



**UNIVERSIDAD DEL MAR
CAMPUS PUERTO ESCONDIDO**

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

TESIS

**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