



Universidad de León

**Departamento de Ingeniería Eléctrica
y de Sistemas y de Automática**

Tesis Doctoral

**Adaptación Automática del Vocabulario
Español y del China para el Lenguaje
VILA_1**

Jia Fu

Dirigida por:

**Dr. Don Ángel Alonso Álvarez
Dr. Don Héctor Aláiz Moretón**

León, abril de 2010

Dedicatoria

Esta tesis doctoral está dedicada a mis padres y mi abuela.

Agradecimientos

En primer lugar quiero mostrar mi mayor agradecimiento al Dr. D. Ángel Alonso Álvarez, y al Dr. D. Héctor Aláiz Moretón, directores de esta tesis doctoral, ya que sin su orientación y apoyo esta tesis no hubiera sido posible.

En segundo lugar mi agradecimiento es para todas las personas que me han ayudado y alentado a la hora de elaborar esta tesis, de entre todas esas personas quiero hacer especial mención por su significado apoyo, de las siguientes:

A los profesores Félix Riesco Peláez, Francisco Jesús Rodríguez Sedano, José Manuel Alija Pérez, Isaías García Rodríguez, Ramón Ángel Fernández Díaz, Carmen Benavides Cuellar, José Luís Calvo Rolle, Enrique López González, Luís Panizo Alonso y Javier Alfonso Cendón que siempre me aconsejaron y escucharon.

A las profesoras del centro de idiomas de la Universidad de León, sus ayudas fueron decisivas para mí en los momentos iniciales.

Al Señor Zhiwei Wang de Embajada China en España, su apoyo y animo es muy importante para me.

A mis padres, mis tíos, por el gran esfuerzo que han realizado para darme la educación que hoy me ha permitido llegar hasta aquí.

A mi novia Xiaochen Yang, por su continua comprensión y ayuda en los momentos difíciles.

También deseo mostrar mi agradecimiento a diversas instituciones, a la Dirección de la Escuela de Ingenierías Industrial e Informática, y a la Dirección del Departamento de Ingeniería Eléctrica y de Sistemas y de Automática, así como a la propia Universidad de León por haber puesto sus recursos a mi disposición cuando así lo solicité.

Resumen

Esta tesis doctoral parte del lenguaje VILA_1, que es un lenguaje visual para la comunicación genérica entre humanos, o humanos y máquinas. VILA_1 se desarrolló como consecuencia de un cambio de paradigma en la Ingeniería del Conocimiento, que consiste en unificar el lenguaje usado por los seres humanos para generar conocimiento y el lenguaje usado para formalizar el conocimiento en los ordenadores.

El trabajo realizado en esta tesis doctoral ha consistido en diseñar y desarrollar una base de datos que incluye el vocabulario del español y del chino, para realizar la traducción entre chino y español a través del lenguaje VILA_1.

Abstract

This PhD Thesis is based on VILA_1 language, which is a visual language for generic communication between humans, or humans and machines. VILA_1 was developed as the result of a paradigm shift in Knowledge Engineering involving the unification of the language used by humans to generate knowledge with the language devised to formalize knowledge in computers.

The work of this doctoral dissertation is to design and develop a database that includes the vocabulary of Spanish and Chinese, to translate between Chinese and Spanish through by language VILA_1.

摘要

这篇博士论文是 VILA_1 语言的一部分，它是一种应用于人与人之间或者人与机器之间的通用通讯的虚拟语言。VILA_1 的研发作为一种在知识工程领域的改变的成果的示范，包括如何统一人类所使用的普通语言和计算机所使用的规范语言。

在此博士论文中设计并开发了一个含有中文和西班牙词汇的数据库系统，并通过 VILA_1 语言实现中文和西班牙语之间的翻译。

ÍNDICE DE CONTENIDO

DEDICATORIA.....	I
AGRADECIMIENTOS	III
RESUMEN	V
ABSTRACT	VII
摘要.....	IX
ÍNDICES	XI
Índice de contenido	XI
Índice de figuras	XV
Índice de tablas.....	XIX
CAPÍTULO 1 (INTRODUCCIÓN).....	1
1.1 Introducción	1
1.2 La Gramática de VILA_1.....	4
1.2.1 Las expresiones lingüísticas de identificación.....	5
1.2.2 Las expresiones lingüísticas para descubrir características	9
1.2.3 Las expresiones lingüísticas para describir acciones	19
1.3 Conclusiones	22
CAPÍTULO 2 (OBJETIVOS Y JUSTIFICACIÓN)	25
2.1 Justificación.....	25
2.2 Objetivos	26
2.3 Metodología de trabajo.....	26
2.4 Estructura de la tesis.....	27
CAPÍTULO 3 (ESTADO DEL ARTE)	29
3.1 Introducción	29
3.2 Definición de un lenguaje	31
3.2.1 Lenguajes formales	33
3.2.2 Lenguajes naturales.....	34
3.3 Lingüística Computacional	35

3.3.1	Antecedentes históricos	35
3.3.2	Procesamiento del lenguaje natural desde 1960 hasta 1980.....	36
3.3.3	Procesamiento del lenguaje natural desde 1980	38
3.3.3.1	Aplicaciones tradicionales.....	40
3.3.3.2	Nuevas aplicaciones	41
3.3.3.3	Nuevos recursos.....	42
3.4	Lenguajes Visuales	44
3.4.1	Aplicaciones de los lenguajes visuales.....	47
3.4.2	Lenguajes visuales de programación.....	50
3.4.3	Lenguajes visuales aplicados a la comunicación.....	51
3.4.4	Lenguajes Visuales de Dominio Específico	52
3.5	La base de datos.....	53
3.5.1	La evolución de los sistemas de base de datos.....	54
3.5.2	Primeros sistemas de base de datos	55
3.5.3	Sistemas de base de datos relaciones.....	57
3.5.4	Sistema de base de datos orientados a objetos	59
3.5.5	Bases de Datos Relacionales Orientadas a Objetos.....	60
3.5.6	Situación actual de sistemas de base de datos	62
3.6	La traducción automática.....	62
3.6.1	La historia de traducción automática.....	63
3.6.1.1	El período de crear (1946-1964).....	63
3.6.1.2	El período de retroceso (1964-1975).....	64
3.6.1.3	El período de recuperación (1975-1989).....	64
3.6.1.4	El período de nuevo (1990-actualidad)	64
3.6.2	La teoría básica de traducción automática.....	65
3.6.3	Los problemas de traducción automática	65
3.6.4	Los productos típicos de traducción automática.....	66
3.7	Conclusiones.....	67

CAPÍTULO 4 (DISEÑO Y DESARROLLO DE LA BASE DE DATOS) 69

4.1	Introducción.....	69
4.2	El diseño de la estructura.....	70
4.2.1	El sistema de gestión de base de datos: MYSQL	71
4.2.2	Los idiomas	71
4.2.2.1	El uso de los idiomas	71
4.2.2.2	Los problemas de gramática entre el chino y el español	74
4.2.2.2.1	Los diferentes en los verbos	75
4.2.2.2.2	Género y número	78
4.2.3	Los procedimientos de desarrollo.....	79
4.2.3.1	Determinar las entidades	80
4.2.3.2	Determinar los datos necesarios	83
4.2.3.3	Normalizar los datos.....	92
4.2.3.4	Importar los datos	95

4.2.3.5 Mantener la base de datos	97
4.3 Uso de la Base de Datos	98
CAPÍTULO 5 (VALIDACIÓN EMPÍRICA)	131
5.1 Introducción	131
5.2 Metodología	132
5.2.1 Selección de la población.....	132
5.2.2 Enteramiento de la población.....	134
5.2.3 Diseño de cuestionarios.....	135
5.2.4 Resolución de cuestionarios por parte de los usuarios.....	141
5.2.5 Análisis de resultados y conclusiones	141
CAPITULO 6 (CONCLUSIONES FINALES Y TRABAJOS FUTUROS).....	145
6.1 Conclusiones finales.....	145
6.2 Trabajos futuros	146
ABREVIATURAS	147
REFERENCIAS	149
ANEXO	157

ÍNDICE DE FIGURAS

CAPÍTULO 1 (INTRODUCCIÓN).

Figura 1.1 Paradigma clásico de la Ingeniería del Conocimiento	2
Figura 1.2 Los ejemplos de la expresión lingüística de identificación.....	7
Figura 1.3 Los ejemplos de la aplicación de las expresiones lingüísticas de identificación	8
Figura 1.4 Los ejemplos de las agrupaciones tipo "Y" y tipo "O"	9
Figura 1.5 Los ejemplos de la expresión lingüística de la característica "Adjetivo"	10
Figura 1.6 Los ejemplos de la expresión lingüística de la característica "Valor"	11
Figura 1.7 Los ejemplos de la expresión lingüística de la característica "Relación"	12
Figura 1.8 Los ejemplos de la expresión lingüística de la característica "Espacio"	15
Figura 1.9 Los ejemplos de la expresión lingüística de la característica "Tiempo"	18
Figura 1.10 Los ejemplos de la expresión lingüística de la agrupación de características	19
Figura 1.11 Los ejemplos de la expresión lingüística para describir acciones	21
Figura 1.12 Los ejemplos de las expresiones complementarias de las acciones	22

CAPÍTULO 3 (ESTADO DEL ARTE).

Figura 3.1 Tipos de lenguajes. Clasificación de VILA_1	32
--	----

CAPÍTULO 4 (DISEÑO Y DESARROLLO DE LA BASE DE DATOS).

Figura 4.1 Dos procesos de construir la base de datos.....	70
Figura 4.2 Los más usos de los idiomas en el mundo	72
Figura 4.3 Los más usos de los idiomas en Internet.....	73
Figura 4.4 Los primero tres usos de los idiomas.....	73
Figura 4.5 La estructural de la frase de tipo SVO	81
Figura 4.6 Las tablas de base de datos	82
Figura 4.7 Expresiones en VILA_1.....	93

Figura 4.8 El adjetivo en VILA_1	94
Figura 4.9 Frase en VILA_1	94
Figura 4.10 La interfaz de importar palabras.....	96
Figura 4.11 Las mensajes realizan la traducción a través por la base de datos.....	98
Figura 4.12 La frase española en VILA_1.....	99
Figura 4.13 Representación gráfica del XML Schema de una estructura de tipo “Acción”	102
Figura 4.14 XML del archivo “Acción” del ejemplo	103
Figura 4.15 XML del archivo “Acción” del ejemplo en chino	105
Figura 4.16 La frase china en VILA_1	108
Figura 4.17 La frase española compleja en VILA_1	109
Figura 4.18 Representación gráfica del XML Schema de una estructura de tipo “Acción”	109
Figura 4.19 XML del archivo “Acción” del ejemplo	110
Figura 4.20 Representación gráfica del XML Schema de una estructura de tipo “Adjetivo”	113
Figura 4.21 XML del archivo “Adjetivo” del ejemplo.....	114
Figura 4.22 Representación gráfica del XML Schema de una estructura de tipo “Valor”	115
Figura 4.23 XML del archivo “Valor” del ejemplo.....	116
Figura 4.24 Representación gráfica del XML Schema de una estructura de tipo “Espacio”.....	117
Figura 4.25 XML del archivo “Espacio” del ejemplo	118
Figura 4.26 Representación gráfica del XML Schema de una estructura de tipo “Relación”	120
Figura 4.27 XML del archivo “Relación” del ejemplo.....	121
Figura 4.28 XML del archivo “Acción” del ejemplo en chino.....	122
Figura 4.29 XML del archivo “Adjetivo” del ejemplo en chino	124
Figura 4.30 XML del archivo “Valor” del ejemplo en chino	125
Figura 4.31 XML del archivo “Espacio” del ejemplo en chino	126
Figura 4.32 XML del archivo “Relación” del ejemplo en chino.....	128
Figura 4.33 La frase china compleja en VILA_1	129

CAPÍTULO 5 (VALIDACIÓN EMPÍRICA)

Figura 5.1 La expresión de formato VILA_1 español a chino	136
Figura 5.2 La expresión de formato VILA_1 español a chino	136
Figura 5.3 La expresión de formato VILA_1 español a chino	137
Figura 5.4 La expresión de formato VILA_1 español a chino	137
Figura 5.5 La expresión de formato VILA_1 español a chino	138
Figura 5.6 La expresión de formato VILA_1 chino a español	138
Figura 5.7 La expresión de formato VILA_1 chino a español	139
Figura 5.8 La expresión de formato VILA_1 chino a español	139

Figura 5.9 La expresión de formato VILA_1 chino a español.....	140
Figura 5.10 La expresión de formato VILA_1 chino a español.....	140
Figura 5.11 Gráfico de resultados de la evaluación de la traducción de chino a español.....	142
Figura 5.12 Gráfico de resultados de la evaluación de la traducción de español a chino.....	142

ÍNDICE DE TABLAS

CAPÍTULO 4 (DISEÑO DE LA APLICACIÓN DE INTERCAMBIO DE MENSAJES CON VILA_1).

Tabla 4.1 La estructura de la tabla	84
Tabla 4.2 Los dominios de sustantivos	87
Tabla 4.3 La estructura de la tabla sustantivos.....	89
Tabla 4.4 La estructura de la tabla verbos.....	90
Tabla 4.5 La estructura de la tabla adjetivos	90
Tabla 4.6 La estructura de la tabla adverbios.....	91
Tabla 4.7 La estructura de la tabla numeral	91
Tabla 4.8 La estructura de la tabla lugar	92
Tabla 4.9 La estructura de la tabla tiempo	92
Tabla 4.10 Las tablas en la base de datos.....	97

CAPÍTULO 5 (VALIDACIÓN EMPÍRICA).

Tabla 5.1 Participantes chinos.....	133
Tabla 5.2 Participantes españoles.....	134

CAPÍTULO 1

Introducción

1.1 Introducción

Esta tesis doctoral forma parte de un proyecto de investigación desarrollado por el grupo de Cognomática de la Universidad de León. En el contexto de ese proyecto se ha creado un nuevo lenguaje visual al que se ha denominado VILA_1 [1]. El desarrollo de este lenguaje se ha apoyado en varios trabajos previos de Alonso [2] [3] [4] [5] [6] [7] [8]. El objetivo de este lenguaje es solucionar alguno de los problemas que tiene planteados la Ingeniería del Conocimiento. La Ingeniería del Conocimiento pretende representar el conocimiento de forma que pueda ser procesado automáticamente mediante un ordenador.

Son muchas las dificultades encontradas para automatizar el conocimiento, hasta el punto que, después de treinta años de investigaciones, el tema no está resuelto. En el grupo de investigación

sobre Cognomática de la Universidad de León ha llegado a la conclusión de que para seguir avanzando es preciso cambiar de paradigma. A continuación se analizan las razones. En la (*Figura 1.1*) se sintetiza la estrategia seguida hasta la fecha:

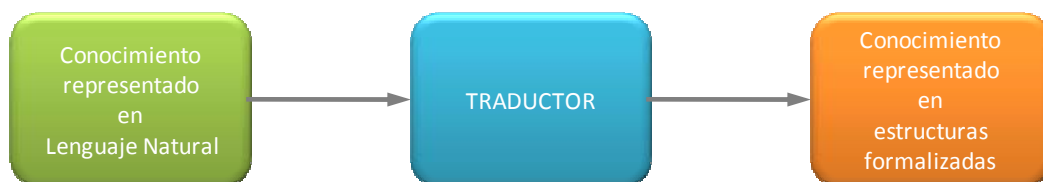


Figura 1.1 Paradigma clásico de la Ingeniería del Conocimiento

El conocimiento es generado en Lenguaje Natural por los seres humanos. Para poder automatizarlo en los ordenadores es preciso representar ese conocimiento en estructuras formalizadas. El paso del conocimiento expresado en Lenguaje Natural a estructuras formalizadas requiere un proceso de traducción. Esta traducción es realizada por los seres humanos.

¿Dónde está el problema? En el proceso de traducción. Para formalizar el conocimiento el traductor necesita eliminar las ambigüedades propias del Lenguaje Natural, su falta de precisión, las diversas acepciones de los términos, su borrosidad estructural. Y para realizar esta tarea no hay normas, no hay estandarización. Cada traductor aplica sus propios criterios y eso conduce a resultados muy subjetivos. Las estructuras formalizadas por un traductor son difícilmente compatibles con las realizadas por otro. El objetivo básico de la reusabilidad no se consigue.

Además del grave problema citado hay más, entre ellos:

- El traductor debe ser un experto en el dominio de conocimiento que se propone traducir y también debe ser un experto en las técnicas de formalización del conocimiento. Actualmente es difícil encontrar perfiles profesionales que cumplan esta condición.

- La generación de conocimiento en Lenguaje Natural es tan dinámica y voluminosa que se necesitaría una legión de traductores para formalizar el conocimiento generado continuamente.

Después de varios años de reflexión hemos llegado a la conclusión de que es necesario eliminar el proceso de traducción, causante de los males citados. Nuestra propuesta consiste en generar el conocimiento directamente en un lenguaje formalizado, un lenguaje común para el hombre y para la máquina. Y esto representa un cambio de paradigma. Este nuevo lenguaje debe tener resueltos en su estructura los problemas de ambigüedad y borrosidad propios del lenguaje natural. ¿Cómo hacerlo? Para responder a esta pregunta se ha seguido el siguiente razonamiento:

- Puesto que la inmensa mayoría de las transacciones de conocimiento hombre-máquina y hombre-hombre son a través de una pantalla electrónica el lenguaje debe tener una orientación visual.
- Para obtener buen rendimiento de la gran capacidad de procesamiento que ofrece el sentido de la vista el lenguaje debe formularse al menos en dos dimensiones.
- Para que el nuevo lenguaje sea fácilmente aceptado debe introducirse de forma progresiva.

Se ha bautizado al nuevo lenguaje como VILA (Visual Language), y se ha dividido en tres niveles que se denominan VILA_1, VILA_2 y VILA_3.

VILA_1 utiliza muchas palabras de los lenguajes naturales e incorpora un conjunto amplio de elementos gráficos, inicialmente en dos dimensiones, pero que con su desarrollo es previsible que se amplíe con elementos tridimensionales y animados.

VILA_2 mantiene los elementos gráficos de VILA_1 y elimina todas las palabras. Se produce una desconexión total con los lenguajes naturales hasta el punto de que se generará una grafía específica para representar los conceptos que se definan.

VILA_3 es una ampliación de VILA_2. Incorporará representación gráfica directa de los elementos físicos definidos mediante el lenguaje, así como los cálculos matemáticos necesarios para obtener y representar las magnitudes y los parámetros propios de la ciencia y de la tecnología.

VILA_1 ya ha sido desarrollado. Su aprendizaje y manejo es inmediato, ya que los elementos gráficos que incorpora son muy intuitivos y las palabras que usa coinciden con las de los lenguajes naturales, pueden usarse las del idioma que cada uno desee.

En síntesis, lo que se está diciendo es que el Lenguaje Natural, gran artífice del desarrollo de la Humanidad, ha tocado techo. La ciencia y la tecnología necesitan un lenguaje estructurado científicamente. Es una necesidad vital. El Lenguaje Natural fue creado en otro contexto histórico, con unas necesidades muy diferentes a las actuales y sin las herramientas electrónicas que existen y las que van a existir. El lenguaje es tan importante en el desarrollo del conocimiento que si no evoluciona, el lenguaje, el desarrollo del conocimiento quedará seriamente amenazado.

1.2 La Gramática de VILA_1

La gramática de VILA_1 [1] se estructura sobre los conceptos de *expresión lingüística* y de *rectángulo*. Hay muchos tipos de expresiones lingüísticas. Todas ellas se construyen mediante agrupaciones de los tres tipos básicos. Estos son:

- Las expresiones lingüísticas de identificación. Sirven para identificar las entidades y los conceptos, desde el punto de vista

del destinatario de la expresión. Se entiende por entidad cualquier elemento diferenciado de otros, tanto real como virtual.

- Las expresiones lingüísticas para describir características. Sirven para describir las características de las entidades y de los conceptos.
- Las expresiones lingüísticas para describir acciones. Sirven para describir las acciones que producen las fuerzas de la Naturaleza. Cualquier fuerza.

En las expresiones de identificación se supone que todo lo expresado es conocido por el destinatario. Sin embargo en los otros dos tipos de expresiones se supone que se exponen conceptos desconocidas para el destinatario.

El rectángulo sirve para ubicar las expresiones lingüísticas en el plano y para referenciarlas. Hay tantos tipos de rectángulos como tipos de expresiones lingüísticas. Cada expresión lingüística lleva asociado su correspondiente rectángulo. Puede decirse que cada rectángulo es el contenedor de una expresión lingüística.

1.2.1 Las expresiones lingüísticas de identificación

Son el primero de los tres tipos básicos de expresiones lingüísticas. El objetivo de este tipo de expresiones es el de identificar tanto los conceptos como las entidades.

➤ *Identificación mediante término*

Consiste en utilizar un solo término. Puede ser una palabra u otro elemento gráfico cualquiera (signos, dibujos, fotografías, etc.). Algunos ejemplos son los siguientes:



Cuando se usan palabras de un idioma, cualquiera, sólo se utiliza una forma de la misma, la más común. No se usa el plural y la diferenciación entre masculino y femenino sólo se hace al referirse a seres vivos que tengan diferenciación de sexos (Ej.: vaca, buey, etc.). Pero a los objetos o conceptos no se les asigna género.

También se incluyen en este apartado los términos destinados a identificar individualidades o entidades concretas. Es el caso de los denominados nombres propios en los lenguajes naturales. Aunque aquí se incluye cualquier tipo de identificador de carácter alfanumérico como por ejemplo la matrícula de un coche. En todos los casos el identificador aparecerá entre comillas. Esos identificadores pueden constar de cualquier conjunto de caracteres alfanuméricos, incluido el espacio en blanco. Ejemplos:

"España"

"Juan Pérez"

"NCV 33465"

"Pablo VI"

Otro tipo de identificadores que se consideran de término único son aquellos identificadores de carácter genérico, es decir, que sirven para representar conceptos o conjuntos de individualidades, y que constan de varias palabras unidas en forma de frases hechas. En este caso se unirán todas las palabras de la expresión mediante una barra baja “_”. Ejemplos:

carnet_identidad

Número_bastidor

Colegio_profesional

Escuela_ingeniería

➤ *Identificación mediante término único con especificadores*

Consiste en concretar el ámbito de identificación de un término único mediante una serie de especificadores gráficos. Entre esos especificadores están los siguientes:

- Negación (X);
- Interrogación (?);

- Singular/plural: por defecto se entiende singular. Para especificar plural (☑).
- Cuantificación numérica: un número; En combinación con los números se pueden combinar algunos signos matemáticos como: < ; > ; ≈.
- Cuantificación cualitativa: ninguno (☐); poco/s (☑); bastante/s (☑); mucho/s (☑); todo/s (☑).
- Determinación (☑); La ausencia de determinación se considera indeterminación.

