con el resumen de la opción seleccionada. Dar clic sobre el botón Install para iniciar el proceso de instalación.

Se finaliza la instalación del conector ODBC mostrando la figura D.4.

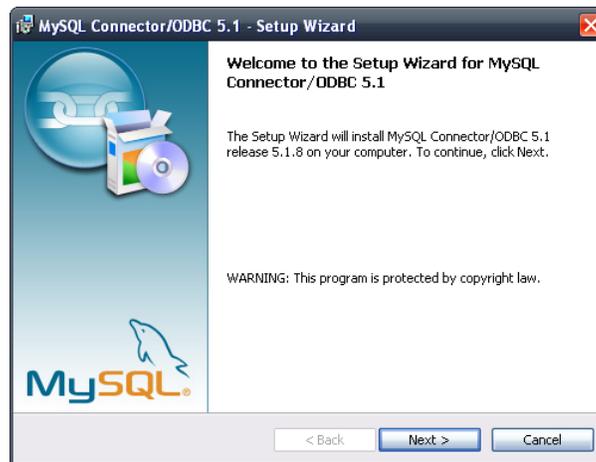


Figura D.1. Pantalla de bienvenida del conector ODBC.



Figura D.2. Tipo de instalación del conector ODBC.

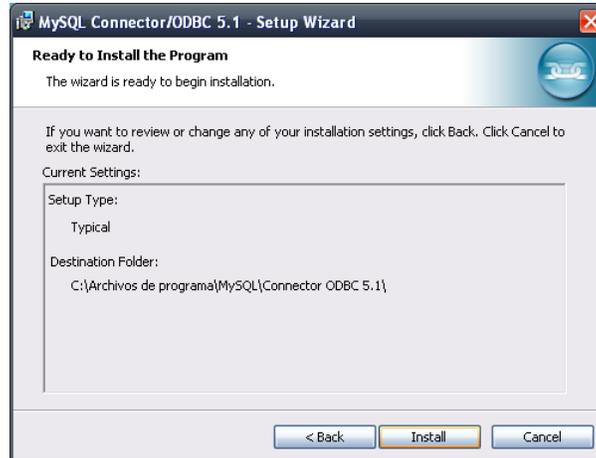


Figura D.3. Resumen del tipo de instalación del conector ODBC.

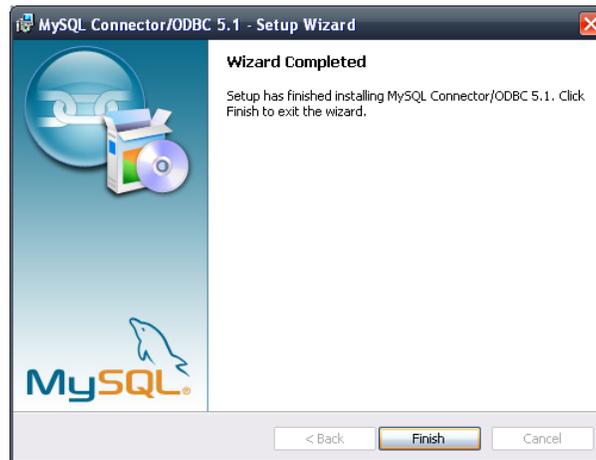


Figura D.4. Finalizar la instalación del conector ODBC.

Enseguida se instala el archivo “mysql-essential-5.0.51a-win32”, que es el servidor de MYSQL, este archivo se muestra en la figura E.5. Se da doble clic para que aparezca el cuadro de diálogo de bienvenida y dar clic sobre el botón **Next** (Fig. D.5).

Aparecerá una pantalla donde se debe seleccionar el tipo de instalación, se puede elegir entre una instalación típica (**Typical**), completa (**Complete**) o a la medida (**Custom**). Para este caso, ya que no se necesitan características especiales, se selecciona **Typical** y se da clic en el botón **Next** (Fig. D.6).

Antes de iniciarse la instalación mostrará una pantalla (Fig. D.7) con el resumen de la opción seleccionada. Dar clic sobre el botón `Install` para iniciar el proceso de instalación.

Empieza la instalación mostrando la pantalla de la figura D.8.

Después de instalar el servidor, muestra las características y publicidad del producto (Figs. D.9 y D.10). Posteriormente muestra una pantalla que preguntará si se quiere configurar MySQL en este momento. Se marca la opción `Configure the MySQL Server now` para empezar a configurar el servidor MySQL y dar clic en `Finish` (Fig. D.11).



Figura D.5. Pantalla de bienvenida del servidor MySQL.



Figura D.6. Tipo de instalación del servidor MySQL.



Figura D.7. Resumen del tipo de instalación del servidor MySQL.

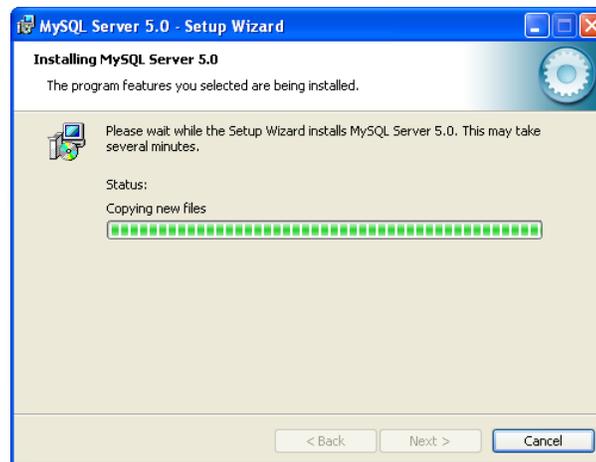


Figura D.8. Instalación del servidor MySQL.