Algunos ejemplos son los siguientes:

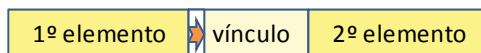
Expresiones en VILA_1	Traducción a lenguaje natural
☑ niño	El niño
6 elefante	Seis elefantes
☐ árbol	Ningún árbol
☑ juguete	Muchos juguetes
☑ ilusión	Las ilusiones
> 10 mesa	Más de diez mesas
monumento	Un monumento

Figura 1.2 Los ejemplos de la expresión lingüística de identificación

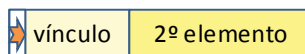
➤ ***Ampliación de las expresiones lingüísticas de identificación***

Tanto las expresiones lingüísticas para describir características, como las destinadas a describir acciones pueden usarse para identificar elementos. Cuando así sea se colocará delante el siguiente icono (☑). En el gráfico anterior se muestran algunos ejemplos.

En el caso de las expresiones lingüísticas para expresar características de relación es frecuente utilizar el vínculo y el segundo elemento para identificar el primer elemento. La estructura de la expresión es la siguiente:



Y en este caso se transforma en esta otra:



Algunos ejemplos de este tipo de expresiones son las siguientes:











Expresiones en VILA_1	Traducción a lenguaje natural
  madre Ana	La madre de Ana
  autor  trabajo	El autor del trabajo
  capital España	La capital de España
  veterinario  perro	El veterinario del perro

Figura 1.3 Los ejemplos de la aplicación de las expresiones lingüísticas de identificación

➤ **Agupaciones tipo “Y” y tipo “O”**

Las expresiones lingüísticas del mismo nivel pueden agruparse mediante los operadores tipo “y” y tipo “o”. Los signos gráficos utilizados para cada tipo son los siguientes:

- Agrupación tipo “Y”: ,
- Agrupación tipo “O”: |

Algunos ejemplos que incluyen agrupaciones de este tipo son los siguientes:

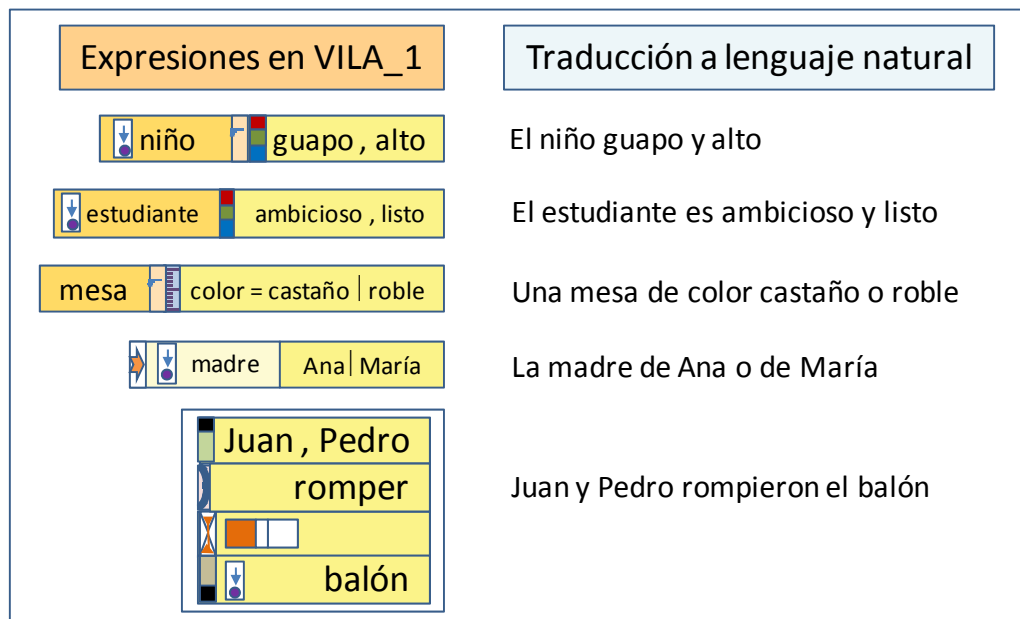


Figura 1.4 Los ejemplos de las agupaciones tipo "Y" y tipo "O"

1.2.2 Las expresiones lingüísticas para describir características

El segundo tipo de expresiones lingüísticas básicas lo constituyen las que se utilizan para describir características. Son expresiones que en el lenguaje natural llevan asociados los verbos **ser**, **estar**, **tener** y algunos otros. Estos verbos tienen grandes particularidades y diferencias entre unos idiomas y otros. En VILA_1 no se utilizan verbos para expresar las características de algo. Las características siempre se aplican a uno o más elementos ya identificados, mediante la correspondiente expresión lingüística. Las características tienen el objetivo de trasladar información, en principio desconocida, para el destinatario de la expresión lingüística. Estas se distribuyen en algunos grupos.

➤ *La característica "Adjetivo"*

Los adjetivos de los lenguajes naturales sirven para darle valor cualitativo a magnitudes y otros elementos de valoración. El

rectángulo destinado a expresar una característica adjetivo tiene la siguiente forma:



Delante del adjetivo pueden incluirse los siguientes especificadores ya descritos anteriormente: X; ?; [icon]; [icon]; [icon].

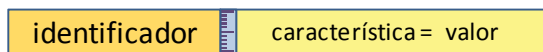
A continuación se describen algunos ejemplos:

Expresiones en VILA_1	Traducción a lenguaje natural
[icon] niño [icon] guapo	El niño es guapo
[icon] estudiante [icon] X ambicioso	El estudiante no es ambicioso
[icon] cuerda [icon] [icon] rígido	La cuerda es muy rígida
Antonio [icon] ? [icon] inteligente	¿Es Antonio muy inteligente?

Figura 1.5 Los ejemplos de la expresión lingüística de la característica "Adjetivo"

➤ **La característica "valor"**

Los adjetivos de los lenguajes naturales, como ya se dijo, engloban en un solo término tanto el tipo de característica como el valor cualitativo asignado a la misma. Pero también es posible especificar la característica y asignarle un valor. Es lo que se hace en el rectángulo que denominamos característica valor. El rectángulo destinado a expresar una característica valor tiene la siguiente forma:



Para la asignación del valor hay varias opciones:

- Que exista un término, palabra o elemento gráfico, que identifique el valor de la característica, por ejemplo "rojo",

“rubio”, etc. En este caso se puede especificar el valor con la palabra correspondiente o con un elemento gráfico que lo represente.

- Que la característica permita cuantificación numérica. En este caso se expresará el número correspondiente, seguido de un espacio y la palabra que identifique la magnitud (Ej.: metro, kilogramo, etc.). El número puede ir precedido de los siguientes especificadores matemáticos: $<$; $>$; \approx . Cuando se quiera especificar un rango se intercalará entre los dos números un guión (-).
- Cuando en una característica no se den ninguno de los dos casos anteriores se recurrirá a los elementos gráficos que representan los conceptos: nada, poco, bastante, mucho, todo.

A continuación se describen algunos ejemplos:

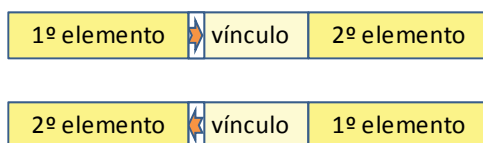
Expresiones en VILA_1	Traducción a lenguaje natural
torre altura = 12 metro	La torre tiene una altura de doce metros
tabla ancho = 25 centímetro	Una tabla tiene veinticinco centímetros de ancho
niño ilusión =	Un niño tiene mucha ilusión
moto matrícula = 5267 NJK	La matrícula de la moto es 5267 NJK
mesa color = castaño	Una mesa es de color castaño
Alberto peso = > 100 kilogramo	Alberto pesa más de cien kilos
señal frecuencia = 1,5 - 2 gigahertzio	La señal tiene una frecuencia comprendida entre 1, 5 y 2 gigahertzios

Figura 1.6 Los ejemplos de la expresión lingüística de la característica “Valor”

➤ *La característica “relación”*

La característica relación permite expresar explícitamente el vínculo que existe entre dos elementos, sean entidades o conceptos. Por

ejemplo el concepto autor permite vincular a un artista con su obra, el concepto hermano permite vincular a dos personas, etc. Esta es una característica que tiene estructura bidireccional, es decir, que puede leerse hacia adelante y hacia atrás. El símbolo de esta característica tiene forma de flecha, la cual indica en la dirección que debe hacerse la lectura. La característica incorpora, además del símbolo que la identifica, otros tres rectángulos. En los dos de los extremos se especifican los elementos que se vinculan y en el central el tipo de vínculo que tienen. Si no se desea especificar el tipo de vínculo puede dejarse este campo en blanco. La estructura es la siguiente:



A continuación se describen algunos ejemplos:

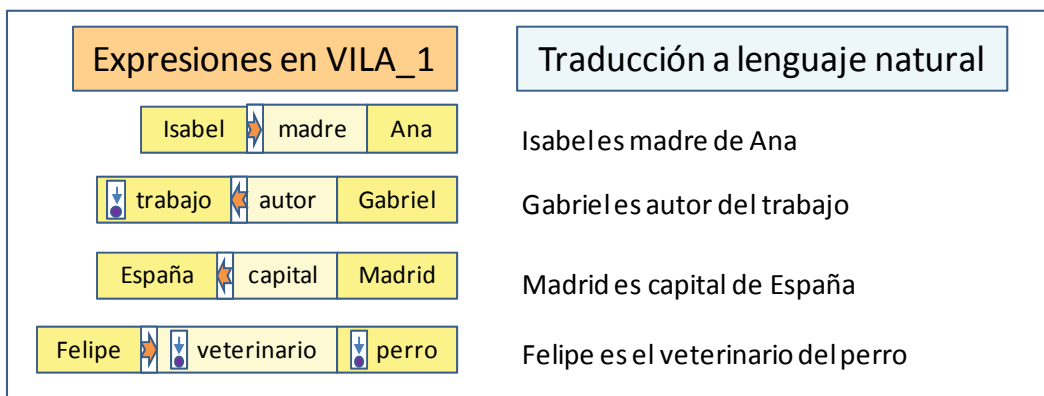


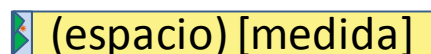
Figura 1.7 Los ejemplos de la expresión lingüística de la característica "Relación"

➤ **La característica "espacio"**

Mediante esta característica puede especificarse el espacio físico o virtual en el que se encuentra ubicada la entidad o el concepto que se está describiendo o donde sucede algo. Hay diversas formas de especificar el espacio. En VILA_1 se usan los siguientes términos relativos como punto de partida para especificar el espacio:

- Dentro
- Fuera
- Arriba
- Debajo
- Derecha
- Izquierda
- Norte
- Sur
- Este
- Oeste
- Al lado
- Cerca
- Lejos
- Alrededor
- Periferia
- Entre
- Desde ... hasta

Exceptuando para los dos últimos términos relativos (*entre y desde ... hasta*), el rectángulo del espacio tiene la siguiente estructura:



La sintaxis del campo espacio es la siguiente:

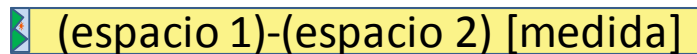
- Siempre comienza por un término relativo (*dentro, fuera, arriba*, etc.), seguido de una flecha. El término relativo *dentro* puede omitirse ya que se considera el término por defecto.
- A continuación se coloca una expresión identificadora de un espacio (armario, jardín, casa, etc.).
- Las expresiones descritas en los dos puntos anteriores pueden secuenciarse tantas veces como se desee. Irán separadas por “;”.
- Pueden usarse como identificadores de espacio figuras geométricas (**círculo, triángulo, cuadrado**, etc.). En este caso podrán incluirse entre corchetes todos los datos de

identificación de la figura que se deseen, separados por “;”. Por ejemplo: **círculo [centro = ciudad “Madrid”; radio = 80 kilómetro]**

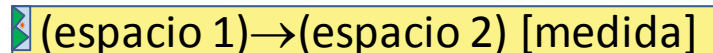
- También podrán usarse como identificadores términos como **ruta y trayectoria**, añadiendo entre corchetes los datos que se deseen.

En el campo medida podrá, si se desea, aportarse información sobre longitud, superficie y volumen del espacio definido previamente.

Para el término relativo *entre* el rectángulo del espacio tiene la siguiente estructura:

 (espacio 1)-(espacio 2) [medida]

Y finalmente para el término relativo *desde ... hasta* el rectángulo del espacio tiene la siguiente estructura:

 (espacio 1)→(espacio 2) [medida]

La sintaxis de **espacio 1**, **espacio 2** y **medida** es igual a la descrita previamente para los otros términos relativos. En el caso de la expresión correspondiente al término relativo *desde ... hasta* podrán dejarse en blanco cualquiera de los dos espacios, con lo que la expresión pasa a convertirse en *desde* o *hasta* respectivamente. Estos términos relativos pueden anidarse con los anteriores, formando expresiones todo lo complejas que se desee.

A continuación se describen algunos ejemplos:

Expresiones en VILA_1	Traducción a lenguaje natural
(Cathedral)	Dentro de la Catedral
(encima → armario; habitación)	Encima del armario que está en la habitación
(lejos → ciudad "Madrid")	Lejos de la ciudad de Madrid
(dentro → oeste → jardín)	Dentro y al oeste del jardín

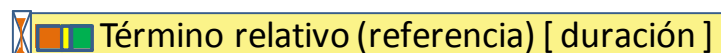
Figura 1.8 Los ejemplos de la expresión lingüística de la característica "Espacio"

➤ **La característica "tiempo"**

Mediante esta característica puede especificarse el momento o el intervalo temporal en el que algo se define o algo sucede. A lo largo del tiempo todo cambia. Pero el ritmo del cambio es diferente o muy diferente para unas cosas y para otras. Hay diversas formas de especificar un intervalo temporal. En VILA_1 se usan los siguientes términos relativos como punto de partida para especificar el tiempo:

- Antes
- Después
- Durante
- Desde
- Hasta
- Desde Hasta
- Entre

En cualquiera de los casos anteriores se puede añadir la duración del intervalo si se desea. La estructura del rectángulo tiempo varía según los términos relativos que se usen. Para los términos *antes*, *después*, *durante desde* y *hasta* la estructura es la siguiente:



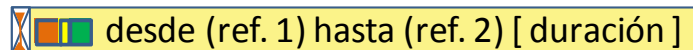
Además del icono de identificación del rectángulo tiempo hay otros cuatro campos en el interior del rectángulo. En el primero se puede indicar si se está refiriendo al pasado (marrón), al presente (amarillo) o al futuro (verde). En el segundo se especifica el término relativo a la referencia temporal. En el tercero se indica la referencia temporal y en el cuarto la duración del intervalo.

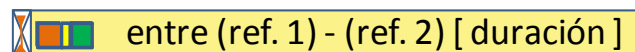
La sintaxis de los distintos campos es la siguiente:

- Campo genérico:
 - Marrón = pasado
 - Amarillo = presente
 - Verde = futuro
 - Marrón-amarillo-verde = pasado, presente y futuro
 - Marrón-amarillo = pasado y presente
 - Amarillo-verde = presente y futuro
 - Marrón-verde = pasado y futuro
- Los términos relativos pueden ser cualquiera de los siguientes:
 - Antes
 - Después
 - Durante. *Si no expresa nada se sobreentiende este término.*
 - Desde
 - Hasta
- En el campo referencia se indicará una expresión temporal que puede tener los siguientes significados:
 - Intervalo temporal cuando se usa con *durante*.
 - Referencia anterior cuando se usa con *después*.
 - Referencia posterior cuando se usa con *antes*.
 - Punto de partida cuando se usa con *desde*.
 - Punto final cuando se usa con *hasta*.
- En todos los casos la expresión temporal puede adoptar las siguientes formas:
 - Mediante un nombre. Ej.: enero, lunes, noche, etc.

- Mediante una unidad temporal y un número. Ej.: (día = 3), (año = 1987), (hora = 12), etc.
- Mediante una secuencia, separada por “;” de identificadores individuales. Se colocaran primero los identificadores temporales de menor duración. Ej.: (día = 3; marzo), (día = 15; mes = agosto; año = 1990;), (mediodía; mañana), etc.
- Mediante una expresión lingüística tipo acción. Se analizará posteriormente.
- Campo “duración”. Existen las siguientes posibilidades para identificar la duración:
 - Mediante un número y la correspondiente unidad temporal. Ej.: (3 día), (2 semana), (15 minuto), etc.
 - Combinando varias unidades temporales. Ej.: (1 mes, 6 día)
 - Mediante una expresión lingüística tipo acción. Se analizará posteriormente.

Para los términos *desde ... hasta* y *entre* la estructura es la siguiente:

 desde (ref. 1) hasta (ref. 2) [duración]

 entre (ref. 1) - (ref. 2) [duración]

La sintaxis de las referencias temporales y de la duración es la misma que la descrita anteriormente.

A continuación se describen algunos ejemplos:







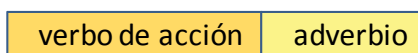



Expresiones en VILA_1	Traducción a lenguaje natural
 desde (hora = 10; ayer) hasta (hora =16; ayer)	Ayer desde las 10 de la mañana hasta las 4 de la tarde
 desde (mañana) [3 mes]	Desde mañana y durante 3 meses
  (año =2004)	Durante el año 2004
  después (agosto)	Después de agosto

Figura 1.9 Los ejemplos de la expresión lingüística de la característica "Tiempo"

➤ **La característica "adverbio"**

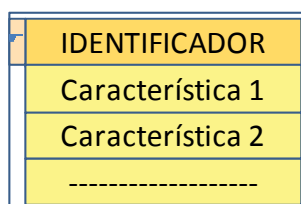
Los adverbios sirven para calificar la forma en la que se realiza una acción. Van siempre asociados a un verbo de acción y por eso no es necesario adjudicarles un icono de identificación. El rectángulo destinado a expresar una característica adverbio tiene la siguiente forma:




Delante del adverbio pueden incluirse los siguientes especificadores ya descritos anteriormente: X; ?; ; ; .

➤ **Agrupación de características**

Cuando se desee expresar varias características de un mismo elemento pueden agruparse todas ellas en la siguiente estructura:



Puesto que el identificador de todas las características es común se especifica una sola vez y de esa forma puede eliminarse de cada una de las características. Teniendo en cuenta que el elemento aglutinador de este tipo de expresiones es un identificador, se le asigna el icono de identificación (), para facilitar la comprensión de la expresión. En el siguiente gráfico se muestran algunos ejemplos. El segundo ejemplo contiene el símbolo (ϵ), para indicar el concepto de pertenencia.

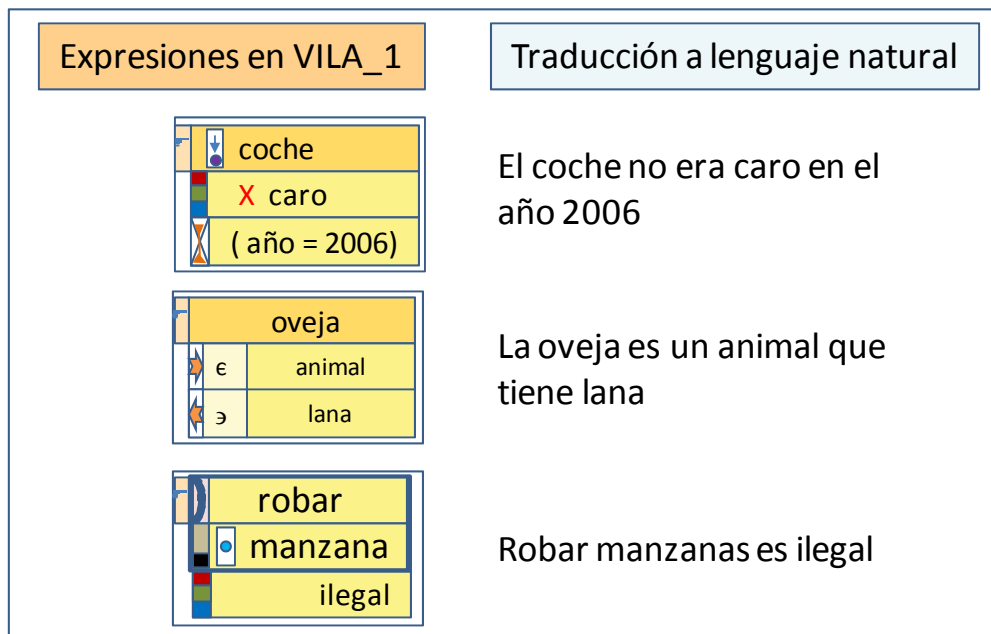


Figura 1.10 Los ejemplos de la expresión lingüística de la agrupación de características

1.2.3 Las expresiones lingüísticas para describir acciones

El tercer bloque de las expresiones lingüísticas básicas lo constituyen las que se utilizan para describir acciones. Son expresiones que sirven para describir los cambios de características de un concepto o entidad. Estos cambios se producen por la acción de fuerzas y siempre están

asociados a un intervalo temporal. La estructura del rectángulo para describir acciones es la siguiente:



El elemento central de este tipo de expresiones es el *verbo*. Con él se describe la acción correspondiente. El verbo se expresa siempre en infinitivo. El *sujeto* es el elemento generador de las fuerzas que producen la acción. El sujeto se expresa con una expresión de identificación. Mediante el *tiempo* se expresa el momento o el intervalo temporal en el que se produce la acción. La estructura del rectángulo tiempo ya se explicó en el apartado correspondiente a las expresiones para describir características. Solo cabe añadir aquí que para expresar el imperativo se dejarán en blanco los rectángulos correspondientes al pasado, al presente y al futuro. Cuando se quiera utilizar el verbo en infinitivo no se usará el rectángulo tiempo.

El *objeto* de la acción es el elemento sobre el que actúan directamente las fuerzas propias de la acción. En algunos casos el objeto de una acción puede ir destinado o dirigido a otro elemento. A este segundo elemento se le denomina *destinatario* de la acción. El rectángulo *espacio* sirve para especificar el lugar donde se desarrolla la acción. Sus características ya se describieron en el apartado correspondiente a las expresiones para describir características. En el gráfico siguiente se muestran algunos ejemplos.