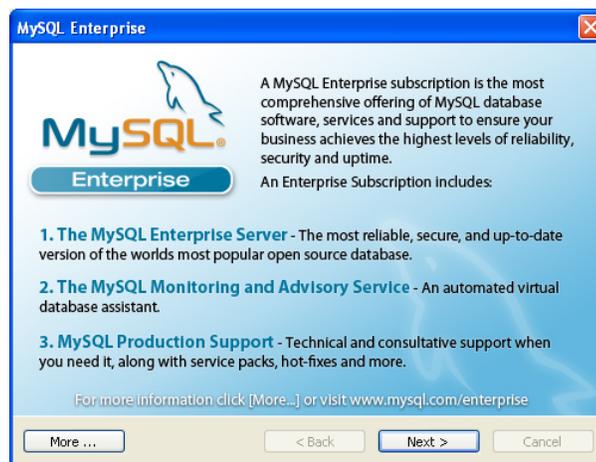


Figura D.9. Primera pantalla de publicidad del servidor MySQL.

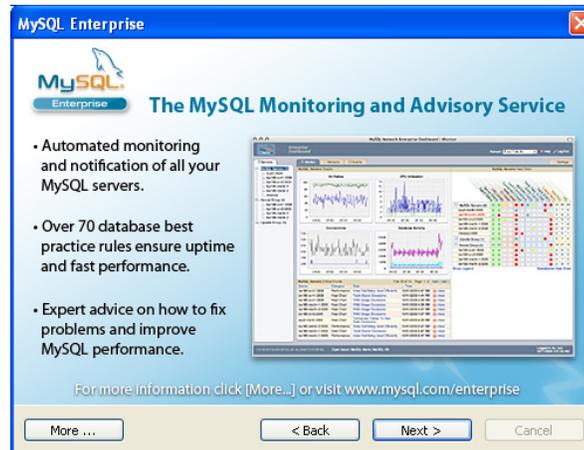


Figura D.10. Segunda pantalla de publicidad del servidor MySQL.



Figura D.11. Fin de la instalación del servidor MySQL.

Al comenzar la configuración del servidor MySQL muestra la pantalla de bienvenida del servidor MySQL, como se muestra en la figura D.12, dar clic en Next.

En la siguiente pantalla (Fig. D.13) se debe seleccionar el tipo de configuración deseada: detallada (Detailed Configuration) o estándar (Standard Configuration). Se selecciona la opción de configuración detallada y se da clic en Next. De esta forma se podrá configurar más opciones de MySQL utilizando el asistente. Si se marca la configuración estándar, el asistente pedirá menos información pero habría que configurar algunas opciones manualmente.

La siguiente pantalla (Fig. D.14) muestra tres opciones. Dependiendo del uso que se le quiera dar al equipo en el que se instala, se marca una de las tres: Developer

Machine, se marca esta opción si en el equipo donde se ha instalado el servidor MySQL se utiliza también para otras aplicaciones. El servidor MySQL utilizará la memoria mínima necesaria. Server Machine, se marca esta opción si se va a utilizar el equipo para algunas aplicaciones (no demasiadas). Con esta opción el servidor MySQL utilizará un nivel medio de memoria y Dedicated MySQL Server Machine, se marca esta opción sólo si se quisiera utilizar el equipo como un servidor dedicado exclusivamente a MySQL. Con esta opción el servidor MySQL utilizará el máximo de memoria disponible. Se obtendrá un rendimiento elevado pero el equipo sólo servirá para MySQL. Dado que no se requiere un servidor dedicado, se selecciona la opción Developer Machine y se da clic en Next.

Dependiendo del uso que se le quiera dar a la base de datos se marca una de las tres opciones que se muestran en la figura D.15, y se da clic en Next. Normalmente se marca Multifunctional Database salvo que se quiera utilizar MySQL como base de datos para transacciones de otra base de datos MySQL.

En la pantalla de la figura D.16, se selecciona la unidad y la carpeta donde se desee guardar los archivos de datos de las tablas InnoDB, un tipo de tabla con soporte para la integridad referencial. Para este ejemplo, se dejan las opciones por defecto y se da clic en Next.

En la pantalla de la figura D.17, se selecciona el número aproximado de conexiones concurrentes (varios clientes conectados a la vez) que tendrá el servidor de MySQL. La primera opción asume unas 20, la segunda unas 500 y la tercera permite especificarlas manualmente. Se elige Decision Support (DSS)/OLAP (primera opción) y se da clic en Next.

En la siguiente pantalla (Fig. D.18) se deja marcada la opción Enable TCP/IP Networking, si se desea que los clientes se puedan conectar mediante TCP/IP al equipo servidor de MySQL. Se puede cambiar el puerto por el que lo harán. Por defecto se suele dejar 3306. También se deja seleccionada la opción Enable Strict Mode, si se quiere que el servidor se comporte como un servidor de base de datos tradicional, que será lo normal, y pulsar el botón Next.

La pantalla que muestra a continuación (Fig. D.19) permite elegir el juego de caracteres que se quiera utilizar. Se deja la opción por default y se da clic en Next.

En la siguiente pantalla (Fig. D.20) se especifica el tipo de arranque del servidor MySQL. Si se selecciona la primera opción, `Install As Windows Service`, el programa de instalación creará un servicio que será el encargado de ejecutar el servidor MySQL. Esta pantalla también permite especificar el nombre del servicio y si se desea que arranque automáticamente al iniciar el sistema (`Launch the MySQL Server automatically`). La segunda opción, `Include Bin Directory in Windows PATH`, añadirá las variables de entorno necesarias para la ejecución de los archivos necesarios para iniciar MySQL. La opción recomendada es `Install As Windows Service`. Una vez elegida una opción, pulsar el botón `Next`.