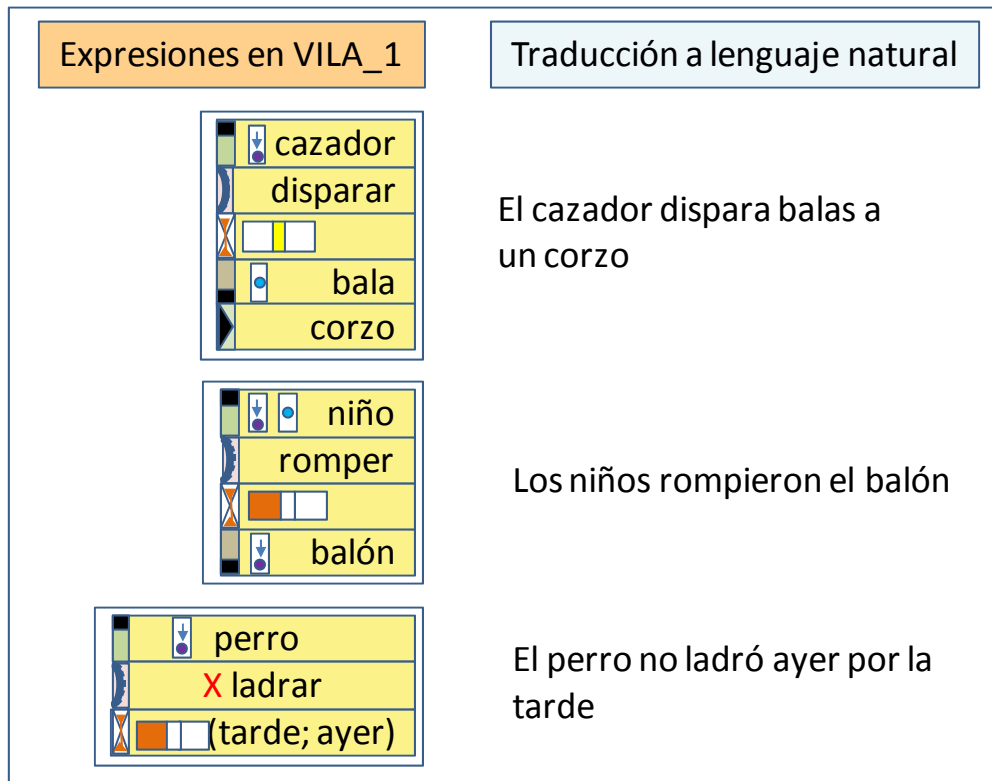


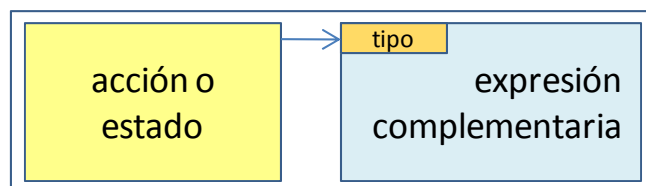
Figura 1.11 Los ejemplos de la expresión lingüística para describir acciones

➤ ***Expresiones complementarias de las acciones***

Con frecuencia es preciso complementar las acciones. Son expresiones que sirven para especificar:

- El objetivo de la acción.
- La causa de la acción.
- La condición de la acción.
- Las consecuencias de la acción.

La estructura general es la siguiente:



Algunos ejemplos de este tipo de expresiones son los siguientes:

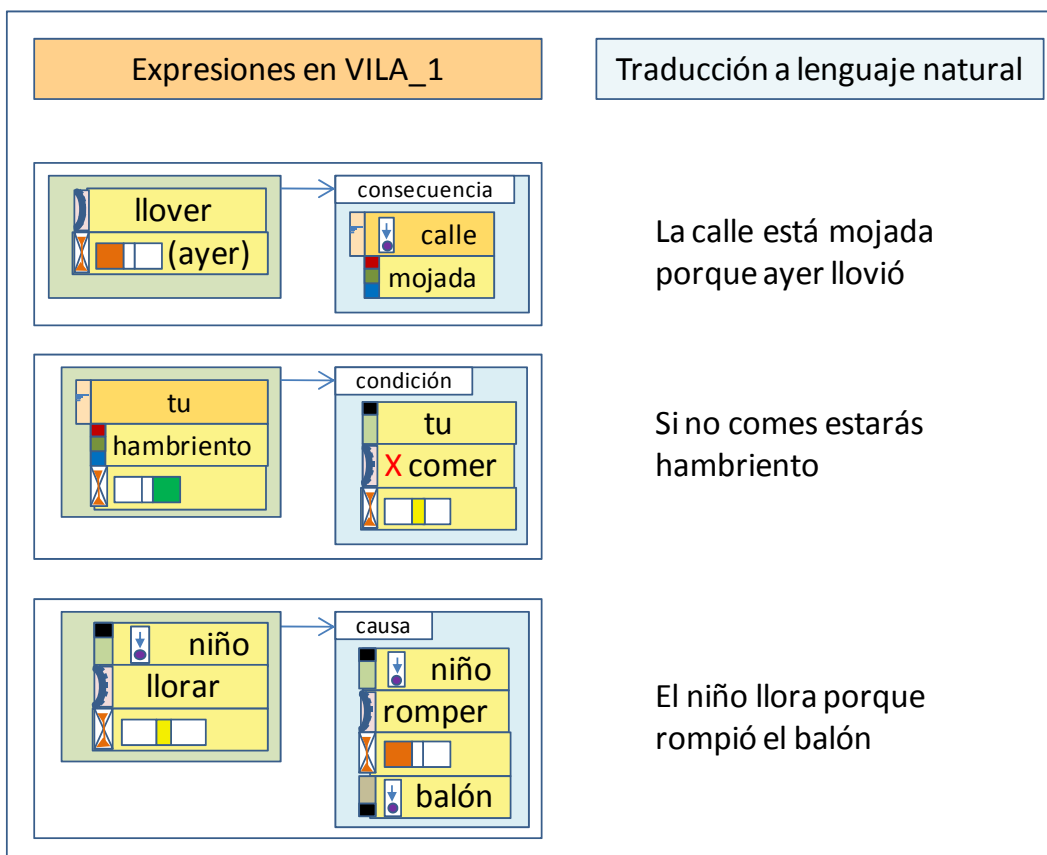


Figura 1.12 Los ejemplos de las expresiones complementarias de las acciones

1.3 Conclusiones

De lo dicho en las páginas anteriores puede deducirse lo siguiente:

- VILA_1 es un lenguaje nuevo, creado para la comunicación tanto hombre-hombre, como hombre-máquina, siempre mediante el uso de una pantalla electrónica.
- VILA_1 es el lenguaje de la ACCESIBILIDAD, y es así porque comparado con los lenguajes naturales es:
 - Mucho menos complejo. Al utilizar 2 dimensiones todo es más sencillo.
 - Menos ambiguo. Porque tiene una estructura semi-formal.

- Es universal. Porque permite la comunicación entre personas de cualquier idioma.
- VILA_1 tiene dos características básicas que le diferencian de los lenguajes naturales:
 - Es un lenguaje diseñado para optimizar la percepción visual, que admite las tres dimensiones espaciales y el tiempo, es decir, la animación.
 - Es un lenguaje semi-formal. Por lo tanto permite un nivel importante de automatización, incluidas todas las búsquedas conceptuales vinculadas a la estructura sintáctica del lenguaje.
- VILA_1 no es un traductor de lenguajes naturales.
- VILA_1 es un lenguaje visual que tiene tantas variantes como idiomas naturales (Hay VILA_1_Español; VILA_1_Chino; VILA_1_Inglés; etc.).
- La traducción entre las distintas variantes de VILA_1 es automática e inmediata.
- El aprendizaje de VILA_1 es extremadamente sencillo.

CAPÍTULO 2

Objetivos, justificación y estructura de la tesis

2.1 Justificación

El grupo de investigación dirigido por el Dr. Don Ángel Alonso Álvarez ha desarrollado un lenguaje visual, denominado VILA_1, como se introdujo en el capítulo anterior.

Las características de VILA_1, son muy diferentes a las del resto de los lenguajes visuales existentes en la actualidad, para soportar las aplicaciones basadas en estas características hay que desarrollar una base de datos especial, con atención especial a los procesos de comunicación y traducción, esta necesidad es la que justifica la investigación de esta tesis doctoral.

2.2 Objetivos

El objetivo de la tesis es crear una base de datos que relacione eficazmente las palabras del español y el chino de acuerdo con las características del lenguaje VILA_1

Para cualquier el objetivo citado se realizarán las siguientes frases:

- Estudiar la estructura y características de VILA_1.
- Crear un formato serializado para la representación y transmisión a través de Internet de VILA_1.
- Estudiar y diseñar una base de datos que puede incluir el vocabulario en varios idiomas para la comunicación mediante el lenguaje VILA_1. La arquitectura incluirá especificaciones para:
 - La selección de los lenguajes más utilizados.
 - La división de los elementos del lenguaje natural.
 - La división de los dominios del vocabulario del lenguaje natural.
 - El formato de las palabras del vocabulario.
 - La traducción entre palabras.
- Desarrollar un prototipo que implemente la arquitectura y formato mencionados anteriormente con el fin de comprobar las metodologías y procesos adecuados para su construcción, así como su viabilidad.
- Diseñar y llevar a cabo un proceso de validación utilizando el prototipo mencionado en el punto anterior.

2.3 Metodología de trabajo

La presente tesis se ha elaborado según la metodología presentada en [9], cuyos pasos principales son: determinación del problema, formulación de la hipótesis, validación de la hipótesis y análisis de resultados, los cuales se detallan a continuación.

- Determinación del problema (Capítulo 3). Este trabajo surgió tras la creación del lenguaje VILA_1, y la necesidad de desarrollar una base de datos que soporte las aplicaciones. Después de la identificación del problema, se estudio la literatura más relevante del los lenguajes, con especial atención en los Lenguajes Visuales, de la Lingüística Computacional, de las diferentes metodologías existentes en la traducción automática, y de las Arquitecturas de base de datos.
- Formulación de la hipótesis (Capítulo 4). Así pues, la hipótesis de la que se partió en esta tesis es que un lenguaje visual con las características de VILA_1 necesita, para poder ser aprovechado, un diseño de la base de datos que incluye el vocabulario en varios idiomas para soportar las aplicaciones, por ejemplo comunicación o traducción.
- Validación de la hipótesis (Capítulo 5). Partiendo de la hipótesis formulada en el capítulo anterior, se definió un proceso de validación empírica, que desarrolló una probación y evaluación en el que se realizaron pruebas con la base de datos desarrollada en la Universidad de León.
- Análisis de resultados (Capítulo 6). Se analizaron los distintos resultados obtenidos en el desarrollo de la presente tesis, estableciendo una serie de conclusiones y líneas futuras.

2.4 Estructura de la tesis

A continuación se realizará un recorrido global por los capítulos que conforman esta memoria.

- En el capítulo 3, se presenta una síntesis del estado del arte necesario para acometer esta tesis. Incluye aspectos de los principales temas o áreas de conocimiento relacionados con la investigación, de la siguiente manera:
 - Lenguajes, prestando especial atención a los denominados lenguajes visuales.

- Aspectos metodológicos de la base de datos, necesarios en el desarrollo de este proyecto.
- Traducción automática, así como los formatos serializados existentes en la actualidad, donde se analizarán sus aspectos generales, tipos de servicios existentes, arquitecturas, y metodologías.

Es importante señalar, que de cada uno de los temas presentados, se hará una breve valoración y reseña de por qué han sido incluidos en esta tesis, y su relación directa con ella. Adicionalmente, para terminar este capítulo se presentarán las conclusiones más importantes del estado del arte.

- En el capítulo 4, se especifican los requerimientos y el diseño de la base de datos, y su funcionamiento más eficiente para soportar la aplicación de VILA_1, así como las distintas fases de desarrollo del prototipo que lo implementa.
- En el capítulo 5, se presentará el proceso de validación empírico de esta tesis doctoral, a través de una evaluación entre un grupo de participantes, cuyo objetivo principal es identificar las ventajas y también los problemas prácticos del proceso de traducción automática basado en el lenguaje VILA_1 gracias a las estructuras definidas mediante una base de datos relacional.
- En el capítulo 6, se presentarán todas las conclusiones que se han obtenido al terminar esta investigación y los trabajos que se podrán iniciar partiendo de los resultados de la misma.
- Para finalizar, se mostrará la lista de abreviaturas usadas en esta memoria. Después, se señalarán las referencias bibliográficas empleadas durante la investigación.

CAPÍTULO 3

Estado del arte

3.1 Introducción

Como fundamento de investigación para esta tesis doctoral, se debe partir de una serie de tecnologías claves para el planteamiento de las consideraciones metodológicas que permitan alcanzar los objetivos descritos en el capítulo anterior.

Las tecnologías clave de las que tratará este estado del arte son las siguientes:

- Lenguajes, con especial incidencia en su clasificación, características y en sus diferentes metodologías de procesamiento, prestando especial atención a los denominados lenguajes visuales.
- Aspectos metodológicos de la base de datos, necesarios en el desarrollo de este proyecto, para establecer con la mayor corrección posible las especificaciones, el diseño, el desarrollo

y la validación de la arquitectura de la base de datos que soporta las aplicaciones de VILA_1.

- Aspectos arquitectónicos y de funcionamiento de la base traducción automática, así como los formatos serializados existentes en la actualidad, donde se analizarán sus aspectos generales, tipos de servicios existentes, arquitecturas, y metodologías, para así poder enseñar los problemas actuales y las ventajas de VILA_1, y cómo solucionar estos problemas con otra forma en VILA_1.

Este capítulo está estructurado de la siguiente forma; la sección 3.2 introduce el concepto de lenguaje, así como la distinción entre su naturaleza formal o natural, se tratan temas como la gramática, expresividad, y computabilidad, realizando un repaso a la Jerarquía de Chomsky. A continuación en el apartado 3.3 se detallan los aspectos más importantes de la lingüística computacional, su evolución histórica, y sus principales aplicaciones. En la sección 3.4 se realiza un repaso del estado del arte de los Lenguajes Visuales, atendiendo a su clasificación, y aplicaciones principales. La sección 3.5 contiene la literatura más relevante de la base de datos, atendiendo a sus diferentes metodologías, y etapas de desarrollo. La sección 3.6 se encarga de realizar un repaso sobre la traducción automática, enumerando las arquitecturas más importantes existentes en la actualidad, así como sus ventajas e inconvenientes. Finalmente en el apartado 3.7 se enumeran las conclusiones que se han obtenido del estudio del presente estado del arte.

Después del análisis de la estructura de este capítulo, resulta fundamental comenzar estableciendo los aspectos más relevantes de los distintos tipos de lenguajes, debido a que uno de los objetivos de esta tesis doctoral es estudiar la estructura y características de VILA_1, y para ello es necesario conocer su clasificación.

3.2 Definición de un lenguaje

Los lenguajes son sistemas más o menos complejos, que asocian contenidos de pensamiento y significación a manifestaciones simbólicas tanto orales, como escritas [10]. Los lenguajes están compuestos por secuencias de símbolos, pertenecientes a alguna colección finita. Por ejemplo, en el caso de los lenguajes naturales, la colección finita es el conjunto de letras del alfabeto, junto a los símbolos que se utilizan para construir palabras, y para el caso de la representación de números naturales, serán secuencias de caracteres de los dígitos $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Un conjunto no vacío y finito de símbolos se conoce como alfabeto. Si Σ es un alfabeto, y $\sigma \in \Sigma$ quiere decir que σ es un símbolo de Σ . Por lo que, si $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, podemos decir que $0 \in \Sigma$. Entonces si un alfabeto es simplemente un conjunto finito no vacío, dados Σ_1 y Σ_2 alfabetos, se tiene que $\Sigma_1 \cup \Sigma_2$ también lo es, incluso $\Sigma_1 \cap \Sigma_2$, $\Sigma_1 + \Sigma_2$, y $\Sigma_1 - \Sigma_2$ también son alfabetos.

Una secuencia finita de símbolos en un determinado alfabeto, se denomina palabra de dicho alfabeto. Cada símbolo de un alfabeto, es una cadena del mismo. La cadena vacía, es una palabra sobre cualquier alfabeto.

Los elementos más simples de un lenguaje, son los símbolos que constituyen el alfabeto Σ , que es el conjunto finito de símbolos $\{\sigma_1, \sigma_2, \sigma_3 \dots \sigma_n\}$. Con la concatenación de los símbolos, se formarán palabras que determinarán un conjunto Σ^* . Por lo tanto, el conjunto de palabras que tengan significado, constituirán el diccionario del lenguaje, así que un lenguaje se considera como el conjunto de oraciones, que usualmente es infinito, y que se forman con palabras del diccionario. Se pueden por tanto distinguir dos clases de lenguajes; los lenguajes naturales y los lenguajes formales, y una subclase híbrida; los lenguajes semiformalizados, que se encontrarían a medio camino entre los lenguajes naturales, y los lenguajes formales.

El origen y desarrollo de los lenguajes naturales; como por ejemplo el español, el chino, el inglés, etc. fue natural, es decir sin el control de ninguna teoría, por lo tanto sus gramáticas, como sus teorías, fueron establecidas después de que los lenguajes ya se hablaran, y hubieran madurado. Por otro lado, los lenguajes formales como las matemáticas, la lógica, etc. fueron desarrollados generalmente a través de una teoría, que sentaba las bases de los mismos.

Como se puede observar en la (Figura 3.1), el lenguaje VILA_1 se encuentra entre los lenguajes formales, y los lenguajes naturales, esto es debido a que utiliza por un lado una sintaxis propia, específica y formal, y por otro la semántica del lenguaje natural, por todo esto se puede definir como un lenguaje semiformal.



Figura 3.1 Tipos de lenguajes. Clasificación de VILA_1.

En el lenguaje natural, la formación de oraciones precedió a la formalización del lenguaje por medio de una teoría o una gramática, por lo tanto un lenguaje es denominado “natural”, porque es no artificial o no construido. El calificativo de “natural”, se opone al de “formal” el cual determina un lenguaje que se ha construido a partir de una teoría, y por lo tanto sería artificial.

Los principales términos que se usan para describir un lenguaje son; *Gramática*, que se encarga de definir los usos correctos de una lenguaje mediante preceptos. *Sintaxis*, es la parte de la gramática que enseña a coordinar y unir las palabras para formar las oraciones y expresar conceptos. *Semántica*, es lo perteneciente o relativo a la significación de las palabras. *Pragmática*, es la disciplina que estudia

el lenguaje en su relación con los usuarios y las circunstancias de la comunicación [11].

En los lenguajes naturales, las palabras que forman una oración poseen un significado¹, y tienen un significante², es decir, que independientemente del significado de cada palabra, debemos tener en cuenta el sentido correcto de las mismas, según el contexto. Uno de los objetivos de las ciencias de la computación, es poder especificar rigurosamente estos significados, por los métodos de interpretación de los sistemas formales, métodos que constituyen las semánticas del lenguaje formal.

Los lenguajes naturales se distinguen significativamente de los lenguajes formales, tanto por su origen, como por su área de aplicación.

3.2.1 Lenguajes formales

El proceso de generación y desarrollo de un lenguaje formal, es justo el inverso de los lenguajes naturales, ya que como se señaló en la sección anterior, fueron precedidos por su definición, y por el establecimiento de una serie de propiedades que permiten su definición. Las palabras y las oraciones de los lenguajes formales, están perfectamente definidas, y estas palabras mantienen el significado independientemente del contexto. Las gramáticas que definen este tipo de lenguajes, son las gramáticas libres del contexto³. El significado de los símbolos, está determinado exclusivamente por la sintaxis, sin tener en cuenta la semántica, ya que la única semántica existente será la que proporcionen sus operadores. En resumen, los lenguajes formales, se caracterizan por las siguientes propiedades [12]:

¹ Según (RAE, 2001), es el contenido semántico de cualquier tipo de signo, condicionado por el sistema y el contexto.

² Según (RAE, 2001), es el fonema o secuencia de fonemas que, asociados con un significado, constituyen un signo lingüístico.

³ Las gramáticas libres del contexto tienen mucha relevancia, debido a su conveniencia para la realización de analizadores sintácticos. Sin embargo, las gramáticas libres del contexto tienen algunas limitaciones, ya que no pueden modelar todos los fenómenos gramaticales presentes en los lenguajes naturales.

- Se desarrollan a partir de una teoría establecida.
- Tienen un componente semántico mínimo.
- Su sintaxis produce oraciones no ambiguas.
- Su formalización es completa.
- Poseen un gran potencial computacional.

3.2.2 Lenguajes naturales

Los lenguajes naturales tienen un gran poder expresivo, y pueden ser utilizados para analizar situaciones altamente complejas, y razonar sutilmente, todo esto es debido a la riqueza de su componente semántico, y su cerrada relación con los aspectos prácticos de los contextos en los que son usados [13].

La formalización del componente semántico de un lenguaje natural, es muy compleja, sin embargo la sintaxis del mismo, puede ser fácilmente modelada por un lenguaje formal, similar a los utilizados en las matemáticas o en la lógica. La polisemia es otra propiedad única de los lenguajes naturales, es decir la posibilidad de que una palabra, tenga distintos significados. El carácter polisémico de los lenguajes naturales, incrementa su componente semántico, lo que hasta el momento ha hecho imposible su formalización.

Después de lo anterior, se puede concluir que los lenguajes naturales se caracterizan por las siguientes propiedades:

- Desarrollados por enriquecimiento progresivo, antes de la formación de una teoría.
- La importancia de su carácter expresivo, es debido a su riqueza semántica.
- Dificultad o imposibilidad de una formalización completa.

El área que estudia el Procesamiento del Lenguaje Natural, y que por tanto es objeto de estudio en este estado del arte, es la Lingüística Computacional, y se detalla a continuación.

3.3 Lingüística Computacional

Los primeros equipos informáticos tenían como función principal, el cálculo matemático, sin embargo, gracias a los avances realizados tanto de software, como de hardware, la informática es aplicable a una amplia gama de áreas, entre las que se encuentra la lingüística computacional [14].

La lingüística computacional es un campo multidisciplinar, que combina la lingüística y la informática para el estudio y el procesamiento del lenguaje natural, desde un punto de vista computacional. Algunas de las aplicaciones más destacadas de la lingüística computacional son [15]:

- Apoyo en la preparación de documentos.
- Recuperación y almacenamiento de la información.
- Reconocimiento y síntesis del habla.
- Traducción automática.

En las siguientes páginas se realizará un análisis del estado del arte de la lingüística computacional.

3.3.1 Antecedentes históricos

El procesamiento del lenguaje natural se remonta hasta la década de 1940, en plena Segunda Guerra Mundial para el cálculo de tablas de artillería y criptoanálisis, aunque al final de la Guerra los investigadores se dedicaron a otro tipo de aplicaciones. Hasta el año 1960 no resulta demasiado difícil recopilar trabajos, debido a la pequeña cantidad de los mismos, sin embargo desde ese momento se han realizado muchas investigaciones en esta materia.

Entre los primeros estudios sobre el procesamiento del lenguaje natural se encuentran las propuestas para la mecanización de la traducción del lenguaje, que son anteriores a la invención de los ordenadores digitales. La primera aplicación reconocible fue un

diccionario desarrollado por el Birkbeck College. Más adelante, surgieron grupos de desarrollo en el campo de la traducción con máquinas en Estados Unidos, Reino Unido, Francia y la Unión Soviética. Los primeros investigadores de esta área eran matemáticos que trabajaban con las primitivas máquinas de computación de la época, en un principio estos científicos eran bilingües, y se suponían capaces de escribir programas que pudieran traducir al menos textos técnicos, pero pronto se dieron cuenta que era una tarea extremadamente difícil, así que pidieron ayuda a los lingüistas, y fue cuando a partir de los años 50 se produjo un importante crecimiento de los trabajos en traducción automática.

En 1957, Noam Chomsky publica *Syntactic Structures* [16], donde introduce la idea de la gramática generativa, que consiste en la descripción de las estructuras sintácticas basándose en reglas. Algunos autores no están de acuerdo con las teorías de Chomsky, pero es importante destacar como su idea revolucionó la lingüística computacional, ya que casi todos los trabajos a partir de 1957 han sido realizados bajo su influencia.

A mediados de los años 60 las inversiones públicas estadounidenses para el desarrollo de la traducción automática alcanzaban los 20 millones de dólares, y el Comité Asesor para el Procesamiento Automático del Lenguaje (ALPAC: Automatic Language Processing Advisory Committee) realizó un informe sobre los resultados de dicha inversión, concluyendo que “no existe una máquina de traducción de textos científicos y no se espera que aparezca ninguna a corto plazo” [17].

3.3.2 Procesamiento del lenguaje natural desde 1960 hasta 1980

Muchos expertos coinciden en que las investigaciones relacionadas con el Procesamiento del Lenguaje Natural fueron muy cuestionadas a

raíz del informe ALPAC⁴. Hay que señalar algunos proyectos que siguen teniendo cierta relevancia hoy en día, entre ellos destacan los siguientes:

- **Redes de Transición Aumentadas.** Este tipo de redes son un sistema para representar el conocimiento que permiten representar gramáticas complejas para procesar su sintaxis. Este sistema supuso en su momento un gran avance a la hora de representar el conocimiento de una manera formal en un dominio en concreto [18].
- **Gramática de Casos.** Las gramáticas de casos están basadas en la información que va asociada en unos casos con el uso de preposiciones, conectores, etcétera y en otros en la propia información de la palabra, o de la posición que ocupan en una determinada oración. Por ejemplo, algunos lenguajes como el español, expresan las relaciones entre verbos y nombres por medio de preposiciones de concatenación [19].
- **Representaciones Semánticas.** Durante este periodo se realizaron numerosos avances en el procesamiento semántico, también se introduce la idea de red semántica que está constituida por nodos y enlaces; este tipo de redes son usadas para la representación del conocimiento en muchos sistemas. William Woods empleó la idea de semántica procedimental como una representación intermedia entre el sistema de procesamiento del lenguaje y un sistema de bases de datos [20]. Los sistemas más destacados son:
- **SHRDLU.** El sistema *SHRDLU* realizaba la simulación de un robot que movía bloques sobre una mesa, su importancia radicaba en que las instrucciones estaban escritas en lenguaje natural, y que con una combinación de sintaxis, semántica y razonamiento, éste lenguaje podía ser procesado [21].

⁴ En 1966, la Sección de Ciencias del Comportamiento de la Academia Nacional de Ciencias norteamericana pidió un informe sobre el estado de la traducción automática y sus perspectivas de futuro. Para ello se formó el Comité ALPAC (Automatic Language Processing Advisory Committee). El informe emitido, Informe ALPAC, fue tan demoledor que durante los siguientes años se fueron denegando todos los fondos oficiales para investigaciones dedicados a este fin, hasta que en 1975 quedaron reducidos a cero.

- *LUNAR*. Es un sistema de interfaz con bases de datos, que trabaja con Redes de Transición Aumentadas y la semántica procedimental de Woods, respecto a su nombre señalar que viene de una base de datos que contenía información sobre la Luna [22].
- *LIFER/LADDER*. El sistema *LIFER/LADDER* fue diseñado como una interfaz de lenguaje natural con una base de datos de los barcos de marina estadounidense. Usa una gramática orientada a la semántica, y se limita a trabajar en un dominio restringido al igual que *SHRDLU* [23].

3.3.3 Procesamiento del lenguaje natural desde 1980

A partir de 1980 el número de estudios relacionados con Lingüística Computacional aumentó de una manera importante. Aparecieron infinidad de formalismos gramaticales para la representación del lenguaje natural, que permitían su representación y su tratamiento. En los próximos apartados se hará referencia a las investigaciones más relevantes, pero antes de continuar es importante señalar algunas características distintivas de esta nueva etapa [24]:

- Se abandonan las especulaciones lingüísticas, obteniendo vital importancia el desarrollo de recursos a gran escala, especialmente léxicos y gramáticas, con información básica, independiente de cualquier aplicación.
- Desarrollo de metodologías basadas en el análisis de corpus lingüísticos, que pasan a ocupar un lugar privilegiado, ya que constituyen una fuente de información valiosa para la creación de diccionarios, léxicos computacionales y gramáticas, se utilizan como recurso en los sistemas de anotación automática, se utilizan como banco de pruebas para los sistemas de procesamiento del lenguaje, etc. Como resultado, aparece una nueva disciplina, denominada la Lingüística del Corpus [25].

- Aplicación de criterios de estandarización de los datos lingüísticos que faciliten su reutilización e integración. En esta línea se han promovido algunas acciones como; *TEI* (Text Encoding Initiative) [26], y el programa *EAGLES* (European Advisory Group for Language Engineering Standards), etc. que definen códigos estándar para el etiquetado de corpus textuales, y de voz, léxicos y formalismos para el análisis del lenguaje [27].
- Desarrollo de aplicaciones con un claro interés práctico.
- Se promueven proyectos para la integración de los recursos desarrollados, entre los que destacan; *Interval* [28], que está orientado a la validación de recursos terminológicos multilingües, esto es debido fundamentalmente a que la mayoría de los textos a los que se aplican programas de traducción automática, y de recuperación de la información sean de áreas temáticas especializadas, por lo que existe un gran interés en recopilar y estandarizar sus terminologías específicas. *Parole*, dedicado a la recopilación y codificación de corpus y léxicos para 14 lenguas: alemán, catalán, danés, finlandés, friego, español, francés, holandés, inglés, italiano y portugués, y cuyo objetivo principal es la codificación de estos recursos según unos estándares comunes de modo que puedan servir de referencia a la comunidad que trabaja en procesamiento de lenguaje [29]. *Speechdat*, orientado al tratamiento de voz [30]. *EuroWordNet*, cuyo objetivo es la construcción de redes léxico-semánticas de diversas lenguas, este proyecto se basa en una ontología léxico-conceptual para la lengua inglesa desarrollada en la Universidad de Princeton, y de amplia utilización en diversos sistemas de tratamiento de la lengua [31].

Es importante destacar, como la conjunción de los cinco factores anteriores, ha incidido en el rumbo que tomaron las aplicaciones tradiciones de la Lingüística Computacional, y como han surgido nuevas aplicaciones, trasladándose el grueso de la investigación en

este campo hacia el desarrollo de métodos de adquisición de datos y conocimiento lingüístico que hagan posible el desarrollo de los recursos básicos necesarios en la Sociedad del Conocimiento⁵.

3.3.3.1 Aplicaciones tradicionales

Las aplicaciones tradicionales de la Lingüística Computacional, como son las interfaces en lenguaje natural y la traducción automática, deberán tratar dominios más generales, y no restringirse a unos campos muy específicos. Esto es debido a que actualmente se exige, además, de que estas aplicaciones sean competitivas, que funcionen en tiempo real [32].

Los sistemas tradicionales de traducción automática tenían como usuarios a grandes empresas e instituciones que necesitaban traducir documentación técnica, manuales, etc. Estos sistemas necesitaban unas importantes y sofisticadas infraestructuras para su funcionamiento, sin embargo actualmente, existen cientos de productos software para la traducción que funcionan sobre ordenadores personales, algunos de los más populares son *Babylom* [33], *SYSTRAM* [34], y *PROMT* [35]. Es importante destacar también como muchos sistemas de traducción son accesibles desde páginas Web, correo electrónico, etc.

En esta línea de trabajo es importante destacar algunos proyectos como; *EURAMIS* (European Advanced Multilingual Information System) [36], cuyo objetivo principal es la creación de una plataforma para la traducción automática basada en memorias de traducción. *Multimeteo* [37], que se dedicaba a generar informes meteorológicos en diferentes idiomas, y así tratar de dar respuesta a las necesidades

⁵ El término “Sociedad del Conocimiento” surge en el ámbito de las organizaciones empresariales a principios de los años 1970 aunque la idea no fue desarrollada hasta los años 1990, haciendo referencia a la importancia del conocimiento o “valor intangible” que radica en cada una de las personas que forman una organización. Bajo este punto de vista el término se relaciona directamente con el de “gestión del conocimiento”. El término también aparece algunas veces citado en relación con el denominado proceso de globalización de las economías; de este modo la sociedad del conocimiento será aquella en la que el acceso al mismo sea universal. Ambos puntos de vista son compatibles con la acepción más general expresada aquí.

que tienen diferentes tipos de usuarios, como por ejemplo los profesionales de la agricultura, del turismo, etc.

3.3.3.2 *Nuevas aplicaciones*

La aparición de Internet como red de telecomunicaciones y de la WWW ha cambiado el concepto de lo que se considera información, y su modo de tratarla; es famosa la expresión “*no está mejor informado quien más datos tiene, sino quien dispone de los mejores medios para obtener sólo y exclusivamente aquellos que necesita*”.

La recuperación de la información, es una cuestión que hasta hace poco tiempo afectaba a colectivos de profesionales muy concretos, pero que ha pasado a ser uno de los problemas más graves de la Sociedad del Conocimiento [38].

El problema de la recuperación de información se centra en dos aspectos fundamentalmente: el filtrado de los datos y el multilingüismo. En el primer caso se trata de diseñar sistemas de recuperación de información que recuperen solo documentos interesantes al usuario, y que por tanto eliminen el “ruido”, en el segundo caso resulta necesario dotar a los sistemas de recursos de capacidad de acceso multilingüe a los datos [32].

Estas necesidades de la Sociedad del Conocimiento, están incidiendo de manera directa en las directrices de la Lingüística Computacional, señalando como unos de sus líneas principales de investigación el tratamiento de textos lingüísticos no restringidos. La confección de resúmenes es una aplicación directamente relacionada con la recuperación de información, tratando de producir una representación concisa para el lector que contenga lo esencial del documento, para lo que se aplican técnicas de Inteligencia Artificial. Otra línea importante de investigación, es la confección de textos de documentación técnica mediante un lenguaje controlado, y con una serie de restricciones, como por ejemplo el uso de alfabetos reducidos.

Las nuevas directrices que han adoptado las aplicaciones tradicionales de lingüística computacional, así como las nuevas opciones que están emergiendo, requieren disponer de recursos lingüísticos a gran escala, y para el mayor número de lenguas posible.

3.3.3.3 Nuevos recursos

La construcción de gramáticas generales de la lengua requiere dispone de la información necesaria sobre el uso real de la misma por parte de los hablantes: los léxicos y las reglas gramaticales que las componen deberán procesar textos reales, no las frases que aparecen como ejemplos en los tratados de lingüística. Por este motivo resulta necesario el desarrollo de sistemas automáticos y semiautomáticos de adquisición de conocimiento lingüístico.

Los corpus y los diccionarios constituyen las fuentes fundamentales de información sobre las lenguas. Los corpus contienen muestras efectivas del uso que los hablantes hacen de la lengua, y los diccionarios contienen datos semicodificados sobre el lenguaje que facilitan el proceso de extracción.

Los corpus

Un corpus lingüístico es un conjunto de textos almacenados en formato electrónico y agrupado con el fin de estudiar un lenguaje o una determinada variedad lingüística. Su objetivo principal es constituirse en elementos de referencia para el estudio de una frase concreta o un cierto aspecto de la lengua [39].

Históricamente se puede decir, que el origen de los corpus está en las colecciones de textos se empleaban ya en el siglo XIX, y con anterioridad para describir los cambios en una lengua, para así poder justificar enunciados gramaticales, para documentar la adquisición de una lengua, elaborar diccionarios o para realizar comparaciones entre

diversas lenguas. Los primeros corpus electrónicos aparecieron durante los años 60, y están formados por textos almacenados de forma digital, y agrupados con el fin de estudiar un idioma específico o una variedad lingüística.

Existen diferentes tipos de corpus, que se pueden clasificar en; corpus orales, que son los corpus están formados por señales de voz, normalmente con sus transcripciones de anotación fonética, suelen estar formados por grabaciones telefónicas, entrevistas, programas de radio, etc. Bases de datos de árboles, que están formados por textos etiquetados sintácticamente. La mayoría de los análisis de este tipo tienen estructura de árbol, aunque en algunos casos más complicados su estructura es de grafo. Algunos ejemplos son el Spanish TreeBank (Universidad Autónoma de Madrid), o el ISST (Italian Syntactic-Semantic TreeBank) [40]. Corpus de textos. Los corpus textuales están formados por lengua escrita, este tipo de corpus son los que más habituales, ya que son los más fáciles de elaborar [41].

Los corpus se pueden aplicar en numerosas áreas de la lingüística, en este apartado vamos a realizar un repaso por las más relevantes [39]; sintaxis, semántica, fonología, lexicografía, y lingüística computacional.

- Sintaxis. Gracias a un corpus etiquetado sintácticamente, se puede verificar una determinada construcción sintáctica.
- Semántica. En el caso de la semántica léxica, un corpus sirve para revisar como se utiliza una determinada palabra, y que sentido y significado tiene.
- Fonología. La fonología se encarga de estudiar cómo se pronuncian los extranjerismos, que gracias a la prosodia⁶ es posible distinguir las lecturas de una palabra, su clasificación, acentos, etc.

⁶ La prosodia es una rama de la lingüística que analiza y representa formalmente aquellos elementos de la expresión oral, tales como el acento, los tonos y la entonación.

- Lexicografía. A la hora de desarrollar un diccionario, los lexicógrafos analizan en los corpus, los contextos en los que aparece una palabra determinada, que palabras no se utilizan, que lecturas existen de una determinada palabra, etc.
- Lingüística computacional. Los lingüistas computacionales utilizan los corpus como recurso para la elaboración automática de diccionarios, para la extracción de los corpus bilingües para elaborar memorias de traducción, para la extracción automática de colocaciones, para la extracción automática de diferencias del lenguaje en todos los niveles (sintaxis, semántica, etc.), como por ejemplo entre el español de España, y el español de Latinoamérica, etc.

Los diccionarios procesables por el ordenador

Un MDR (Machine Dictionary Readable) es un tipo especial de diccionario que puede ser procesable por el ordenador, que puede considerarse como una determinada variedad de corpus lingüístico, esto es debido a que posee unas características muy específicas que los distinguen claramente del resto; han sido elaborados con una clara finalidad: codificar información sobre el léxico de la lengua, tienen una estructura interna predeterminada, existen un cierto grado de codificación en determinados contenidos, contienen relaciones léxicas internas, vocabularios restringidos, etc. [42].

Aunque la información presente en un diccionario es variada en cuanto a forma y contenido, responde a unos patrones estables, lo que ha permitido definir modelos computacionales para su tratamiento.

3.4 Lenguajes Visuales

La comunicación compleja es la seña de identidad de los seres humanos. Cuando se piensa en comunicación entre personas, la mayoría de la gente piensa en la comunicación verbal, y en los

lenguajes escritos. Estos dos tipos de lenguajes son similares, si se tiene en cuenta como son procesados sus símbolos, ya que en ambos casos su procesamiento es secuencial, ya sea en la temporalidad de los lenguajes hablados, o en la secuenciación de las páginas en los lenguajes escritos, sin embargo, no todas las comunicaciones entre humanos son secuenciales por naturaleza. Otro componente importante de las comunicaciones entre humanos son los lenguajes visuales, como los mapas o los diagramas, en estos lenguajes sus símbolos no están ordenados secuencialmente, sino que se encuentran más bien agrupados [43].

Los lenguajes visuales se asocian a la cultura humana desde sus orígenes, desde las pinturas de los aborígenes de los nativos australianos, hasta la utilización de las señales de tráfico actuales, en definitiva los lenguajes visuales han sido utilizados desde la prehistoria hasta el presente, en todas las actividades humanas. Los lenguajes visuales cubren un amplio abanico de la expresión humana, partiendo desde las obras de arte, hasta los mapas de carreteras [44].

Los lenguajes visuales se han convertido en un componente clave de la interacción hombre-máquina, y están abocados a aumentar su importancia todavía más, entre otras cosas por el importante crecimiento de Internet.

Por otro lado es importante destacar como desgraciadamente los aspectos teóricos de los lenguajes visuales no se entienden igual de correctamente, que los relacionados con el resto de tipos de lenguajes, ya que no existen desarrollos como por ejemplo el de la teoría lingüística moderna, que posee un amplio desarrollo teórico, o el de la lingüística computacional visto anteriormente. Esto puede ser debido en gran parte a que las investigaciones que se realizan sobre lenguajes visuales, se producen en áreas y comunidades muy distantes científicamente, y que en la mayoría de los casos se desconocen mutuamente.

Las investigaciones sobre lenguajes visuales son muy variadas [43], aunque la mayoría de los grupos coinciden en la necesidad de lograr una clasificación de los lenguajes visuales, y en establecer directrices para su diseño, y estructuración, tanto sintáctica, como semántica. Una de las motivaciones principales de las investigaciones de los lenguajes visuales se centra en facilitar la interacción entre seres humanos y máquinas, y en conseguir automatizar el procesamiento, y razonamiento de su contenido.

La razón principal para el uso de lenguajes visuales, es que estos lenguajes son mucho más eficientes en las comunicaciones entre personas, que los lenguajes textuales tradicionales. También es importante entender que el lenguaje escrito, es un caso concreto del lenguaje visual, en realidad podemos pensar en múltiples lenguajes visuales con reglas sintácticas muy específicas. Por ejemplo las señales de tráfico o la notación musical proporcionan un lenguaje visual con significados muy precisos. También podemos hablar del lenguaje visual de los gráficos de barras y de pastel o de los organigramas corporativos.

En el caso de las señales de tráfico, la combinación de unas pocas figuras geométricas como triángulos, cuadrados y círculos con unos colores determinados y un conjunto de símbolos como coches, bicicletas, etc. permite expresar un amplio surtido de prohibiciones, obligaciones y situaciones de forma inequívoca y perceptible en fracciones de segundo.

Es importante aclarar lo que es un lenguaje visual, ya que en su literatura se manejan definiciones muy diferentes, y resulta necesario proporcionar una definición consensuada que pueda servir como base teórica para aglutinar las distintas áreas de investigación [43].

Se considera un Lenguaje Visual, a todos los conjuntos de diagramas, y colecciones de símbolos en dos o tres dimensiones que forman frases válidas, y coherentes con las

relaciones espaciales entre los distintos símbolos de un idioma. El significado de una frase, en general está constituido por los símbolos gráficos utilizados en la oración y su disposición espacial.

Después de esta definición, y para finalizar la introducción de esta sección, resulta importante destacar que la mayoría de los proyectos de investigaciones existentes, y la literatura relacionada con los lenguajes visuales se centra en dos ámbitos; los lenguajes visuales de programación, y los lenguajes visuales aplicados a la comunicación.

3.4.1 Aplicaciones de los lenguajes visuales

Cuando se investiga en la razón principal para el uso de lenguajes visuales, la mayoría de las investigaciones coinciden en que su fin es permitir la comunicación y la interacción visual, siempre desde la perspectiva de la interacción hombre-computadora. Los lenguajes visuales abarcan una amplia variedad de entidades que han sido objeto de investigación en diferentes áreas de conocimiento, aunque su filosofía sea la misma.

A continuación se presentan los términos más relevantes relacionados con los lenguajes visuales, haciendo especial énfasis en los lenguajes visuales de programación, y en los lenguajes visuales aplicados a la comunicación, para ello se han seleccionado una serie de términos que contienen la palabra “visual” en la literatura, pero con significados diferentes [45].

- Lenguajes visuales. Son lenguajes con alfabetos consistentes habitualmente en representaciones gráficas, que son usados para comunicaciones humano-humano o humano-ordenador.
- Programación visual. El uso de representaciones gráficas para comunicar datos y operaciones a un ordenador.

- Lenguajes de Programación Visual. Un lenguaje de programación visual con un alfabeto compuesto de representaciones visuales.
- Visualización software. La generación y la representación estática y dinámica de las diferentes características del software, incluyendo estructuras, código, algoritmos, etc.
- Animación de algoritmos. La generación y visualización de animaciones utilizando representaciones dinámicas visuales transmiten a los humanos como operan los algoritmos.
- Representación de diagramas. Consisten en representaciones visuales que permiten la asociación y transmisión de información, mediante la representación de sus referentes.
- Razonamiento utilizando esquemas. El proceso de transmisión y realización de inferencias a partir de representaciones de diagramas y esquemas.

La idea principal en la teoría de los lenguajes visuales es establecer el diseño y la caracterización tanto computacional, como cognitiva de los procesos informáticos que utilizan lenguajes visuales [46], ya que debido a que no existe una teoría, se utilizan reglas lógicas, y del lenguaje natural para establecer sus propiedades y características, lo que implica una importante reducción del potencial de los lenguajes visuales [45].

Después de lo dicho, y para entender mejor los lenguajes visuales, se debe revisar el uso de los mismos por las personas y por los ordenadores, a la hora de comunicarse, interactuar y realizar razonamiento. Para ello debemos de atender a los siguientes usos:

- Manipulación directa. La manipulación directa permite a sus usuarios ejecutar acciones directas interactuando de manera visual con los distintos objetos, sin la necesidad de describir acciones. Por ejemplo cuando se quiere marcar un archivo para su eliminación en una interfaz gráfica, el usuario simplemente tiene que arrastrar el icono del archivo, hasta el icono que

representa la papelera, lo que ejecuta internamente el comando “mover <archivo> papelera”. Con la manipulación directa el usuario puede “representar” la operación que desea, sin la necesidad de escribir un comando que interpretara un intérprete, y que tradujera en una acción [47].