En la siguiente pantalla (Fig. D.21) se introduce la contraseña para el usuario administrador (`root`) y se da clic en el botón `Next`. Se marca la opción `Enable root access from remote machines` si se quiere que se pueda acceder como administrador desde otros equipos. Es importante establecer una contraseña para el usuario `root`, ya que si no se hace, cualquier persona podría acceder al sistema y hacer lo que quisiera con los datos.

Para terminar con la configuración de MySQL dar clic en el botón `Execute` como se muestra en la figura D.22, para que el asistente de configuración de MySQL aplique todas las opciones indicadas anteriormente y luego dar clic en `Finish`, para finalizar la configuración de MySQL (Fig. D.23).



Figura D.12. Pantalla de bienvenida del servidor MySQL.



Figura D.13. Tipo de configuración del servidor MySQL.



Figura D.14. Primera pantalla del uso del servidor MySQL.



Figura D.15. Segunda pantalla del uso del servidor MySQL.



Figura D.16. Elegir unidad para guardar los archivos del servidor MySQL.

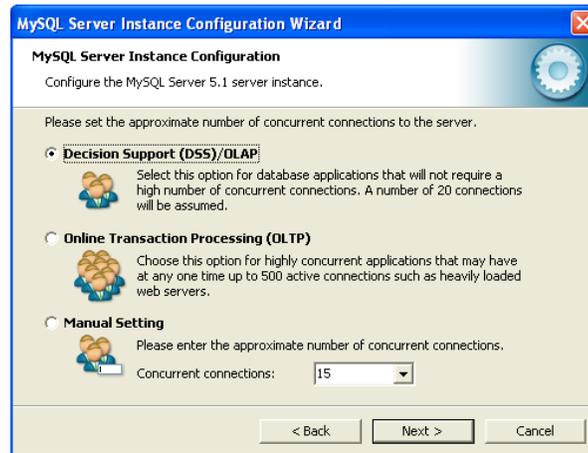


Figura D.17. Número de conexiones del servidor MySQL.

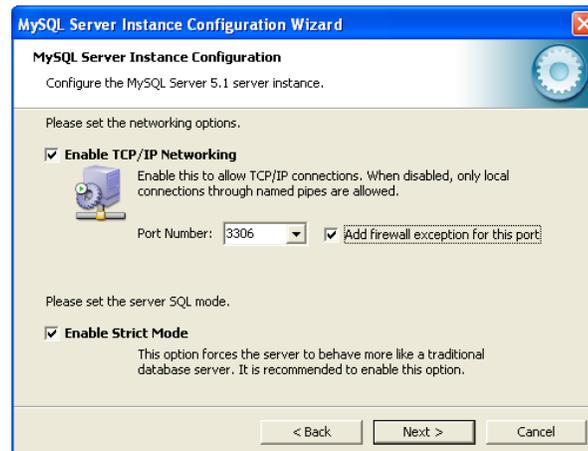


Figura D.18. Número de puerto del servidor MySQL.

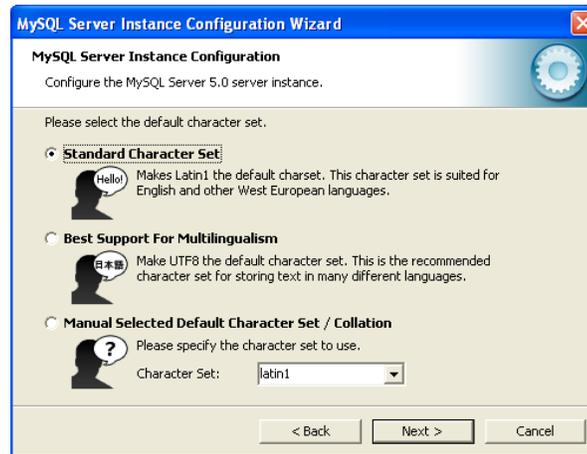


Figura D.19. Elegir el juego de caracteres del servidor MySQL.



Figura D.20. Tipo de arranque del servidor MySQL.



Figura D.21. Ingresar contraseña para el servidor MySQL.

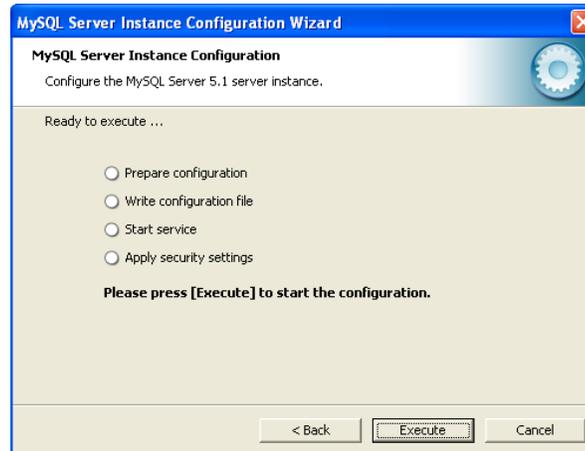


Figura D.22. Opción Execute del servidor MySQL.