- Visualización de información. Las representaciones visuales se utilizan a menudo para hacer que las relaciones entre conceptos, sean más accesibles a la intuición humana [48]. Un gráfico de una función es mucho más explícito, e intuitivo que una tabla de valores de la misma función, a pesar de ser representaciones equivalentes. Habitualmente se utilizan ilustraciones visuales, para ejemplificar las ideas que contiene un determinado texto, o para complementar su información. La visualización de la información es un área asociada con el desarrollo de lenguajes visuales, y más concretamente con la transmisión de las distintas estructuras visuales a los seres humanos, permitiendo la comunicación a través de manipulación directa.
- Visualización de software. Visualizar las características estáticas y dinámicas de los programas puede facilitar la comprensión y depuración para el usuario [49]. Tanto las estructuras de datos, como las herramientas de control de flujos son más fáciles de entender cuando se visualizan mediante gráficos y diagramas de flujo. La estructura jerárquica de archivos de un sistema operativo es fácil de visualizar con un árbol de carpetas y directorios. La visualización del software es esencialmente la inversa de analizar un lenguaje de programación visual.
- Representación de diagramas y razonamiento. La mayoría de las investigaciones sobre la teoría de los lenguajes visuales, están caracterizadas por la búsqueda de especificaciones semánticas y sintácticas de los lenguajes visuales utilizando descripciones proposicionales, lo cual facilita la ejecución de procesos de razonamiento sobre estos lenguajes. Uno de los motivos principales por los que los lenguajes visuales están

presentes en sectores como el comercial, es debido a que facilitan la representación de la información de una manera sencilla y compacta, lo que es difícil de escribir de manera adecuada utilizando descripciones proposicionales [50].

- Simulación gráfica. La noción de diagrama de razonamiento se combina con la manipulación directa en las simulaciones interactivas, convirtiéndose en una herramienta importante en la investigación educativa [51]. Las representaciones dinámicas visuales, como las animaciones son especialmente adecuadas para mostrar la información de manera temporal. Los beneficios cognitivos del uso de simulaciones gráficas radican en la eliminación de representaciones diferentes; la simulación del proceso de representación dinámica visual y su descripción específica textual.

3.4.2 Lenguajes visuales de programación

La programación visual engloba la idea de utilizar un lenguaje de alto nivel que permita mejorar y potenciar las capacidades como programador. La base de los lenguajes de programación visual está en el uso de semántica basada en objetos visuales multidimensionales [52].

El concepto de programación visual, y sus primeras investigaciones surgen a finales de los años 60, con la aparición de las interfaces gráficas, y el creciente ascenso de los computadores. Hasta la actualidad se han desarrollado muchos lenguajes, herramientas de programación, etc. que se pueden englobar dentro de la categoría lenguajes gráficos, muchos de ellos de propósito general, aunque también destacan algunos de propósito específico [53].

Los objetivos de la programación visual han ido evolucionando, y redefiniéndose con el tiempo. En sus orígenes, se trabajaba con la idea de desarrollar un lenguaje visual que permitiera potenciar las

capacidades de programación, y desarrollar programas de propósito general, pero pronto se llegó a la conclusión de la gran dificultad que suponía, aunque en dominios específicos, y ámbitos restringidos se demostró que podían ser muy exitosos. A raíz de esto surgieron en el mercado numerosos paquetes de programación visual en campos como los sistemas de control, sistemas de minería de datos, etc. Con el surgimiento de estas herramientas de programación visual, también aparecieron “entornos de programación visual”, que pueden considerarse con algunas características de lenguaje de programación, pero que están concebidos para dar un soporte visual a lenguajes textuales tradicionales.

3.4.3 Lenguajes visuales aplicados a la comunicación

En la vida cotidiana, existen infinidad de representaciones gráficas que tienen un significado exacto asociado, y que en muchos casos se utiliza para la comunicación. Diferentes estudios han permitido llegar a la conclusión de que los procesos mentales son guiados más rápidamente por estructuras gráficas, por lo que el uso de este tipo de elementos agiliza y simplifica la comunicación [54].

Lo anterior, ha permitido incorporar iconos⁷ al mundo de la computación, en donde han tenido una gran aceptación, y su uso en comunicaciones a través de equipos electrónicos se incrementa rápidamente, como por ejemplo señalar el uso de emoticonos, en sistemas de mensajería instantánea, como el Microsoft Messenger⁸, etc. [55]. Algunas de las ventajas de usar lenguajes visuales en las comunicaciones son:

- Las imágenes son más didácticas que las palabras como un medio de comunicación, pudiendo transmitir mucha más información por unidad de expresión.

⁷ Icono, según el diccionario Webster (Merriam-Webster, 2003), es una imagen, representación, ilustración, grabado, o esquema utilizado para representar un concepto, idea, dato u operación.

⁸ <http://download.live.com/>

- Las imágenes ayudan a entender y recordar, y son más atractivas visualmente.
- Las imágenes no están sujetas a las barreras del lenguaje, cuando se crean correctamente, son independientes al idioma que se hable.

Después de haber realizado este repaso por los conceptos más importantes de los lenguajes visuales de programación, y de los lenguajes visuales aplicados a la comunicación, se ha considerado importante incluir otra sección en la que se habla de lenguajes visuales de dominio específico, de una manera genérica que podría englobar cualquier tipo de aplicación de los mismos.

3.4.4 Lenguajes Visuales de Dominio Específico

Los Lenguajes de Dominio Específico (LDE) son lenguajes dedicados a resolver un problema en un dominio en particular, o a una técnica de representación o resolución de problemas específica [56].

Los Lenguajes de Dominio Específico ofrecen abstracciones de alto nivel para la especificación de la estructura y comportamiento en un dominio concreto. Suelen ser lenguajes pequeños que definen un conjunto de primitivas para representar conceptos del dominio, lo que permite a los expertos en el dominio por tanto realizar un aprendizaje más rápido del lenguaje, así como de la especificación, creación y mantenimiento de sistemas definidos con ellos [57].

Cuando las primitivas que definen el LDE son gráficas, se habla de Lenguajes Visuales de Dominio Específico (LVDE). Los LVDE poseen las ventajas de los LDE, y las del uso de una notación visual, que permite obtener resultados más intuitivos [58].

Cuando se diseña un LVDE habitualmente se diferencia entre la sintaxis abstracta, y la concreta. La sintaxis abstracta incluye los conceptos del lenguaje, y sus relaciones, mientras que la sintaxis

concreta define la apariencia gráfica de los elementos de la sintaxis abstracta.

Existen principalmente dos tipos diferentes de enfoques para definir la sintaxis abstracta de un LVDE, los enfoques declarativos y los operacionales. Los enfoques declarativos son más sencillos.

A la hora de definir la sintaxis concreta de un LVDE existen también varios enfoques. En el caso más sencillo, si la estructura concreta y abstracta son similares, es suficiente con asignar una representación gráfica a cada elemento de la sintaxis concreta, sin embargo en ocasiones es posible encontrar relaciones arbitrarias entre la sintaxis abstracta y la concreta. Es habitual encontrarse con conceptos de la sintaxis abstracta que no tienen una representación gráfica, o que por el contrario puedan representarse de distinta forma dependiendo de ciertos factores. También es posible encontrar casos donde ciertos elementos de la sintaxis concreta no están asociados a ningún elemento de la sintaxis abstracta.

3.5 La base de datos

El nacimiento y desarrollo de un sistema de base de datos impulsó una revolución de la gestión de información. En poco más de tres décadas, en el mundo ya han aparecido miles de bases de datos, ya sea en la infraestructura de las empresas, gobierno incluso en trabajos de carácter. Al mismo tiempo, la expansión y complejidad de las aplicaciones han originado, las áreas de investigación también se amplía. En los años 30, el área de base de datos ya obtiene tres computer turning [59], por eso la base de datos es un área con mucha energía y novedad.

3.5.1 La evolución de los sistemas de base de datos

Los sistemas de información existen desde las primeras civilizaciones. El concepto más esencial de sistema de información no ha variado desde los censos romanos, por poner un ejemplo. Los datos se recopilaban, se estructuraban, se centralizaban y se almacenaban convenientemente. El objetivo inmediato de este proceso era poder recuperar estos mismos datos u otros, derivados de ellos en cualquier momento, sin necesidad de volverlos a recopilar, ya que solía ser el más costoso o incluso irrepetible. El objetivo ulterior de un sistema de información, no obstante, era proporcionar a los usuarios información fidedigna sobre el dominio que representaban, con el objetivo de tomar decisiones y realizar acciones más pertinentes que las que se realizarían sin dicha información.

Llamamos base de datos a esta colección de datos recopilados y estructurados que existe durante un periodo de tiempo. Por ejemplo, un libro contable, debido a su estructura, se puede considerar una base de datos. Una novela, por el contrario, no tiene casi estructura, y no se suele considerar una base de datos. Generalmente, un sistema de información consta de una o más bases de datos, junto con los medios para almacenarlas y gestionarlas, sus usuarios y sus administradores.

Hoy en día, sin embargo, solemos asociar las bases de datos con los ordenadores, y su gestión no suele ser manual, sino altamente automatizada. Más concretamente, la tecnología actual insta a la delegación de la gestión de una base de datos a unos tipos de aplicaciones software específico denominadas sistemas de gestión de bases de datos (SGBD) o, simplemente, sistemas de bases de datos. Por esta razón, hablar de la tecnología de bases de datos es prácticamente lo mismo que hablar de la tecnología de los sistemas de gestión de bases de datos. Las funciones básicas de un sistema de gestión de base de datos son:

- a) Permitir a los usuarios crear nuevas bases de datos y especificar su estructura, utilizando un lenguaje o interfaz especializado, llamado lenguaje o interfaz de definición de datos.
- b) Dar a los usuarios la posibilidad de consultar los datos (es decir, recuperarlos parcial o totalmente) y modificarlos, utilizando un lenguaje o interfaz apropiado, generalmente llamado lenguaje de consulta o lenguaje de manipulación de datos.
- c) Permitir el almacenamiento de grandes cantidades de datos durante un largo periodo de tiempo, manteniéndolos seguros de accidentes o uso no autorizado y permitiendo un acceso eficiente a los datos para consultas y modificaciones.
- d) Controlar el acceso a los datos de muchos usuarios a la vez, impidiendo que las acciones de un usuario puedan afectar a las acciones de otro sobre datos diferentes y que el acceso simultáneo no corrompa los datos.

3.5.2 Primeros sistemas de base de datos

La historia de los bases de datos se remontó a cincuenta años atrás. El momento en que la gestión de datos era muy simple. Y se realizaba a través de un gran número máquinas de clasificación, comparación y formularios en forma de tarjetas perforadas para gestionar los datos.

Sin embargo, en el año 1951 un ordenador de la empresa Remington (Remington Rand inc) que se llama Univac, introducía un tipo de unidades de cinta, que importaba cientos de registros en un segundo, en consecuencia desencadenaba una revolución en la gestión de datos. En año 1956 la empresa IBM producía la primera de unidad de disco—*the Model 305 RAMAC*. Esta unidad de disco tenía cincuenta discos, el diámetro de cada disco era dos pies, se podía almacenar 5 MB de datos. La mayor ventaja de unidad de disco es que era de acceso aleatorio a los datos, y las tarjetas perforadas solo de acceso secuencial a los datos.

La revolución del sistema de base de datos aparecía en los años 60. En ese momento, se comenzó a utilizar el ordenador ampliamente en el área de gestión de datos. Las demandas de intercambio de eran datos cada vez mayores. El sistema tradicional de archivos ya no podía satisfacer las necesidades de las personas. Entonces el sistema de gestión de base de datos (*DBMS, data base management system*) podía gestionar y difundir los datos surgía. El modelo de los datos es el núcleo y fundamento del sistema de base de datos. Todo el software *DBMS* se basa en un modelo de datos. Por eso normalmente según las características de modelo de datos, se divide los sistemas tradicionales de base de datos en tres tipos: base de datos en red, base de datos jerárquica, base de datos relacional.

La primera base de datos en red se llama *IDS—Integrated DataStore*, la desarrolló *Bachman* y sus compañeros que trabajaban en *General Electric Corporation (GE)* de E.E.U.U en 1961, también es el primer sistema de base de datos, sentó las bases para base de datos distribuido. El *IDS* tiene las características de modelos de datos y registro. Este sistema solo podía funcionar en el servidor de *GE*, donde hay un documento en la base de datos, todas las tablas de base de datos hay que generarlas a través de codificación manual. Después, la empresa *BF Goodrich chemical* desarrolló otro sistema que se llamó *Integrated Data Management System (IDMS)*.

El modelo de base de datos distribuido se puede simular naturalmente la estructura jerárquica y no estructura jerárquica, antes de aparecer la base de datos relacional, la utilizaba más ampliamente que la base de datos jerárquica. En la historia del desarrollo de la base de datos, la base de datos en red ocupa un lugar importante.

La base de datos jerárquica apareció seguida de la base de datos distribuido. La más famosa y clásica era *IMS--Information Management System*, desarrollado por la empresa *IBM* en 1968, esta base de datos que adapta su servidor. Eso es el primer producto de procedimientos para gran escala de sistema de base de datos que

desarrolló de la empresa IBM. Producida a partir de finales de los años 60 en adelante, hoy ha sido desarrollado a *IMSV6*, este producto de base de datos que tiene 30 años de edad, hoy en día siguen jugando un importante papel.

3.5.3 Sistemas de base de datos relaciones

La base de datos distribuido y jerárquica y soluciona los problemas de concentración e intercambio de datos, pero independencia de datos y abstracta todavía tiene una gran caraca.

En el 1970, el investigador de IBM, Doctorado *E.F.Codd* publicó un artículo “*A Relational Model of Data for large Shared Data Banks*” en la revista <*Communication of the ACM*>, presentó el concepto de modelo relacional, sentó la base teórica de modelo relacional. Aunque en año 1968 *Childs* ya había presentado el modelo orientado a conjuntos, pero el artículo de *E.F.Codd* se considera en general como un hito que tiene un significado que hace época en la historia de sistema de base de datos. Después *E.F.Codd* también publicó muchos artículos, exponía 12 estándares de la teoría de paradigmas y sistemas para medir la relación. Este usaba la teoría matemática a sentar las bases de bases de datos relacionales. El modelo relacional lleve implícito matemáticos estrictos, con alto de abstracto nivel.

En 1974, *Ray Boyce* y *Don Chamberlin* de IBM utilizó la sencilla sintaxis de palabras clave para explicar la definición matemática de las 12 directrices que exponía *Codd* para la base de datos relacional, inventó el lenguaje *SQL (Structured Query Language)*. La función de lenguaje *SQL* incluye consultar, manejar, definir y controlar, es el lenguaje sintético y común de base de datos relacional, también es el lenguaje con alto nivel de no-proceso, los usuarios solo necesitan señalar lo que quieren, y no necesitan saber cómo hacerlo. Con *SQL* se puede realizar todas las operaciones del ciclo vida de base de datos. *SQL* ofrece una manera para interactuar con base de datos relacional,

compartible con el lenguaje ensamblador. A partir de su origen, lenguaje es la piedra de toque para verificar la base de datos relacional, y todos los cambios de estándar de lenguaje SQL dirigen la dirección de desarrollo del productor de base de datos relacional. En año 1986, ANSI (*American National Standards Institute*) utilizó SQL como estándar de E.E.U.U de base de datos relacional, y en mismo año publicó SQL estándar texto. Actualmente hay 3 versiones de SQL estándares. La definición básica es *ANSIX3135-95 (Database Language - SQL with Integrity Enhancement[ANS89])*, normalmente se llama *SQL-89*. Este estándar ha definido la definición de modelo, la operación de datos y la transacción de procesamiento. *SQL-89* y *ANSIX3168-1989(Database language-Embedded SQL)* posterior, integraban la primera generación SQL estándar. Y después, *ANSIX3135-1992[ANS92]* describía el SQL con la función reforzada. *SQL-92* incluye las características reforzadas como modo de operación, creación dinámica, efectuación dinámica de SQL, apoyo del entorno de la red, etc. Después de terminar la estándar de *SQL-92*, ANSI e ISO empezaron a desarrollar juntos la estándar *SQL3*. La característica principal de *SQL3* es el apoyo de datos abstractos, ha proporcionado estándar para el nuevo tipo de base de datos--Bases de Datos Relacionales Orientadas a Objetos.

Después de establecer el modelo relacional en año 1970, la empresa IBM empezó un programa de investigación que se llamó *System R*, que terminó en 1979, este fue el primer *DBMS* que había realizado *SQL*. Hasta 1980 no apreció como un producto comercial. Pero al mismo tiempo, en año 1973 *Michael Stonebraker* y *Eugene Wong* de *Universidad de California at Berkeley* a partir de las materiales de *System R*, empezó a desarrollar sus sistemas de base de datos relacional—*Ingres*. Este programa se comercializó por la empresa *Oracle*, empresa *Ingres*. En 1976, la empresa *Honeywell* desarrolló el primer sistema comercial de base de datos relacional--*Multics Relational Data Store*.

3.5.4 Sistema de base de datos orientados a objetos

Aunque la tecnología del sistema de base de datos relacional es bastante madura, sus limitaciones también son evidentes: él puede arreglar los que se llama “datos de tablas”, pero para los datos de tipos complejos que han aparecido cada vez más. Después de la década de los noventa, se está investigando y buscando un nuevo tipo de sistema de base de datos. Las bases de datos orientadas a objetos (BDOO) son aquellas cuyo modelo de datos está orientado a objetos y almacenan y recuperan objetos. Su origen se debe a que en los modelos clásicos de datos tienen problemas para representar cierta información, puesto que aunque permiten representar gran cantidad de datos, las operaciones que se pueden realizar con ellos son bastante simples. Las clases utilizadas en un determinado lenguaje de programación orientado a objetos son las mismas clases que serán utilizadas en una BDOO; de tal manera, que no es necesaria una transformación del modelo de objetos para ser utilizado por un SGBDOO. De forma contraria, el modelo relacional requiere abstraerse lo suficiente como para adaptar los objetos del mundo real a tablas. Las bases de datos orientadas a objetos surgen para evitar los problemas que surgen al tratar de representar cierta información, aprovechar las ventajas del paradigma orientado a objetos en el campo de las bases de datos y para evitar transformaciones entre modelos de datos (usar el mismo modelo de objetos).

En la actualidad, las arquitecturas tradicionales de un sistema de información delegan la responsabilidad de la persistencia de los datos a un motor de base de datos relacional, mientras que en la parte lógica, y en general en la ejecución del sistema, se mantienen familias de objetos para encapsular las entradas y mantener el estado de las operaciones en dichos sistemas. Esto obliga a siempre buscar alternativas para extraer la información contenida en una familia de objetos y adaptarla de tal forma que se pueda almacenar el estado de éstos en tablas de un modelo relacional. Las bases de datos O.O.

hacen transparente este proceso, permitiendo hacer la persistencia, consulta y reconstrucción de objetos de forma inmediata.

Una base de datos orientados a objetos posea por definición las siguientes características:

- Manejo de objetos complejos
- Identidad de objetos: Los objetos dentro de una base de datos orientados a objetos no manejan el concepto de llave primaria, es decir, permite dos objetos con los mismos atributos: son iguales, pero no son el mismo.
- Una base de datos orientados a objetos no afecta las características de polimorfismo y encadenamiento dinámico de los objetos.

Por estas características, la base de datos orientados a objetos tiene ventajas tal como el manejo natural de modelos complejos (información estructurada recursivamente, modelos polimorfos, etc.) y el eliminación por completo la necesidad de *ORMs--object-relational mapping tool* (impedancia), con su consecuente mejora en el desempeño. Pero por otro lado, también tiene algunos problemas, por ejemplo, mecanismos de consulta muy primitivos, sin un estándar independiente de la plataforma; Imposibilidad de utilizar procedimientos almacenados, ya que los objetos solo pueden ser consultados en el cliente; Ineficiencia para algunas operaciones optimizadas en los RDBMS, como consulta de listados.

3.5.5 Bases de Datos Relacionales Orientadas a Objetos

La base de datos del modelo objeto-relaciona es un nuevo desarrollo con gran impacto. No es una tecnología en sí, sino una aglutinación del modelo relacional y orientado a objetos. Algunas extensiones hacia el paradigma de los sistemas relacionales se pueden datar en los principios de los ochenta.

Como se ha comentado anteriormente, existía la necesidad imperiosa de la industria y de sus clientes de tratar con nuevos tipos de datos: audio, imágenes y vídeo, además de tipos definidos por el usuario con sus propias propiedades. Por otra parte, las organizaciones eran reticentes a migrar un *SGBDR* a un *SGBDO* por diversos motivos.

Además, el mantenimiento de los *SGBDR* empezaba a crear *desajustes* con el uso de los lenguajes orientados a objetos. Los programadores en estos lenguajes tenían que realizar una serie de pasos de traducción de la estructura objetual del programa y de los datos a la estructura relacional de los datos [60].

Los proveedores de Bases de datos relacionales no ignoran la aparición del objeto. Claramente, especifica que los datos complejos no tienen ningún sentido. Hace un ejemplo extremo, si desea estandarizar la forma de una imagen de mapa de bits, se presenta por una serie de píxeles, así necesita establecer una tabla bastante compleja. Obviamente, almacenarla como un objeto es mucho mejor.

Ellos presentan la idea como Base de Datos Relaciona Orientada a Objeto, en esta idea se mantiene una estructura de base de datos relacional, pero permite en las columnas de tablas relaciones contener un objeto complejo. En estos objetos se pueden integrar el procesado de los datos complejos (un tipo de procesamientos de almacenar). Mientras que SQL puede llamar a los “objetos” igual como los tipos de relación.

La teoría de la relaciones de datos, ignora completamente los datos complejos (mapas, vectores, gráficos o incluso toda la tabla) como un elemento almacenado en la estructura relacional. Sin embargo, estas funciones se implementan y se comercializa, gracias a empresa como *Informix* que incorpora procesos como *Datablades*, Por otra parte Oracle utiliza *Cartridges* para implementar una aproximación de Base de Datos Relaciona Oriendado a Objetos.

3.5.6 Situación actual de sistemas de base de datos

La investigación de sistema de gestión de base de datos, a través del desarrollo en más de 30 años, ha logrado grandes éxitos, desarrollado como una disciplina rica en contenido, convirtiendo una industria de software que tiene un valor de decenas de miles de millones de dólares. En la actualidad, los productores de sistemas de base de datos representatividades son Oracle (empresa Oracle), DB2 (empresa IBM), SQL server (empresa Microsoft) y etc.

En el actual, la tecnología principal de sistema de base de datos todavía es sistema de base de datos relacional, aunque a principios de los 90 era lanzado un gran reto de base de datos orientadas a objetos, pero el mercado al fin elegía la base de datos relacional. Tanto *Oracle 9i* (empresa Oracle), *DB2* (empresa IBM) como *SQL Server* (empresa Microsoft) son bases de datos relacionales. Debido a las aplicaciones de internet, aparece la gran cantidad de datos en formato XML, algunos investigadores consideran que la siguiente generación de base de datos es un tipo nuevo que apoya el modelo XML. Desde el punto de vista, que la tecnología relacional todavía es el núcleo, todas las nuevas por ejemplo la gestión de contenidos multimedia, apoyo en formato XML u objeto complejo, son desarrollas en la tecnología basada de sistema relacional.

3.6 La traducción automática

La traducción automática, referencia a la utilización de la potencia de procesamiento rápido de las máquinas (ordenadores), para ayudar a trabajar en la labor de traducción de documentos. El sistema de traducción automática puede transformar las normas del lenguaje natural a los algoritmos de máquinas, las máquinas a través de utilizar estos algoritmos, traducen el idioma original de entrada al idioma de destino de salida.

Hoy día, debido a la rápida expansión de información, la comunicación internacional es cada vez más frecuente, rápidamente maneja una gran cantidad de información de lenguas extrañas, ya es una exigencia universal. Por eso, aplicar el sistema de traducción automática para ayudar a traducir rápidamente, se ha convertido en una tendencia inevitable [61].

3.6.1 La historia de traducción automática

La historia de la investigación en traducción automática puede remontarse a los años 40 o 50. En el año 1946, nació el primer ordenador *ENIAC*, después el científico americano *W.Weaver* y ingeniero británico *A.D.Booth* formularon la idea de utilizar el ordenador para traducir los lenguajes naturales automáticamente. En año 1949, *W.Weaver* propuso oficialmente la idea de la traducción automática. Durante los últimos 60 años, la traducción automática ha sufrido un largo y tortuoso camino que generalmente se divide en cuatro etapas:

3.6.1.1 El período de crear (1946-1964)

Año 1954, la empresa *IBM* cooperó con La Universidad de Georgetown, que utilizó el ordenador *IBM-701*, y realizó la prueba de traducción automática entre inglés y ruso, mostró la viabilidad de la traducción automática, así empezó la investigación de traducción automática.

Desde los años 50 hasta la primera mitad de los años 60, la investigación de traducción automática mostró una tendencia al alza. E.E.U.U, y Rusia por la meta de ser una potencia, habían ofrecido una gran cantidad de apoyo financiero a programa de traducción automática. Y los países europeos también habían dado importancia a la investigación de traducción automática. En esta década, la teoría de traducción automática ya desarrolló mucho.

3.6.1.2 El período de retroceso (1964-1975)

En el año 1966 ALAPC (*Automatic Language Processing Advisory Committee*) presentaba un informe, que decía que dentro de los 10 años, la investigación de la traducción automática avanzaría muy lenta.

3.6.1.3 El período de recuperación (1975-1989)

Después de entrar en la década de los 70, con el desarrollo de la tecnología y el intercambio de información entre los países se hace cada vez más frecuente, las barreras lingüísticas entre los países es aún más grave, la traducción humana tradicional ya no puede satisfacer las necesidades. Por eso, utiliza el ordenador para ayudar a traducir los documentos. Al mismo tiempo, según el desarrollo de los ciencias de la computación y la investigación de la lingüística, en particularmente, el aumento sustancial de la tecnología hardware y las aplicaciones de inteligencia artificial en lenguaje natural, promovieron la recuperación de investigación de traducción automática en nivel técnico, los proyectos comenzaron a desarrollarse. Una variedad de sistemas de prácticas o pruebas se introdujeron, por ejemplo sistema *Weinder*, multi-idioma sistema de traducción *EURPORTA*, *TAUM-METEO* sistema.

3.6.1.4 El período de nuevo (1990-actualidad)

Con la aplicación generalizada de Internet, el proceso de integración económica mundial se lleva a un ritmo acelerado y los intercambios internacionales son cada vez más frecuentes, la demanda de traducción automática se aumenta muchísimo, así se inicia una nueva oportunidad de desarrollo para la traducción automática, y impulsada por la demanda del mercado, el sistema comercial de traducción automática ha entrado en una fase de prácticas [62].

3.6.2 La teoría básica de traducción automática

La traducción automática tradicional se puede dividir en tres categorías: *Direct Translation*, *Interlingua approach*, y *Transfer approach*:

- A. *Direct Translation* solo cambia las palabras o locuciones del lenguaje original a componentes correspondientes del lenguaje objetivo. Este enfoque demasiado simplista para comprender el proceso de traducción.
- B. La idea de *Interlingua approach*. Esta categoría es a través de analizar el lenguaje original, según sus características de la gramática, resume una semántica que vale para todos los otros lenguajes. Este método en teoría es muy económico para diseñar la traducción automática sistema de traducción multilingüe.
- C. *Transfer approach*. Utiliza dos expresiones intermedias y tres procesos para traducir. El primero proceso es la conversión del lenguaje original a la expresión interior de lenguaje original, el segundo es la conversión de la expresión interior de lenguaje original a objetivo, y último es sobre la expresión interior de lenguaje objetivo a generar el lenguaje objetivo. Actualmente muchos sistemas de uso práctico utilizan esta estructura general [63].

3.6.3 Los problemas de traducción automática

Las mayores dificultades del procesamiento del lenguaje natural, es el lenguaje natural en sí mismo ya es bastante complejo, seguirá cambiando, siempre añade nuevas palabras y nuevos usos, y además hay numerosas excepciones. Por eso, los principales problemas que encuentra la traducción automática son:

1. Ambigüedad.
2. Poca estructura formal.

En la sintaxis y la semántica del lenguaje natural, a veces hay ambigüedad y la incertidumbre, necesita el contexto o el conocimiento compartido entre los que leen el texto para entender.

La producción de sistemas de traducción automática con calidad alta, necesita una gran cantidad de conocimientos. La adquisición y gestión de conocimientos, es un cuello de botella en la investigación de sistema de traducción. Según la experiencia anterior, la complejidad de este trabajo es alta, la mayoría de las unidades de investigación no puede encargarse con su costo.

3.6.4 Los productos típicos de traducción automática

En la actualidad, una fuerte demanda del mercado contribuido al actual recrudescimiento de la traducción automática. Muchas empresas e instituciones de investigación se unieron al equipo de la investigación y el desarrollo de productos de traducción automática, la comunidad presta mucha atención a las aplicaciones de traducción automática, por eso, hay una gran cantidad de los productos de traducción automática en el mercado. Estos productos se dividen en las siguientes categorías:

1. Los productos tradicionales de la traducción automática de texto.
2. Traducción de páginas web, siempre junta con la búsqueda de web, por ejemplo la traducción automática de Google.
3. Los productos de traducción automática auxiliar, ayuda a los traductores profesionales para mejorar la eficiencia de la traducción.
4. Los diccionarios bilingües electrónicos, es principalmente para ayudar a los usuarios que conocen el idioma extranjero bien para entender las informaciones de lenguaje extranjero en el ordenador.
5. La traducción automática.

3.7 Conclusiones

Los lenguajes formales permiten un tratamiento informático cómodo y generalizado, ya que dada una secuencia perteneciente a un lenguaje, y un conjunto de propiedades que deben de satisfacer, solo será necesario aplicarlas para verificar su gramaticalidad. Para llevar a cabo el objetivo de formalizar los lenguajes, habrá que definir una gramática, esta gramática será un conjunto de reglas que definirán si una secuencia arbitraria de símbolos es correcta, y por tanto pertenecerá al lenguaje.

Los lenguajes visuales son interesantes para los seres humanos porque les permiten representar, comprender, modificar y hacer inferencias de manera directa gracias a las representaciones visuales. Si el elemento humano no fuera importante en las comunicaciones a través del ordenador, los lenguajes visuales no tendrían ningún sentido, ya que los ordenadores trabajan mucho mejor con lenguajes puramente formales. Por este motivo los lenguajes visuales no solo deben de estudiarse desde la perspectiva de las especificaciones formales, sino que también deben de atender a las especificaciones de la interacción hombre-máquina. Por lo tanto hay que tener en cuenta el contexto formado por el hombre y la máquina, y el papel de los lenguajes visuales para facilitar su comunicación, y la interacción entre agentes computacionales y cognitivos, tanto en el desarrollo de aplicaciones, como en la investigación de sus propiedades formales.

Las teorías de los lenguajes visuales han existido durante bastante tiempo, el pensamiento visual [64], la semiótica⁹ de los lenguajes visuales [65], etc. Sin embargo, es necesario establecer nuevas teorías para los lenguajes visuales en el contexto de la interacción hombre-máquina.

Las tecnologías de la base de datos en principio investigan como guarda, utilizar y gestionar los datos, es uno de las tecnologías del

⁹ La semiótica es la teoría general de los signos.

ordenador que ha desarrollado y aplicado muy amplio. En el día de hoy, la tecnología de información ha desarrollado muy rápido, las aplicaciones de la base de datos ya en todas las áreas. Actualmente, la base de datos relacional todavía es la mayoría opción en el mercado, y en futuro, la tecnología de la Bases de Datos Relacionales Orientadas a Objetos será el siguiente genera.

Desde año 1946 hasta ahora, aunque las teoría de la traducción automática ya ha desarrollado mucho, también hay muchos productos se aplican en varias áreas. Pero, la traducción automática todavía necesita afrontar a muchos problemas.

CAPÍTULO 4

Diseño y desarrollo de la base de datos

4.1 La introducción

La meta de esta tesis doctoral es construir una base de datos que sirva para el proyecto “VILA-1”. En el capítulo anterior ya se explicó que este proyecto tenía como función la traducción entre lenguajes diferentes. Es decir, construimos el lenguaje nuevo y se aplica para comunicar entre los usuarios que tienen diferentes idiomas nativos. También podemos entender lo de otra forma, el “VILA-1” es un puente, conecta los usuarios, ayuda a entrar en el mundo de otros. Y la base de datos de vocabulario, sin duda, es el pilar del puente. Porque todas las palabras se almacenan en la base de datos, y todas las aplicaciones del proyecto, tanto correo electrónico como mensajería instantánea, o en el futuro, por ejemplo la aplicación de generar el resumen automática, tienen que coger las palabras en la base de datos. Si hacemos una comparación más viva, el sistema maestro es la comandancia del militar, la base de datos es su arsenal. Si falta municiones, ¿cómo ganamos en la guerra?

Así, para construir esta base de datos, hay dos procesos básicos (figura 4. 1):

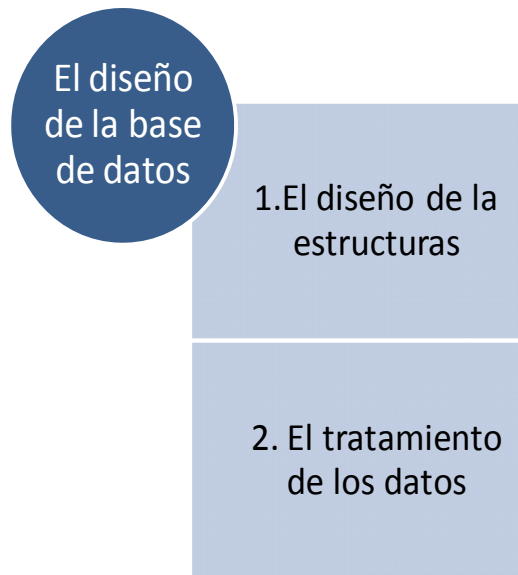


Figura 4.1 Dos procesos de construir la base de datos

Sobre la figura 4.1 se puede ver que todo el trabajo se divide en dos partes, el diseño de la base de datos y el tratamiento de los datos. Continuamos la comparación anterior, el diseño de la base de datos es como el edificio del arsenal, tiene que ser sólida y seguro, los datos son las municiones, tienen que ser útiles y coherentes. Por esta razón, para construir la base de datos bien, hay varios problemas que resolver

- ¿Qué sistema de gestión de base de datos seleccionamos?
- ¿Qué tipo de base de datos construimos?
- ¿Cómo se relacionan las palabras de los diferentes idiomas?

4.2. El diseño de la estructura

En esta sección se explica los elementos básicos de la base de datos. Incluye la introducción del sistema de gestión de base de de datos, y los elementos de la base de datos.

4.2.1 El sistema de gestión de base de datos: MYSQL

El objetivo principio de esta tesis doctoral es la construcción de una base de datos que soporta las aplicaciones de lenguaje “VILA_1”, especialmente es la aplicación de traducción entre diferentes idiomas. Por esta idea, de momento necesita una base de datos de texto. Sintetiza todos los aspectos, elige el sistema de gestión de base de datos: MYSQL.

La historia de MYSQL se puede remontar a año 1979, un hombre se llama *Monty Widenius*, trabajó en la empresa TcX, y utilizó lenguaje BASIC a escribir un programa, esto es la origina de MYSQL. En año 1996, la versión MYSQL 1.0 publicó, fui una versión internas, y en octubre en el mismo año, la versión MYSQL 3.11.1, esta es la versión simbólica, proporciona apoyo básico de SQL. Hasta ahora, ya desarrolla a la versión 5.0, MYSQL es el más popular de código abierto del sistema de gestión de bases de datos SQL, es un base de datos servido rápido, multi-hilos, multi-usuario y robusto de tipo SQL. Según estas ventajas, el sistema de base de datos MYSQL es muy corresponder para construir la base de datos que sirve para lenguaje “VILA_1”.

4.2.2 Los idiomas

Aunque el objetivo de esta tesis se centra los idiomas en relacionar adecuadamente el vocabulario chino con el español, conviene enmarcar esta relación en un ámbito más amplio, ya que VILA_1 está pensado para poder relacionar entre sí cualquier idioma.

4.2.2.1 El uso de los idiomas

En el mundo, los tres idiomas que más uso tienen son el chino, el inglés y el español.

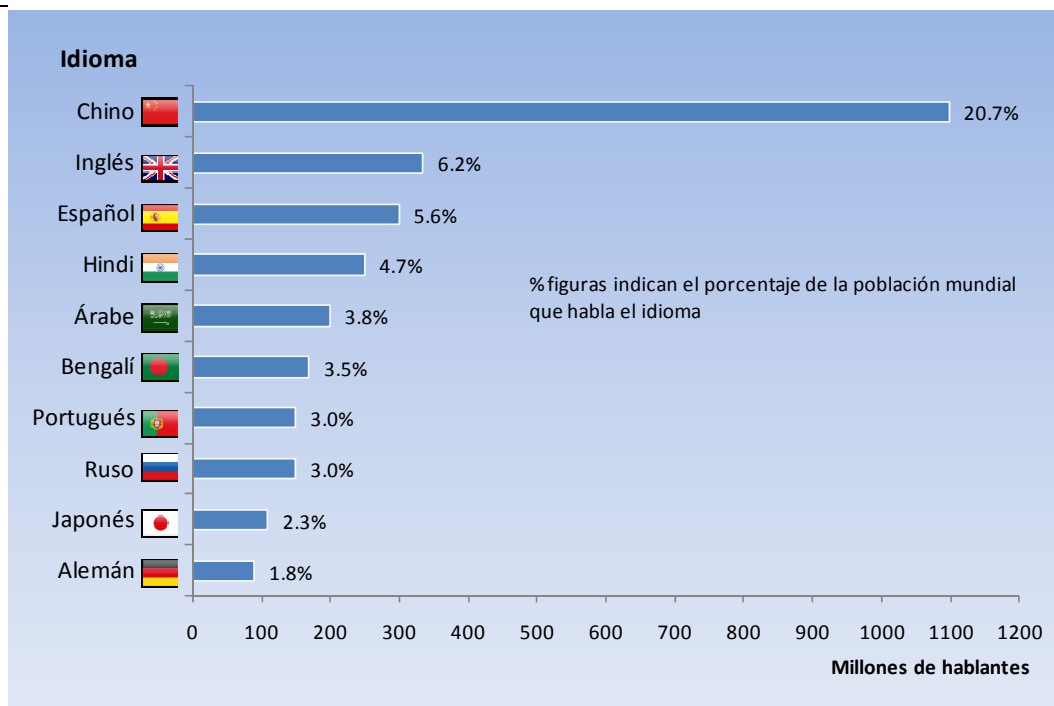


Figura 4.2 Los más usos de los idiomas en el mundo

Según la “*The World Almanac and Book of Facts*” de Estados Unidos, el chino mandarín es el idioma primero utilizado en todo el mundo, aproximadamente son 874 millones personas lo habla. Y el hindi es usado por 366 millones personas siendo segundo, y el tercero es el inglés con 340 millones y español es el cuarto con 320 millones.

Teniendo en cuenta que VILA_1 es un lenguaje orientado para su uso a través de Internet, conviene analizar qué idiomas se usan más en Internet [66].

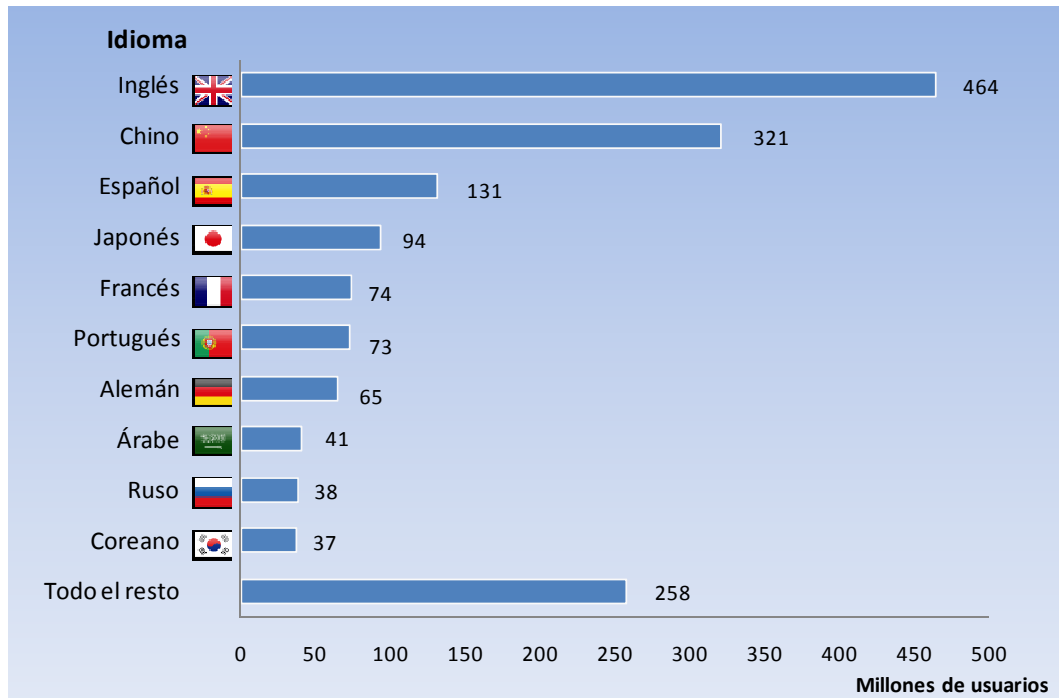


Figura 4.3 Los más usos de los idiomas en Internet

Según el diagrama se puede ver, en todo el mundo, los idiomas más utilizados en Internet son inglés con 464 millones de usuarios, chino con 321 millones de usuarios y español con 131 millones de usuarios, estos tres ocupan 57.8% del total.

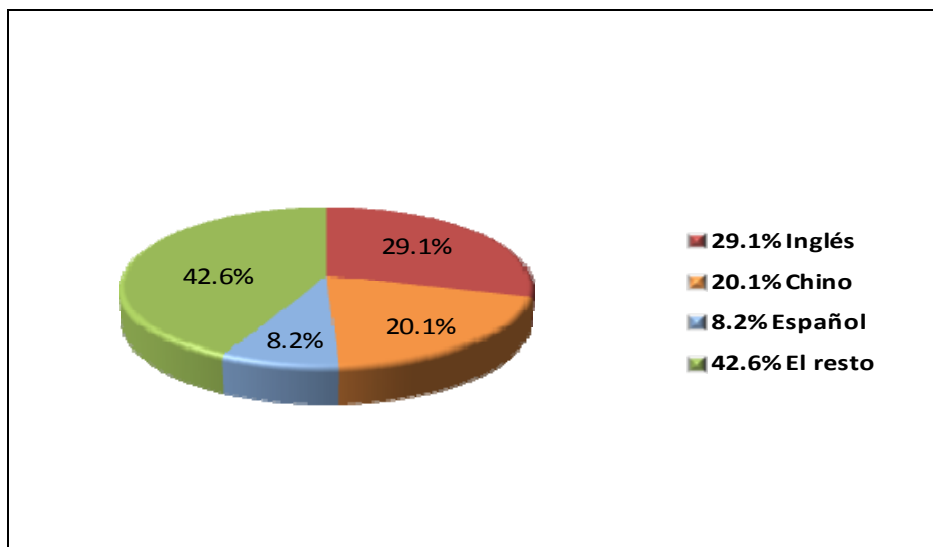


Figura 4.4 Los primero tres usos de los idiomas

- El Español es el tercero idioma en el mundo, casi hay 352, 000, 000 personas lo habla como idioma nativo, es idioma oficial de la Unión Africana, la Unión Europea y las Naciones Unidas. También es el idioma oficial de 25 países. En toda la Sudamérica menos Brasil y Haití utiliza español como idioma oficial.

El Chino es el idioma que tiene más usuarios en el mundo, también es actualmente el que crece más rápidamente. Es el idioma oficial de China, Hong-Kong, Macao, Taiwán y Singapur. En el mundo aproximadamente hay 20% persona utiliza chino como el idioma nativo, concentran principalmente en China.

Si cuenta por el número de persona que utiliza como lengua nativa, inglés es el segundo idioma en el mundo, 380, 000,000, también es el segundo idioma que más utilizado general. Hay 60% cartas y 50% publicaciones escriben en inglés, hay 73 países y regiones utilizan inglés como el idioma oficial.

De lo anterior se deduce que los idiomas más interesantes para VILA_1 son el inglés, el español y el chino.