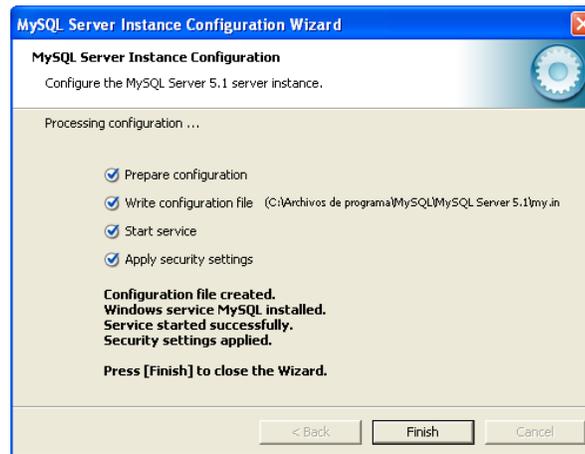


Figura D.23. Finalizar la configuración del servidor MySQL.

Por último se instala el archivo “mysql-gui-tools-5.0-r16-win32”, para que muestre la interfaz gráfica de MySQL, este archivo se muestra en la figura E.5. Se da doble clic para que aparezca el cuadro de diálogo de bienvenida y dar clic sobre el botón Next (Fig. D.24).



Figura D.24. Pantalla de bienvenida de la instalación entorno gráfico de MySQL.

Enseguida aparece una pantalla (Fig. D.25), mostrando la licencia de MySQL, una vez leída la licencia y estar de acuerdo, seleccionar la opción `I accept the terms in the license agreement` y pulsar el botón `Next`.

En la figura D.26, muestra la pantalla de directorio de instalación, da la opción de cambiar el directorio que el programa de instalación propone por defecto, si no se desea cambiar, dar clic en `Next`.

En la siguiente pantalla (Fig. D.27), se elige el tipo de instalación, se puede elegir entre una instalación completa (`Complete`) o a la medida (`Custom`). Se selecciona la opción `Complete` y se da clic en el botón `Next`.

Antes de iniciarse la instalación mostrará una pantalla (Fig. D.28) con el resumen de la opción seleccionada. Dar clic sobre el botón `Install` para iniciar el proceso de instalación (Fig. D.29).

Con la pantalla que aparece en la figura D.30, se termina de instalar el entorno gráfico de MySQL.

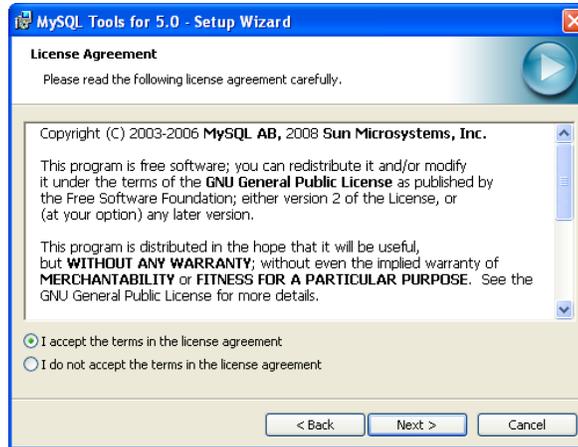


Figura D.25. Licencia del entorno gráfico de MySQL.

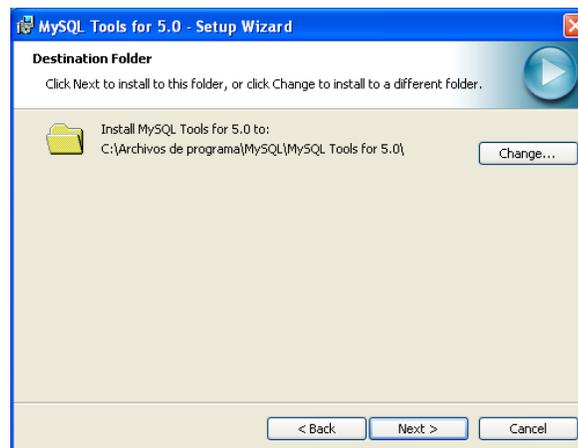


Figura D.26. Directorio de instalación del entorno gráfico de MySQL.



Figura D.27. Tipo de instalación del entorno gráfico de MySQL.



Figura D.28. Resumen del tipo de instalación del entorno gráfico de MySQL.

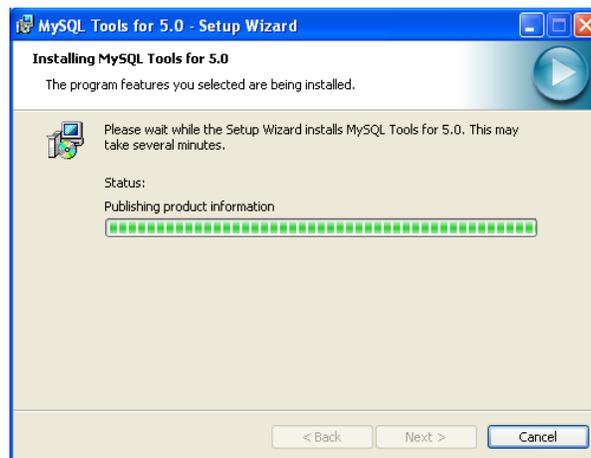


Figura D.29. Instalación del entorno gráfico de MySQL.



Figura D.30. Fin de la instalación del entorno gráfico de MySQL.

Restauración de la base de datos

A continuación se describen los pasos para restaurar la base de datos realizada en este documento de tesis en MySQL 5.0.

Lo primero que se tiene que hacer es entrar a la pantalla principal de MySQL desde la barra de tareas → Todos los programas → MySQL → MySQL Administrator, hasta que aparezca una pantalla como la que muestra en la figura D.31, en donde se escribe el password que se estableció en la pantalla de la figura D.21. Una vez ingresado el password muestra una pantalla en donde se observa la información relativa al servidor de base de datos al que se ha conectado, así como las características de la máquina en la que se está ejecutando el programa (Fig. D.32).