4.2.2.2 Los problemas de gramática entre el chino y el español

VILA_1 no utiliza el concepto tradicional de traducción. Como todos saben, en los últimos años, la traducción automática tradicional aunque se ha desarrollado mucho, todavía le queda por resolver muchos problemas. Las mayores dificultades del procesamiento del lenguaje natural, son lenguaje natural en sí mismo es bastante complejo, seguirá cambiando, siempre añade nuevas palabras y nuevos usos, además hay numerosas excepciones. Por eso, los principales problemas que encuentra la traducción automática son:

1. Ambigüedad.
2. Dependencia del contexto.
3. Multiplicidad de significados.

En la sintaxis y la semántica de lenguaje natural, a veces hay la ambigüedad y la incertidumbre, necesita entenderlo por el contexto o el conocimiento compartido entre los que leen el texto. Vamos a ver unos ejemplos entre chino y español para comprender mejor las dificultades.

4.2.2.2.1 Los diferentes en los verbos

En chino, los verbos siempre tienen forma original, nunca conjugan ni tiempo ni número, por ejemplo:

我去学校。
你去学校。
他去学校。
我们去学校。
你们去学校。
他们去学校。

Se puede ver, no importa qué es la primera persona “**我**”, o segunda persona “**你**” o tercera persona “**他**” o plural, el verbo “**去**” nunca conjuga, siempre con su forma original. Vamos a comparar las frases españolas que tienen el mismo significado:

Yo voy a la escuela.
Tú vas a la escuela.
Él/ella va a la escuela.
Nosotros vamos a la escuela.
Vosotros vais a la escuela.
Ellos/ellas van a la escuela.

En la frase se puede ver, según el cambio de persona yo, tú, él, nosotros, vosotros, ellos, el verbo “**ir**” cambia a forma correspondiente como voy, vas, va, vamos, vais, van. Si usamos la teoría de la traducción automática se necesitan establecer un conjunto de reglas que

permitan establece la relación adecuada en cada caso. Sin duda, este trabajo es complejo.

Además, sobre este problema encontramos otro problema. Es que en chino, los verbos nunca conjugan, así el sujeto es obligatorio, él decide la persona. Usamos el mismo ejemplo:

**我去学校。
你去学校。
他去学校。
我们去学校。
你们去学校。
他们去学校。**

Si eliminamos los sujetos “我”, “你”, “他”, “我们”, “你们”, “他们”, las frase tienen estas formas:

**去学校。
去学校。
去学校。
去学校。
去学校。
去学校。**

Repentinamente podemos notar, todas las frases cambian a la misma, incapaz de describir los sentidos. Pero en español, cómo los verbos conjugan correspondiente con los sujetos, aun eliminamos sujetos, las frases todavía pueden expresar los sentidos:

***Voy a la escuela.
Vas a la escuela.
Va a la escuela.
Vamos a la escuela.
Vais a la escuela.
Van a la escuela.***

Sino también esta es la costumbre de español. Por si usamos el método rutinario de palabra a palabra, cada sujeto de chino traduce a sujeto español, y el texto generado será muy largo y aburrido; Y cuando traduce español a chino, necesita según unas reglas a añadir los sujetos correspondientes en chino, este trabajo es complicado y fácil de producir equivocación.

Además, esto solo es un parte del trabajo, los verbos en español, no sólo conjugan según la persona, también conjugan según el tiempo:

今天我去学校。
昨天我去学校。
明天我去学校。

En el ejemplo podemos ver, en chino, el verbo “去” tampoco conjuga nada, sólo ponemos los palabras de tiempo como “今天”, “明天”, “昨天” para expresar el tiempo concreto. Y en español:

Hoy voy a la escuela.
Ayer fue a la escuela.
Mañana yo iré a la escuela.

Según diferentes tiempos, el verbo ir conjuga a “voy”, “fue”, “iré”. Igual que en el caso de la persona, las palabras de tiempo no hacen falta:

Voy a la escuela.
Fue a la escuela.
Iré a la escuela.

Las frases son totalmente correctas y no tienen polisemias, también esta es la costumbre de español. Pero si eliminamos las palabras de tiempo en chino, mejor no hacemos esta prueba, seguro que trae muchísimos problema.

Cuando se intenta traducir un texto español a chino nos encontramos con el problema de que los sujetos con frecuencia se omiten.

- Al revés, en un texto chino los sujetos no se pueden omitir y las palabras de tiempo están en todos los sitios, después de traducir al español, el resultado será difícil a como se esperan los que hablan español.

4.2.2.2.2 Género y número

En chino, los nombres no tienen diferencia entre géneros. Un profesor o una profesora, todo son “**教师**”. Lo mismo el adverbio o adjetivos tampoco cambian según los cambios de nombre correspondiente. Por ejemplo:

Carmen 是 León 大学的教师，她很忙。
Isaías 是 León 大学的教师，他很忙。

Es fácil a ver, sobre lo profesora Carmen y profesor Isaías, el nombre “**教师**” es igual, y el adjetivo “**忙**” tampoco conjuga, tanto femenino como masculino. Pero en español, vamos a ver estos ejemplos que tienen los mismos significados que el anterior:

Carmen es la profesora de la Universidad de León, está ocupada.
Isaías es el profesor de la Universidad de León, está ocupado.

En estas frases se puede ver, el nombre “**profesora**” es el femenino, “**profesor**” es masculino. También según diferentes géneros de nombre, el adjetivo cambia a “**ocupado**” o “**ocupada**”. Esto es un problema grave para la traducción automática tradicional, especial cuando necesitamos traducir de chino a español, es muy difícil decidir el género. Claro sabemos Carmen es una chica y Javier es un chico, pero en china hay muchos nombres neutros, ¿y la máquina cómo lo sabe?

Y además, en español los nombres, adjetivos y adverbios tienen número singular y plural, por ejemplo:

Ramón ángel es profesor, está ocupado.

Francisco es profesor, está ocupado.

Ramón ángel y Francisco son profesores, están ocupados.

En la primera y segunda frase, nombre “*profesor*” y adjetivo “*ocupado*” son de número singular, y en esta tercera frase, “*profesores*” y “*ocupados*” son de número plural. Pero en chino, vamos a ver las mismas frases:

Ramón ángel 是老师, 他很忙。

Francisco 是老师, 他很忙。

Ramón ángel 和 Francisco 是老师, 他们很忙。

En todas las frases, los nombres y adjetivos nunca cambian. Por eso cuando traducimos una frase de chino a español, cómo decidirse es singular o plural, se una asunto compleja. Y combinando el número con el género, el trabajo será doblemente complicado.

Claro, entre chino y español hay muchas diferencias en la gramática, no puede ejemplar todos. Por eso la traducción automática tiene que afrontar muchas dificultades. Pero actualmente la demanda de comunicación es creciente. Tenemos que satisfacer las necesidades. Hay un gran río (los problemas de lenguaje natural) detiene nuestros avance, y el barco (la solución de traducción automática) todavía no está perfecto, ¿y qué hacemos, abandonamos? Que no, podemos dar un rodeo para llegar a la orilla opuesta. Este camino es el proyecto VILA, y su corazón es la base de datos.

4.2.3 Los procedimientos de desarrollo

El objetivo principal de la base de datos, es para soportar la función de traducción automática de palabras del lenguaje VILA_1. Es decir, hay que establecer una base de datos que incluye un gran número de vocabulario de varios idiomas. Para realiza este objetivo, el diseño de la base de datos se divide en los siguientes pasos:

1. Determinar las entidades.

2. Determinar los datos necesarios.
3. Normalizar los datos.
4. Importar los datos.
5. Mantener la base de datos.

4.2.3.1 Determinar las entidades

Es que el lenguaje VILA_1 es un lenguaje semi-formal, su semántica es la misma que la de los lenguajes naturales. Y en otro lado, el lenguaje VILA_1 también es un lenguaje visual, es más intuitivo y fácil de entender. VILA_1 no traduce todos los elementos de la frase. Por esta razón, hay que determinar las entidades que es necesario traducir sobre la gramática de los lenguajes naturales y la gramática del lenguaje VILA_1.

En todos lenguajes del mundo, hay 75% son de tipo SVO (sujeto + verbo + objeto) y tipo SOV (sujeto + objeto + verbo), y los cinco idiomas oficiales de las naciones unidas, chino, inglés, español, ruso y francés son de tipo SVO. Entonces, de momento considere los aspectos de tipo SVO. Abajo presenta una estructura de la frase de tipo SVO:

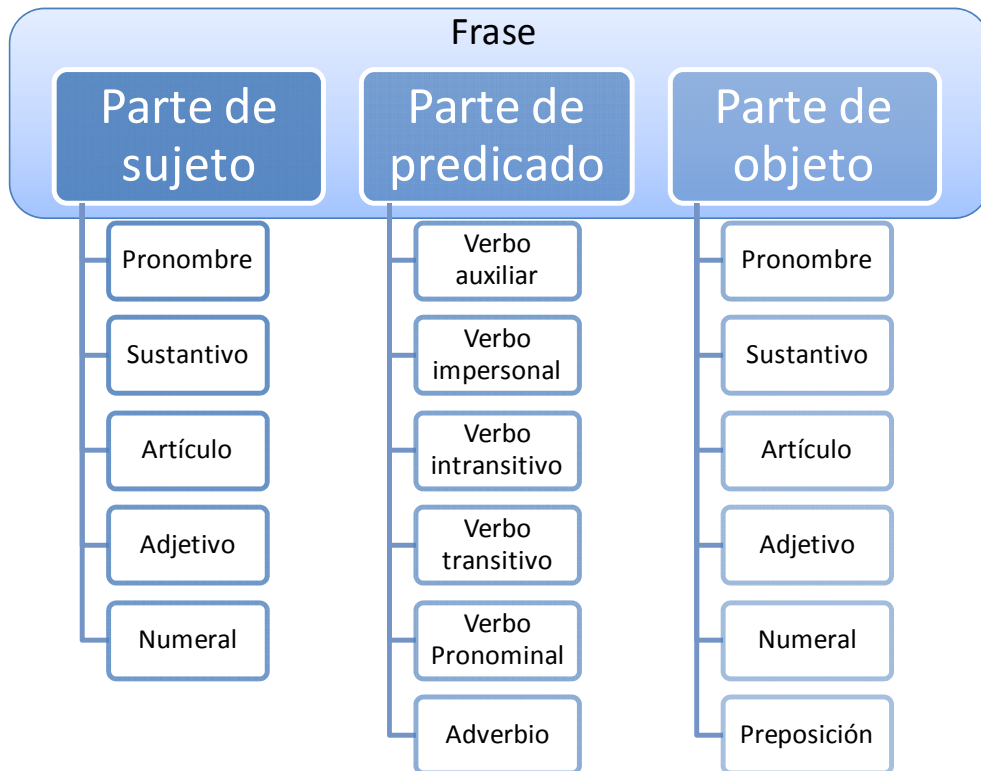


Figura 4.5 La estructural de la frase de tipo SVO

Esa es una frase básica de lenguaje natural de tipo SVO, tiene 3 partes básicas, parte de sujeto, parte de verbo y parte de objeto, y en cada parte tiene algunos elementos, pero cada parte tiene un núcleo importante, son sustantivo de parte de sujeto, verbos de parte de verbo y sustantivo de parte de objeto. Por ejemplo:

Alija come bocadillo.

Esta es una frase típica y sencilla de tipo SVO, incluye los tres partes, Alija es sujeto, come es verbo, bocadillo es objeto. Se puede explicar una situación básica. Pero si queremos describir una situación más compleja, podemos añadir más elementos a la frase:

Alija come rápidamente dos bocadillos grandes.

Esta frase ya es más compleja que la primera. Ya tiene el adverbio “*rápidamente*” en parte de predicado, numeral “*dos*” y adjetivo

“grandes” en parte de objeto. Pero tiene los elementos básicos mismos. Así, todos los elementos de la frase hay que pensarlo si necesita ser una entidad o no.

Por otra parte, el lenguaje VILA_1 es un lenguaje visual, utiliza muchos símbolos gráficos a expresar, hay pocos diferentes que le lenguaje natural, por ejemplo los artículos “el”, “la”, “los”, “las” y etc., hay símbolos especiales para representarlos, así no hace falta establecer una entidad para ellos. Y también según la gramática del lenguaje VILA_1, para fácil a expresar, hay que establecer algunas entidades especiales, por ejemplo “tiempo” y “lugar”. En resumen, toda la base de datos tienen las siguientes entidades:

1. Verbo.
2. Sustantivo.
3. Adjetivo.
4. Adverbio.
5. Numeral.
6. Tiempo
7. Lugar.

Cada entidad es una tabla en la base de datos, como las presenta en el siguiente diagrama:

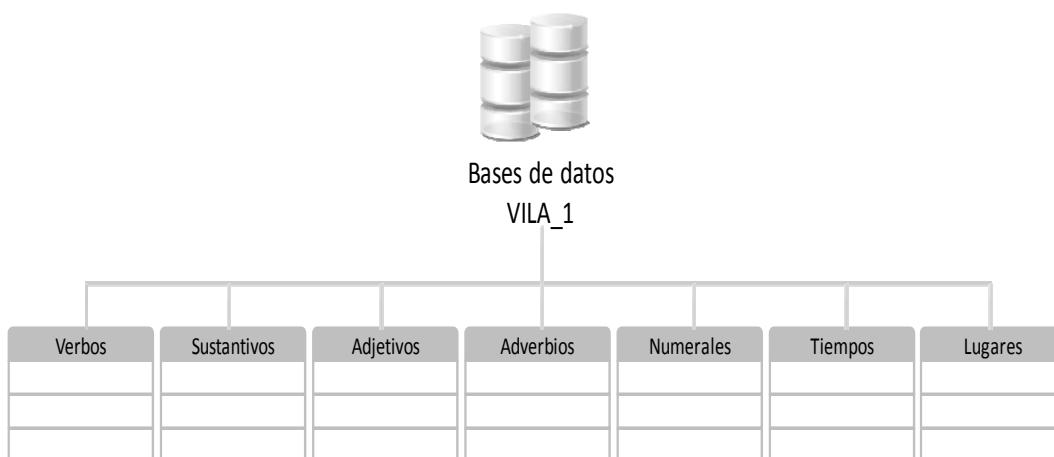


Figura 4.6 Las tablas de base de datos

Estas tablas se pueden satisfacer los requisitos de momento, y también en futuro, según el desarrollo de la teoría del lenguaje VILA_1, puede añadir cualquier tabla para satisfacer los requisitos.

4.2.3.2 Determinar los datos necesarios

El primero, necesita un campo de “Id”. Cada palabra en la base de datos tiene su único número de “ID”, para distinguir de los resto. Este campo utiliza el tipo de “int”, se incrementa automática.

La base de datos se estructura para inducir, además de español y chino, y el inglés. Pero hay una cosa debe pensarla bien. Porque chino es un idioma muy distinto que español e inglés, él no usa el alfabeto latino, usa los caracteres pictográficos, por eso es más difícil a ordenar y buscar en la data base. Entonces, aplica su signo fonético, se llama “Pinyin” o “Hanyupinyin para ayudar. El “Pinyin” o “Hanyu Pinyin”, como se escribiría correctamente es el “deletreo por sonido”, y el sistemas de transcripción oficial del chino mandarín (REFERENCIA), en lugar de utilizar los símbolos chinos, se usan letras del alfabeto latino para escribir el “sonido” de las palabras chinas. Se ha incorporado a las bases de datos porque es el sistema que permite escribir “chino” en el ordenador, y realizar ordenaciones alfabéticas. Y los cuadros campos, español, inglés, chino y Pinying, aplican el tipo de texto.

Así, en la base de datos incluye 3 idiomas, “Pinying” y ID esto 5 campos. La estructura como siguiente:

	字段	类型	整理	属性	Null	默认	额外	操作
<input type="checkbox"/>	Id	int(11)			否		auto_increment	
<input type="checkbox"/>	Espanol	text	utf8_unicode_ci		否			
<input type="checkbox"/>	English	text	utf8_unicode_ci		否			
<input type="checkbox"/>	中文	text	utf8_unicode_ci		否			
<input type="checkbox"/>	Pinyin	text	utf8_unicode_ci		否			

↑ 全选 / 全部不选 选中项:

Tabla 4.1 La estructura de la tabla

Hasta aquí, encuentra un problema muy grave: las palabras polisemias. En el lenguaje natural, algunas palabras tienen dos o más significados muy distintos. Por ejemplo:

La palabra “*gato*”, cuando decimos sobre animal, él significa el animal “*gato*”, traduce a chino es “*猫*”; pero cuando decimos sobre industria o transporte, en este caso él significa un herimiento o un instrumento, traduce a chino es “*千斤顶*”, son muy diferentes. Por eso, si no dividimos los dominio bien, el sistema seguro que tiene problema para cómo elegir el significado correcto. Es fácil a ver que si no tiene el normal serio, el ordenado siempre nos da problema. Un usuario manda un mensaje con palabra “*gato*”, igual es sistema lo traduce a “*千斤顶*”, aunque él está hablando su querida animal.

Por esta razón, tiene que dividir todo el vocabulario entre varios dominios, y tiene que dividir bien, los dominios aunque es imposible cubre todas las palabras, pero debe cubrir los voces y expresiones usuales. Y el sistema según la tema de los mensajes, coge el significado correspondiente de la palabra en el dominio correspondiente.

Después de dividir los dominios, cuando escribimos la palabra “*gato*” al sistema, en la lista hay dos para elegir: “*gato* (Vida y seres vivos)” o “*gato* (sustancias, materiales, objetos y equipamiento)”, el usuario puede elegir lo que quiere a expresar. Si quiere hablar sobre animal, elige “*gato* (Vida y seres vivos)” y cuando habla sobre coche, elige “*gato* (sustancias, materiales, objetos y equipamiento)”, entonces no equivoca nada.

Entonces, a través de investigar de los hábitos diarios de la gente y referencia a algunos diccionarios, divide en los siguientes 14 dominios:

1. Vida y seres vivos
Life and living things
生命与生物
2. El cuerpo: su función y la salud
The Body: its function and welfare
身体：机能与健康
3. Personar y familia
People and the family
人与家庭
4. Edificios, casas, el hogar, ropa, pertenencias y cuidado personal
Buildings, houses, the home, clothes, belongings and personal care
建筑，房屋，家居，衣服，财产与个人护理
5. Alimentos, bebidas, y agricultura
Food, drink and farming
食物，饮料与农业
6. Sentimientos, emociones, actitudes y sensaciones
Feelings, emotions, attitudes and sensations
感觉，感情，态度和知觉
7. Pensamiento y comunicación, lengua y gramática
Thought and communication, language and grammar
思想与交流，语言与语法
8. Sustancias, materiales, objetos y equipamiento
Substances, materials, objects and equipment

-

- 物质，材料，物件与装备
- 9. Artes y oficios, la ciencia y la tecnología, industria y educación
Arts and crafts, science and technology, industry and education
工艺美术，科学技术，工业与教育
- 10. Números, medición, finanzas y comercio
Numbers, measurement, money and commerce
数字，计量，货币和商业
- 11. Entretenimiento, deportes y juegos
Entertainment, sports, and games
娱乐，运动与游戏
- 12. Espacio y tiempo
Space and time
空间与时间
- 13. Movimiento, lugar, viajes y transporte
Movement, location, travel and transport
移动，位置，旅行与运输
- 14. General y abstracto términos
General and abstract terms
泛指与抽象

DOMINIO	领域	DOMAIN
Vida y Seres Vivos	生命与生物	Life and Living Thing
El Cuerpo: su Función y la Salud	身体, 机能与健康	The Body: its Function and Welfare
Persona y Familia	人与家庭	People and the Family
Edificios, Casas, el Hogar, Ropa, Pertenencias, y Cuidado Personal	建筑与家居	Buildings, Houses, the Home, Clothes, Belongings, and Personal Care
Alimentos, Bebidas, y Agricultura	食物, 饮料与农业	Food, Drink, and Farming
Sentimientos, Emociones, Actitudes, y Sensaciones	感觉, 感情, 态度与知觉	Feelings, Emotions, Attitudes, and Sensations
Pensamiento y Comunicación, Lengua, y Gramática	思想与交流, 语言与语法	Thought and Communication, Language and Grammar
Sustancias, Materiales, Objetos, y Equipamientos	物质, 材料, 物件与装备	Substances, Materials, Objects, and Equipment
Artes y Oficios, la Ciencia y la Tecnología, Industria y Educación	工艺美术, 科学技术, 工业与教育	Arts and Crafts, Science and Technology, Industry and Education
Números, Medición, Finanzas, y Comercio	数字, 计量, 货币与商业	Numbers, Measurement, Money, and Commerce
Entretenimiento, Deportes y Juegos	娱乐, 运动与游戏	Entertainment, Sports, and Games
Espacio y Tiempo	空间与时间	Space and Time
Movimiento, Lugar, Viajes, y Transporte	移动, 位置, 旅行与运输	Movement, Location, Travel, and Transport
General y Abstracto Términos	泛指与抽象	General and Abstract Terms

Tabla 4.2 Los dominios de sustantivos

Hay una cosa que debe explicarse bien, es que las divisiones de dominio sólo existen en la tabla de sustantivos. Porque en las frases de tipo SVO, el sustantivo incluye el accionado (sujeto) y el accionado (objeto), es el núcleo de la frase. Los restos elementos se puede entender y elegir su significado correspondiente con el sustantivo, tanto verbos como adjetivos. Por ejemplo:

Sobre el verbo “*tirar*”, se puede hace dos frases

1. *Él tira el balón.*
2. *Él tira la puerta.*

En estas dos frases, “*tirar*” significa las acciones diferentes, pero según sus objetos “*balón*” o “puerta” podemos entender perfectamente, y no nos equivoca. Y en chino, hay diferentes traducciones.

1. **他踢足球。**
2. **他拉门。**

Otro ejemplo sobre el adjetivo “rico”:

1. ***Ella nació en una familia rica.***
2. ***Ella hizo la comida rica.***

La palabra “rico” en la primera frase describe la economía, y en la segunda describe el sabor. Cambia su significado según los diferentes objetos “familia” y “comida”. Y también cuando se traduce a chino, hay diferentes palabras para corresponder:

1. **她出生在一个富裕的家庭。**
2. **她做过美味的食物。**

En los ejemplos se puede ver, solo se necesita dividir los dominios en la tabla de sustantivos. En las restantes tablas no hace falta. Entonces, en la tabla de los sustantivos, tiene ocho campos de datos, son:

1. Id
2. Español
3. Dominio
4. English
5. Domain.
6. 中文
7. 领域
8. Pinying

Los campos como dominio, domain y 领域 también son de tipos texto. Es que el campo “Pinying” sirve para ordenar y buscar, se puede ver

como un parte añadido del campo “中文”, por eso no hace falta indicar sus dominios.