De la pantalla principal, al elegir la opción `Restore`, aparece una nueva pantalla como la que se muestra en la figura D.33. De esta pantalla se elige el componente 1, que contiene la opción `Open Backup File`.

A continuación muestra el cuadro de diálogo (Fig. D.34), donde se elige el nombre del respaldo de la base de datos. Se elige el archivo “Corpus en Lengua Chatina” de la ruta del CD, dentro de la carpeta `Respaldo de la BD`, y se da clic en `Open`. Una vez elegido el archivo, se da clic en el componente 2 de la figura D.33, y quedará restaurada la base de datos, mostrando el mensaje de la figura D.35.

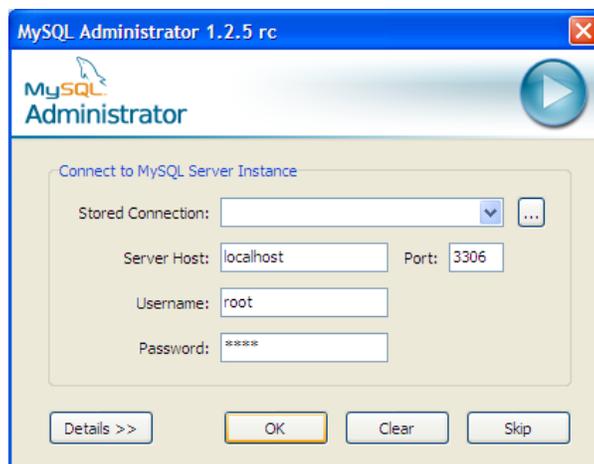


Figura D.31. Ingresar password para entrar a MySQL.

ANEXO D. INSTALACIÓN Y RESTAURACIÓN DE LA BASE DE DATOS

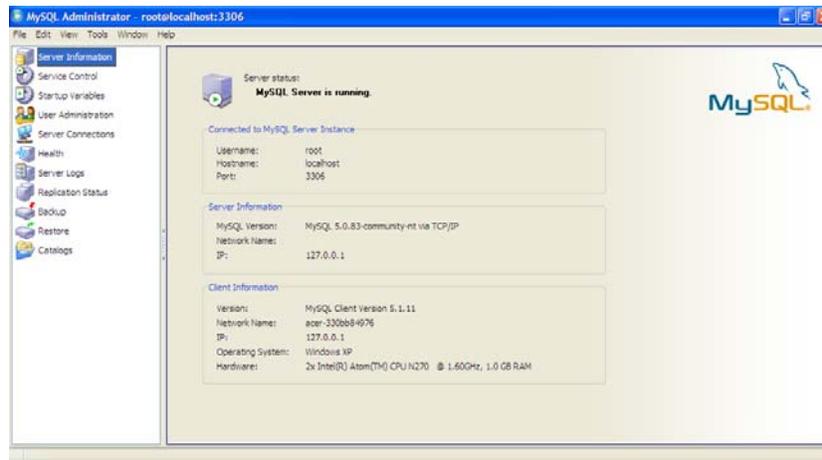


Figura D.32. Pantalla Principal.

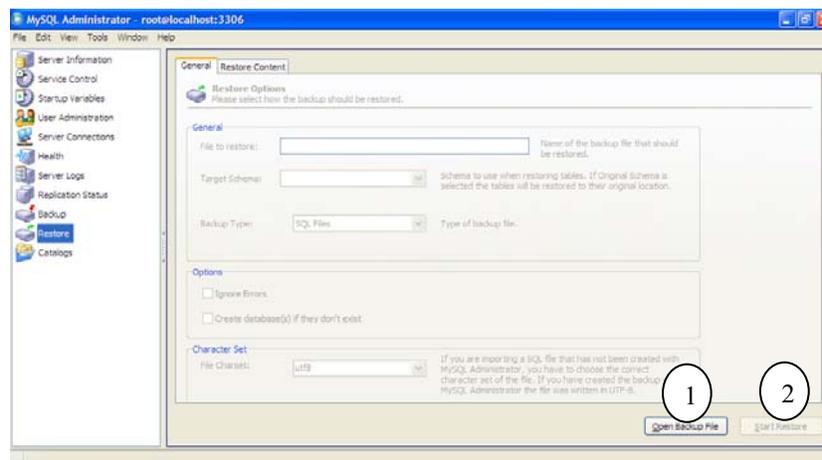


Figura D.33. Pantalla opción Restore.

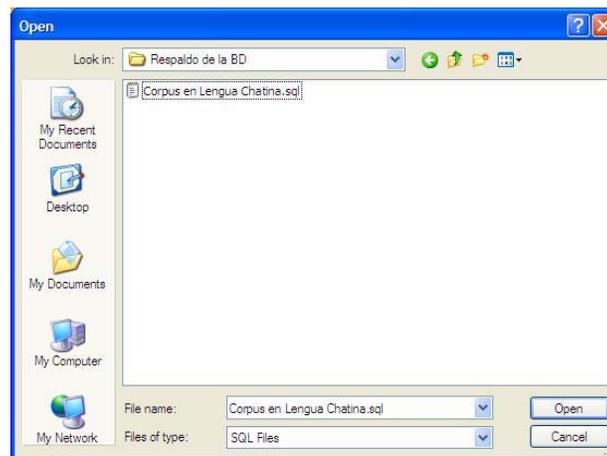


Figura D.34. Cuadro de diálogo abrir archivo sql.

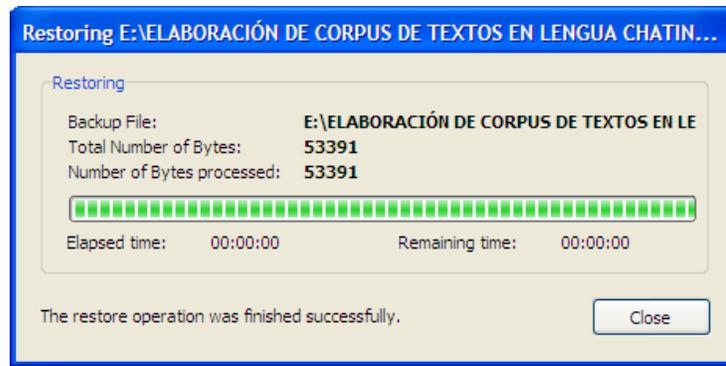


Figura D.35. Restauración de la base de datos.

Una vez presentados todos los pasos para la instalación y restauración de la base de datos, por último en el anexo E se muestra la estructura del CD.

ANEXO E. CONTENIDO DEL CD

Este anexo muestra la estructura de los directorios que contiene el CD que se anexa a este proyecto.

El contenido del CD de este trabajo de tesis está compuesto de seis carpetas, las cuales se muestran en la figura E.1. A continuación se da una breve explicación de cada una de ellas.

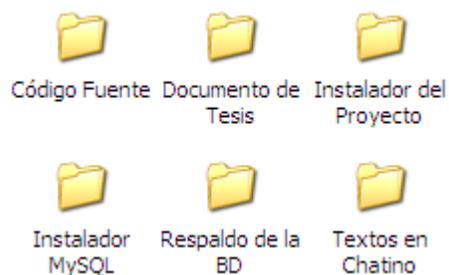


Figura E.1. Contenido del CD.

Carpeta Código Fuente

Esta carpeta está compuesta por diferentes archivos que componen la aplicación “Elaboración de corpus de textos en lengua chatina” (Fig. E.2). Cabe recordar que la aplicación fue realizada en el lenguaje Microsoft VB 6.0. El proyecto se abre ejecutando el archivo Corpus.vbp.

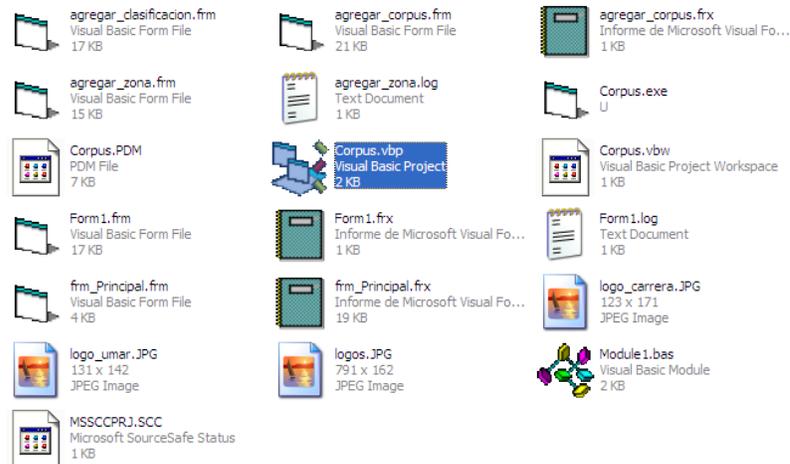


Figura E.2. Archivos fuente de la aplicación.

Carpeta Documento de Tesis

Esta carpeta contiene en formato digital el presente documento de tesis denominado “Elaboración de corpus de textos en lengua chatina” (Fig. E.3).

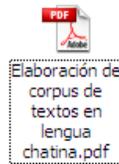


Figura E.3. Documento de tesis.

Carpeta Instalador del Proyecto

Dentro de esta carpeta se encuentran diferentes archivos (Fig. E.4), setup.exe es el archivo de instalación. Para instalar correctamente el proyecto se deben seguir los pasos que muestra el propio instalador.



Figura E.4. Archivos de instalación.

Carpeta Instalador MySQL

Dentro de esta carpeta se encuentran los archivos necesarios para la instalación del manejador de base de datos MySQL (Fig. E.5).

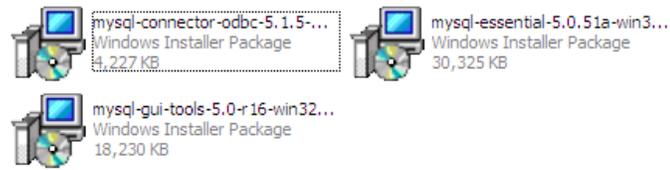


Figura E.5. Instalador MySQL.

El archivo “mysql-connector-odbc-5.1.5-win32” es el instalador del controlador ODBC para MySQL, “mysql-essential-5.0.51a-win32” es el servidor de MySQL y “mysql-gui-tools-5.0-r16-win32” es para que muestre la interfaz gráfica de MySQL.

Cada uno se debe instalar por separado, para más detalle consultar el Anexo D.

Carpeta Respaldo de la BD

En esta carpeta se encuentra el respaldo de la base de datos del corpus en lengua chatina, se abre el archivo “Corpus en Lengua Chatina.sql” (Fig. E.6).

Para restaurar la base de datos consultar el Anexo D.

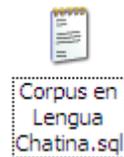


Figura E.6. Respaldo de la base de datos.

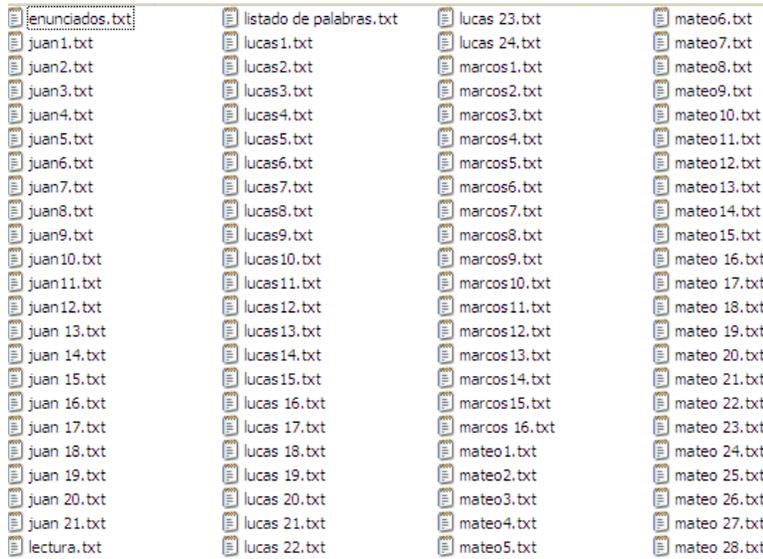
Carpeta Textos en Chatino

Dentro de esta carpeta se encuentran tres carpetas llamadas: Variante de San Juan Quiahije, Variante de Santa Catarina Juquila y Variante de Tataltepec de Valdés, cada carpeta contiene archivos escritos en chatino de la zona correspondiente.



(a)

(b)



(c)

Figura E.7. Textos escritos en chatino. (a) Variante de San Juan Quiahije. (b) Variante de Santa Catarina Juquila. (c) Variante de Tataltepec de Valdés.

REFERENCIAS

Alberto-Robledo, M 1994, *Pueblos Indígenas de México. Chatinos*, Ed. INI, Oaxaca.

Berumen-Barbosa, ME 2003, *Geografía económica de Oaxaca*, Biblioteca Virtual eumed.net, Universidad de Málaga, visitada el 15 de octubre de 2010, <<http://www.eumed.net/cursecon/libreria/mebb/1.htm>>.

Carrizosa, O s.f., *Go-Oaxaca | Grupos étnicos de Oaxaca | Chatinos*, Go-Oaxaca, visitado el 18 de noviembre de 2009, <www.go-oaxaca.com/overview/chatinos.html>.

CIESAS 2008, *Atlas de las Lenguas Indígenas de México*, Centro de Investigación y Estudios superiores en Antropología Social –Unidad Pacífico Sur, visitada el 20 de abril de 2010, <<http://pacificosur.ciesas.edu.mx/anexos/mapas/atlas/chatino.gif>>.

Cruz, H 2010, *Chatino Language Documentation*, Universidad de Texas en Austin, visitada el 20 de junio de 2011, <<http://www.facebook.com/groups/102434476471103/?id=238000399581176#!/groups/102434476471103/>>.

Definición ABC 2007, *Definición de aplicación »concepto en definición ABC*, visitada el 25 de septiembre de 2011, <<http://www.definicionabc.com/tecnologia/aplicacion.php>>.

Delgado-Orta, JF 2005, ‘Representación de una consulta en lenguaje natural a través de un grafo’, tesis de licenciatura, Instituto Tecnológico de Ciudad Madero, Tamaulipas.

Educ.ar 2006, *Lingüística de corpus: Lengua*, Ministerio de Educación de la Nación Argentina, visitada el 03 de marzo de 2010, <http://aportes.educ.ar/lengua/nucleo-teorico/influencia-de-las-tic/tecnologias-de-la-informacion-y-la-comunicacion-tic-y-lingueistica/lingueistica_de_corpus.php>.

Encarta 2008, *Corpus lingüístico*, Microsoft Student con Encarta Premium 2009, Microsoft Corporation.

Guzmán-Ventura, E 2009, *Juquila Memorias de mi pueblo*, 1ª Edición, Productos Gráficos El Castor, Oaxaca.

Inafed 2009, *Oaxaca – San Juan Quiahije*, Enciclopedia de los Municipios de México, Estado de Oaxaca, Instituto Nacional para el federalismo y el desarrollo municipal, visitada el 02 de enero de 2011, <<http://www.inafed.gob.mx/work/templates/enciclo/oaxaca/municipios/20213a.htm>>.

Inafed 2009, *Oaxaca – Santa Catarina Juquila*, Enciclopedia de los Municipios de México, Estado de Oaxaca, Instituto Nacional para el federalismo y el desarrollo municipal, visitada el 02 de enero de 2011, <<http://www.inafed.gob.mx/work/templates/enciclo/oaxaca/municipios/20364a.htm>>.

Inafed 2009, *Oaxaca – Tataltepec de Valdés*, Enciclopedia de los Municipios de México, Estado de Oaxaca, Instituto Nacional para el federalismo y el desarrollo municipal visitada el 02 de enero de 2011, <<http://www.inafed.gob.mx/work/templates/enciclo/oaxaca/municipios/20543a.htm>>.

Inafed 2009, *Regionalización – Oaxaca*, Enciclopedia de los Municipios de México, Estado de Oaxaca, Instituto Nacional para el federalismo y el desarrollo municipal, visitada 02 de mayo 2011, <<http://www.inafed.gob.mx/work/templates/enciclo/oaxaca/regi.htm>>.

INEA 1992, *K'tyi cha'kña ji'i ngu kusu jlo'o*, Libro del Adulto, Instituto Nacional para la Educación de los Adultos, Oaxaca.

Johansson, S 1978, *Manual of information to accompany the lancaster-oslo/bergen corpus of british english, for use with digital computers*, Department of English, University of

Oslo, visitada el 05 de marzo de 2010, <<http://khnt.hit.uib.no/icame/manuals/lob/index.htm>>.

La revista informática 2006, *Lenguajes de programación orientado a eventos*, La revista informática, visitada el 21 de octubre de 2011, <<http://www.larevistainformatica.com/Lenguajes-programacion-orientado-eventos.htm>>.

Martínez-Mena, A 2007, *Procesamiento del Lenguaje Natural para Recuperación de Información*, Procesadores de lenguaje natural, visitada el 20 de septiembre de 2001, <<http://pln-ri-hmm.orgfree.com/pln.html>>.

Mena-Mendoza, G 2005, *RAD: Desarrollo Rápido de Aplicaciones*, visitada el 21 de octubre de 2011, <<http://mena.com.mx/gonzalo/maestria/ingsoft/presenta/rad/>>.

Microsoft 2010, *Introducción a Visual Basic*, Microsoft Corporation, visitada el 18 de diciembre de 2010, <<http://msdn.microsoft.com/es-mx/vbasic/bb466159>>.

Montero-Martínez, JM 2008, *Capítulo 4. Corpus Lingüísticos*, Grupo de Tecnología del Habla, Departamento de Ingeniería Electrónica, Universidad Politécnica de Madrid, visitada el 18 de octubre de 2011, <<http://lorien.die.upm.es/juancho/pfcs/AJP/cap4.pdf>>.

Muñoz-Fernández, S 2007, *Procesamiento del lenguaje natural para recuperación de información*, visitada el 15 de febrero de 2011, <<http://procesamientolenguajerecuperacion.50webs.org/>>.

Oracle 2010, *MySQL :: Why MySQL?*, Oracle corporation, visitada el 16 de diciembre de 2010, <<http://www.mysql.com/why-mysql/>>.

Oxford 2009, *[bnc] British National Corpus*, University of Oxford, visitada el 26 de febrero de 2010, <<http://www.natcorp.ox.ac.uk/>>.

Pérez-Sánchez, JG 2004, *Buscando el origen de los Chatinos de Nopala*, 2ª Edición, Carteles Editores, Oaxaca.

Pérez-Sánchez, JG 2010, *Buscando el origen de los Chatinos de Nopala*, 3ª Edición, Carteles Editores, Oaxaca.

Pride, L & Pride, K 1997, *Chatino de la Zona Alta, Oaxaca*, El Colegio de México, Centro de Estudios Lingüísticos y Literarios, México D.F.

Proyecto Chatino 2008, *Honrar, preservar y promover la Lengua Chatino*, Departamento de lingüística, Universidad de Texas en Austin, visitada el 24 de enero de 2010, <<http://lenguachatino.com>>.

RAE s.f., *Corpus de referencia del español actual (CREA)*, Real Academia Española, visitada el 26 de febrero de 2010, <<http://corpus.rae.es/creanet.html>>.

RAE 2010, *REAL ACADEMIA ESPAÑOLA*, Asociación de Academias de la Lengua Española, visitada el 24 de octubre de 2011, <<http://www.rae.es/rae.html>>.

Ramírez-Ramírez, JF 2001, *Apr enda Visual Basic practicando*, Pearson Educación, México D.F.

Silberschatz, A, Korth, HF & Sudarshan, S 2006, *Fundamentos de Base de Datos*, 5ª Edición, Madrid.

Sosa, E 1997, *Procesamiento del lenguaje natural: revisión del estado actual, bases teóricas y aplicaciones*, Sección científica de Biblioteconomía y Documentación, Universitat Pompeu Fabra, Barcelona, visitada el 19 de octubre de 2011, <http://www.elprofesionaldelainformacion.com/contenidos/1997/enero/procesamiento_del_lenguaje_natural_revisin_del_estado_actual_bases_tericas_y_aplicaciones_parte_i.html>.

Sherzer, J s.f., *AILLA: El archivo de los idiomas indígenas de latinoamérica*, Universidad de Texas en Austin, visitada el 4 de marzo de 2010, <<http://www.ailla.utexas.org/site/welcome-sp.html>>.

Triste-Vásquez, J 2006, *Voces de mi pueblo*, 1ª Edición, Carteles Editores, Oaxaca.

Ullman, L 2003, *Guía de aprendizaje MySQL*, Prentice Hall, Madrid.

Vallez, M & Pedraza-Jiménez, R 2007, *El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines*, Departamento de Comunicación, Universitat Pompeu Fabra, Barcelona, visitada el 19 de octubre de 2011, <<http://www.upf.edu/hipertextnet/numero-5/pln.html>>.

VARIENG 2010, *CoRD | The Brown Corpus (BROWN)*, Research unit for variation, contacts and change in English, visitada el 26 de febrero de 2010, <<http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/>>.

Vergara-Heidke, A 2006, *Recolección de corpus digitalizados del Instituto de Investigaciones Lingüísticas de la Universidad de Costa Rica*, INIL – Instituto de Investigaciones Lingüísticas, Universidad de Costa Rica, visitada el 24 de octubre de 2011, <<http://inil.ucr.ac.cr/proyecto.php?idProyecto=63>>.