←T→	Id	Espanol	Dominio	English	Domain	中文	领域	Pinyin
<input type="checkbox"/>	1	carne	alimentos, bebidas y agricultura	meat	food, drink and farming	肉	食物, 饮料与农业	rou
<input type="checkbox"/>	2	ternera	alimentos, bebidas y agricultura	calf	food, drink and farming	小牛肉	食物, 饮料与农业	xiaoniurou
<input type="checkbox"/>	3	cordero	alimentos, bebidas y agricultura	mutton	food, drink and farming	羊肉	食物, 饮料与农业	yangrou
<input type="checkbox"/>	5	solomillo	alimentos, bebidas y agricultura	steak	food, drink and farming	牛脊肉	食物, 饮料与农业	niujirou
<input type="checkbox"/>	6	bistec	alimentos, bebidas y agricultura	steak	food, drink and farming	牛排	食物, 饮料与农业	niupai
<input type="checkbox"/>	7	chuleta	alimentos, bebidas y agricultura	cutlet	food, drink and farming	连骨肉	食物, 饮料与农业	liangurou
<input type="checkbox"/>	8	estofado	alimentos, bebidas y agricultura	stew	food, drink and farming	炖肉	食物, 饮料与农业	dunrou
<input type="checkbox"/>	11	jamón	alimentos, bebidas y agricultura	ham	food, drink and farming	火腿	食物, 饮料与农业	huotui
<input type="checkbox"/>	12	tocino	alimentos, bebidas y agricultura	bacon	food, drink and farming	咸肉	食物, 饮料与农业	xianrou
<input type="checkbox"/>	13	salchicha	alimentos, bebidas y agricultura	sausage	food, drink and farming	香肠	食物, 饮料与农业	xiangchang
<input type="checkbox"/>	14	morcilla	alimentos, bebidas y agricultura	sausage	food, drink and farming	血肠	食物, 饮料与农业	xuechang
<input type="checkbox"/>	15	fiambres	alimentos, bebidas y agricultura	deli meats	food, drink and farming	冷盘	食物, 饮料与农业	lengpan
<input type="checkbox"/>	16	pollo	alimentos, bebidas y agricultura	chicken	food, drink and farming	鸡	食物, 饮料与农业	ji
<input type="checkbox"/>	17	pavo	alimentos, bebidas y agricultura	turkey	food, drink and farming	火鸡	食物, 饮料与农业	huoji
<input type="checkbox"/>	18	pato	alimentos, bebidas y agricultura	duck	food, drink and farming	鸭	食物, 饮料与农业	ya
<input type="checkbox"/>	19	pescado	alimentos, bebidas y agricultura	fish	food, drink and farming	鱼肉	食物, 饮料与农业	yurou
<input type="checkbox"/>	20	verdura	alimentos, bebidas y agricultura	vegetable	food, drink and farming	蔬菜	食物, 饮料与农业	shucai
<input type="checkbox"/>	21	patucos	edificios, casas, hogar, ropa, pertenencias y cuid...	booties	buildings, houses, home, clothes, belongings and perso...	婴儿鞋	建筑, 房屋, 家居, 服装, 财产与个人护理	yingerxie
<input type="checkbox"/>	22	patata	alimentos, bebidas y agricultura	potato	food, drink and farming	土豆	食物, 饮料与农业	tudou

Tabla 4.3 La estructura de la tabla sustantivos

Y los restos tablas tienen 5 campos de datos, son

1. Id
2. Español
3. English
4. 中文
5. Pinying

←T→			Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>			1	venir	to come	来	lai
<input type="checkbox"/>			2	correr	to run	跑	pao
<input type="checkbox"/>			3	ver	to look	看	kan
<input type="checkbox"/>			4	traer	to bring	带来	dailai
<input type="checkbox"/>			6	auxiliar	to assist	辅助	fuzhu
<input type="checkbox"/>			7	aprender	to learn	学习	xuexi
<input type="checkbox"/>			8	catar	to degustate	品尝	pinchang
<input type="checkbox"/>			9	golpear	to beat	打	da
<input type="checkbox"/>			10	comer	to eat	吃	chi
<input type="checkbox"/>			11	tomar	to take	拿	na
<input type="checkbox"/>			12	beber	to drink	喝	he

Tabla 4.4 La estructura de la tabla verbos

←T→			Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>			1	absoluto	absolute	绝对的	jueduide
<input type="checkbox"/>			2	actual	present	现在的	xianzaide
<input type="checkbox"/>			3	amplio	ample	宽敞的	kuanchangde
<input type="checkbox"/>			4	anterior	previous	在前的	zaiqiande
<input type="checkbox"/>			5	bajo	low	低的	dide
<input type="checkbox"/>			6	bastante	enough	足够的	zugoude
<input type="checkbox"/>			7	bueno	good;well	好的	haode
<input type="checkbox"/>			8	capaz	able	有能力的	younenglide
<input type="checkbox"/>			9	central	central	中心的	zhongxinde
<input type="checkbox"/>			10	cientifico	scientific	科学的	kexuede
<input type="checkbox"/>			11	civil	civil	城市的	chengshide
<input type="checkbox"/>			12	claro	clearly	明亮的	mingliangde

Tabla 4.5 La estructura de la tabla adjetivos

←T→	Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>  	1	abajo	below	在下	zaixia
<input type="checkbox"/>  	2	en absoluto	not at all	一点也不	yidianyebu
<input type="checkbox"/>  	3	acá	here	在这里	zaizheli
<input type="checkbox"/>  	4	acaso	perhaps	或许	huoxu
<input type="checkbox"/>  	5	actualmente	at present	现在	xianzai
<input type="checkbox"/>  	6	adelante	forward	向前地	xiangqiandi
<input type="checkbox"/>  	7	además	besides	此外	ciwai
<input type="checkbox"/>  	8	adentro	inside	在里面	zailimian
<input type="checkbox"/>  	9	adónde	where	什么地方	shenmedifang
<input type="checkbox"/>  	10	adrede	on purpose	故意	guyi
<input type="checkbox"/>  	11	afuera	outside	在外面	zaiwaimian
<input type="checkbox"/>  	12	ahí	there	在那里	zainali
<input type="checkbox"/>  	13	ahora	now	现在	xianzai
<input type="checkbox"/>  	14	ahorita	right now	立刻	like
<input type="checkbox"/>  	15	algo	somewhat	稍微	shaowei

Tabla 4.6 La estructura de la tabla adverbios

←T→	Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>  	1	uno	one	一	yi
<input type="checkbox"/>  	2	dos	two	二	er
<input type="checkbox"/>  	3	tres	three	三	san
<input type="checkbox"/>  	4	cuadro	four	四	si
<input type="checkbox"/>  	5	cinco	five	五	wu
<input type="checkbox"/>  	6	seis	six	六	liu
<input type="checkbox"/>  	7	siete	seven	七	qi
<input type="checkbox"/>  	8	ocho	eight	八	ba
<input type="checkbox"/>  	9	nueve	nine	九	jiu
<input type="checkbox"/>  	10	diez	ten	十	shi
<input type="checkbox"/>  	11	once	eleven	十一	shiyi
<input type="checkbox"/>  	12	doce	twelve	十二	shier
<input type="checkbox"/>  	13	trece	thirteen	十三	shisan
<input type="checkbox"/>  	14	catorce	fourteen	十四	shisi
<input type="checkbox"/>  	15	quince	fifteen	十五	shiwu

Tabla 4.7 La estructura de la tabla numeral

				Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>				1	habitación	room	房间	fangjian
<input type="checkbox"/>				2	piso	flat	套间	taojian
<input type="checkbox"/>				3	casa	house	房屋	fangwu
<input type="checkbox"/>				4	chalet	villa	别墅	bieshu
<input type="checkbox"/>				5	edificio	building	建筑	jianzhu
<input type="checkbox"/>				6	calle	street	街	jie
<input type="checkbox"/>				7	barrios	neighborhood	区	qu
<input type="checkbox"/>				8	pueblo	village	村镇	cunzhen
<input type="checkbox"/>				9	ciudad	city	城市	chengshi
<input type="checkbox"/>				10	provincia	province	省	sheng
<input type="checkbox"/>				11	capital	capital	首都	shoudu
<input type="checkbox"/>				12	país	country	国家	guojia

Tabla 4.8 La estructura de la tabla lugar

				Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>				1	milenio	millennium	千年	qiannian
<input type="checkbox"/>				2	siglo	century	世纪	shiji
<input type="checkbox"/>				3	década	decade	十年	shinian
<input type="checkbox"/>				4	año	year	年	nian
<input type="checkbox"/>				5	mes	month	月	yue
<input type="checkbox"/>				6	día	day	日	ri
<input type="checkbox"/>				7	semana	week	星期	xingqi
<input type="checkbox"/>				8	fecha	date	日期	riqi
<input type="checkbox"/>				10	horario	schedule	日程表	richengbiao
<input type="checkbox"/>				11	almanaque	calendar	日历	rili
<input type="checkbox"/>				12	lunes	Monday	星期一	xingqiyi

Tabla 4.9 La estructura de la tabla tiempo

4.2.3.3 Normalizar los datos

Es que en la base de datos hay tres idiomas, español, chino e inglés, también tiene posibilidad en el futuro de meter más idiomas. Sobre las compatibilidades de la base de datos, aplica la codificación de UTF-

8(8-bit Unicode Transformation Format). UTF-8 es un formato de codificación de caracteres Unicode e ISO 10646 utilizando símbolos de longitud variable. UTF-8 permite codificar cualquier carácter Unicode, se puede soportar tanto las letras como los jeroglíficos, por ejemplo chino. Y también se puede mostrar los símbolos especiales para matemáticas o científicos.

Todas las palabras en la base de datos, las ha metido con su forma infinitivo, no hay conjugación de los verbos ni complementos (no verbos). La forma como el diagrama anterior. Porque en la gramática del lenguaje VILA_1, hay una serie de símbolos especiales para distinguir la persona y el tiempo de los verbos, también sirve para distinguir el género y el número de los complementos (no verbos). Aplica esta forma es para solucionar los problemas de gramática entre chino y español.

En primera lugar, vamos a ver algunas expresiones en VILA_1, y sus traducciones a lenguaje natural:

Expresiones en VILA_1	Traducción a lenguaje natural
☑ niño	El niño
6 elefante	Seis elefantes
☐ árbol	Ningún árbol
☑ juguete	Muchos juguetes
☑ ilusión	Las ilusiones
> 10 mesa	Más de diez mesas
monumento	Un monumento


Figura 4.7 Expresiones en VILA_1

En la figura 4.16 se puede ver, hay muchos símbolos que ayudan a expresar el sentido de la frase. Aunque las palabras están en su forma básica, pero combina dos con los símbolos se pueden entender correctamente.

La forma de expresar el adjetivo es la siguiente:



Figura 4.8 El adjetivo en VILA_1

A continuación se muestran algunos ejemplos que utilizan adjetivos. Por ejemplo, en la segunda frase, el símbolo () se considera determinación, el símbolo (X) significa negativo, por eso, la frase en VILA_1 es fácil de entender:










Expresiones en VILA_1	Traducción a lenguaje natural
 niño  guapo	El niño es guapo
 estudiante  X ambicioso	El estudiante no es ambicioso
 cuerda   rígido	La cuerda es muy rígida
Antonio  ?  inteligente	¿Es Antonio muy inteligente?

Figura 4.9 Frase en VILA_1

Es precisamente porque a través de utilizar símbolos especiales se pueden evitar algunos problemas de las traducciones automáticas, que se mencionaron en la sección anterior. Las palabras tanto verbos como complementos (no verbos), están en la base de datos con su forma básicas.

Entonces, se formulan unas normas para los campos y los datos. Para los campos son:

1. El campo “ID” es de tipo “int”, se incrementa automática. Cada dato en la base de datos tiene su “ID” único corresponde.
2. Los restos campos, como “español”, “inglés”, “中文”, “Pinyin”, “dominio”, “domain”, “领域” son de tipo “texto”, con la codificación de UTF-8.

Para los datos que importan a la base de datos:

1. Todas palabras guardan con minúsculo.
2. Los verbos guardan con forma de infinitivo, no conjuga con el cambio de la personar ni tiempo. Los verbos pronominales Como “llamarse”, “sentarse” y etc., los guardan sin “se” en la base de datos. Por ejemplo, “llamarse” guarda como “llamar”, “sentarse” guardar como “sentar”.
3. Los sustantivos, adjetivos, adverbios, y numerales guardan en la base de datos con su forma de infinitivo, no conjugan del cambio de género ni numero. Por ejemplo la palabra

4.2.3.4 Importar los datos

En esta base de datos, los datos son las palabras de los tres idiomas que se han elegido, español, chino, e inglés. Los requisitos principales de importar los datos, son cómo garantizar los vocabularios que son útiles y modernos. Porque esta base de datos sirve para las aplicaciones de lenguaje VILA_1, son aplicaciones en línea. Por eso tiene que importar las palabras comunes en internet, y algunas palabras muy especiales y poco comunes no hacen falta. Y en otro lado, de momento las aplicaciones en principio sirven en las áreas de negocio electrónico, viajes y amistad, entonces importa más vocabularios especiales en estas áreas.

En el proceso de importar los datos, se ha elegido español como la lengua estándar, primero según las tablas de la base de datos se eligen las palabras españolas, y se luego importan las palabras correspondientes de inglés, chino y Pinying. Si la palabra tiene varios significados, hay que actuar conforme a las siguientes normas:

1. Sí la palabra tiene los significados parecidos, importar uno de ellos, el más general y común. Se puede entender la palabra sin ambigüedad.
2. Sí la palabra tiene los significados con gran diferencia, se

importan todos los significados.

E importan los tres idiomas y Pinying de una palabra en el mismo tiempo, como indica la siguiente figura:

字段	类型	函数	Null	值
Id	int(11)	<input type="text"/>		<input type="text"/>
Espanol	text	<input type="text"/>		españa
English	text	<input type="text"/>		spain
中文	text	<input type="text"/>		西班牙
Pinyin	text	<input type="text"/>		xibanya

Figura 4.10 La interfaz de importar palabras

Después de importar los datos, la base de datos esta básicamente completado, tiene la forma siguiente:










































表	操作	记录数	类型	整理	大小	多余
<input type="checkbox"/> adjetivos	     	186	MyISAM	utf8_unicode_ci	12.3 KB	-
<input type="checkbox"/> adverbios	     	209	MyISAM	utf8_unicode_ci	15.1 KB	-
<input type="checkbox"/> lugar	     	248	MyISAM	utf8_unicode_ci	17.9 KB	-
<input type="checkbox"/> numerales	     	30	MyISAM	utf8_unicode_ci	3.3 KB	-
<input type="checkbox"/> sustantivos	     	2,492	MyISAM	latin1_swedish_ci	192.1 KB	-
<input type="checkbox"/> tiempo	     	71	MyISAM	utf8_unicode_ci	5.1 KB	-
<input type="checkbox"/> verbos	     	2,518	MyISAM	latin1_swedish_ci	148.6 KB	-
7 个表	总计	5,754	MyISAM	utf8_unicode_ci	394.4 KB	0 字节

Tabla 4.10 Las tablas en la base de datos

4.2.3.5 Mantener la base de datos

En los procesos del desarrollo de la base de datos, cómo mantenerlo bien, es muy importante e indispensable. Una base de datos, aunque tiene una estructura perfecta, incluye una gran cantidad de datos, pero sin mantener y mejorar en período posterior continuamente, habrá un montón de problemas en poco tiempo.

Entonces, para mantener la base de datos bien, y mejorar la calidad constantemente, hay varios trabajos que se deben hacer:

1. Control de errores de la base de datos. Es que los datos se han importado a mano por el grupo de investigación, es inevitable a tener algunos fallos. Po eso, hay que examinar todas las tablas regular a través de probar en las aplicaciones del lenguaje VILA_1.
2. Según las reacciones de los usuarios, se importan las palabras demandadas y se borran las palabras inútiles. Así se garantiza que la base de datos funciona bien.
3. Regularmente hacer copias de seguridad.

4.3 Uso de la Base de Datos

Como ya se ha dicho la Base de Datos sirve para realizar la traducción de las palabras que usa VILA_1 entre los distintos idiomas

En el servidor de VILA_1 hay un módulo especial que sirve para traducir, se llama módulo de traducción. Este módulo se encarga de la gestión de todo el proceso de traducción de palabras. Es decir, todo el proceso de traducción se realiza en el servidor VILA_1, enviando al servidor de base de datos solamente las consultas necesarias.

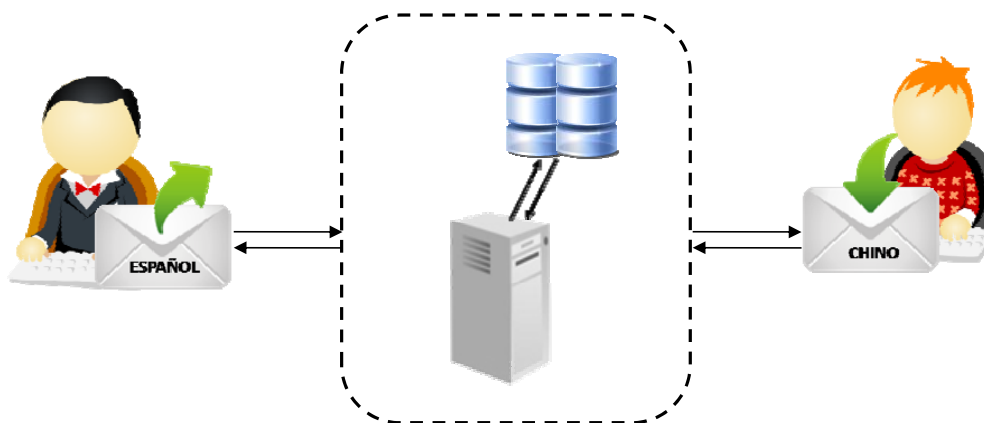


Figura 4.11 Las mensajes realizan la traducción a través por la base de datos.

Por ejemplo, el usuario español quiere mandar un mensaje que tiene la frase “Jueves Ramón va a la Universidad”, al otro usuario chino, el sistema tiene varios procesamientos.

Primero, este mensaje sobre la sintaxis del lenguaje VILA_1 es una expresión de acción, incluyen 4 elementos básicos: el tiempo “jueves”, el destinatario “Universidad”, el sujeto “Ramón”, y la acción “ir”. Los términos como “a” o “la” en el lenguaje VILA_1 son símbolos gráficos que no es necesario traducir:

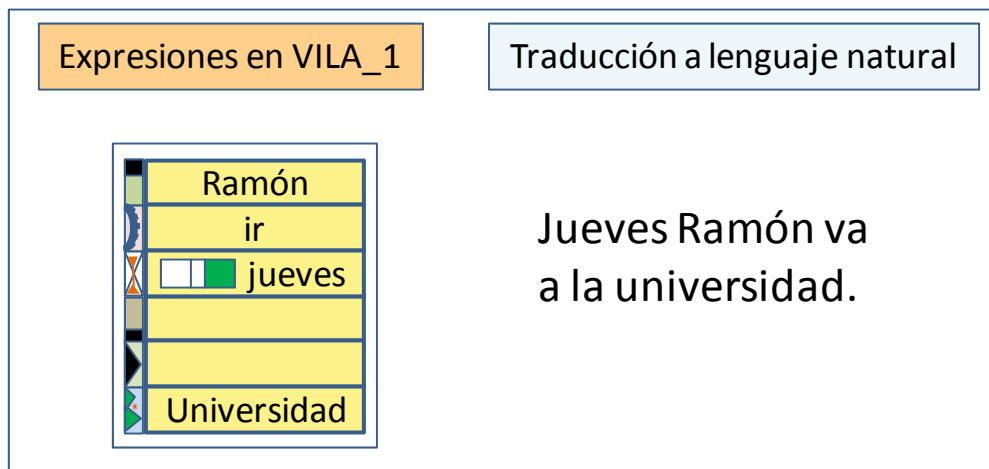


Figura 4.12 La frase española en VILA_1

Una estructura VILA_1 de tipo “Acción” se utilizan para describir acciones, sirven para describir los cambios de características de un concepto o entidad. Estos cambios se producen por la acción de fuerzas y siempre están asociados a un intervalo temporal. Este tipo de estructura está formada por los campos de tipo “Identificación”, tipo “Sujeto”, tipo “Verbo”, tipo “Tiempo”, tipo “Objeto”, tipo “Destinatario”, y tipo “Espacio”.

En la figura se puede ver, delante de la expresión de “Acción” hay un símbolo identificador, significa entre estas 5 expresiones, él es el núcleo, otras expresiones son para describirlo. Si quiere un expresión ser el núcleo, solo necesita poner “true” en el valor de Identificación.

El campo de tipo “Sujeto” contiene los siguientes elementos:

- “EspecificadoresA” (*Complex*). Almacena los especificadores como “Negación” (*booleano*), “Interrogación” (*booleano*) y “Determinación” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- “EspecificadoresB” (*Complex*). Almacena los especificadores como “Singular” (*booleano*), “Plural” (*booleano*), “MenosDe” (*booleano*), “MasDe” (*booleano*), “MasOMenos” (*booleano*), “Ninguno” (*booleano*), “Poco” (*booleano*), “Bastante”

(*booleano*), “Mucho” (*booleano*), “Todo” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).

- “Palabra” (*string*). Almacena los sujetos.

El campo de tipo “Verbo” está compuesto por los siguientes elementos:

- “EspecificadoresC” (*Complex*). Almacena los especificadores como “Negación” (*booleano*) y “Interrogación” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- “Palabra” (*string*). Almacena los verbos.

El campo de tipo “Tiempo” está formado por los siguientes elementos:

- “Pasado” (*boolean*). Almacena el tiempo pasado, si está elegido de manera aseverativa (*ture*), o de manera negativa (*false*).
- “Presente” (*boolean*). Almacena el tiempo presente, si está elegido de manera aseverativa (*ture*), o de manera negativa (*false*).
- “Futuro” (*boolean*). Almacena el tiempo futuro, si está elegido de manera aseverativa (*ture*), o de manera negativa (*false*).

El campo de tipo “Objeto” está compuesto por los siguientes elementos:

- “EspecificadoresA” (*Complex*).
- “EspecificadoresB” (*Complex*).
- “Palabra” (*string*). Almacena los objetos.

El campo de tipo “Destinatario” está compuesto por los siguientes elementos:

- “EspecificadoresA” (*Complex*).
- “EspecificadoresB” (*Complex*).
- “Palabra” (*string*). Almacena los destinatarios.

El campo de tipo “Espacio” está compuesto por los siguientes elementos:

- “EspecificadoresA” (*Complex*).
- “EspecificadoresB” (*Complex*).
- “Palabra” (*string*). Almacena los espacios.

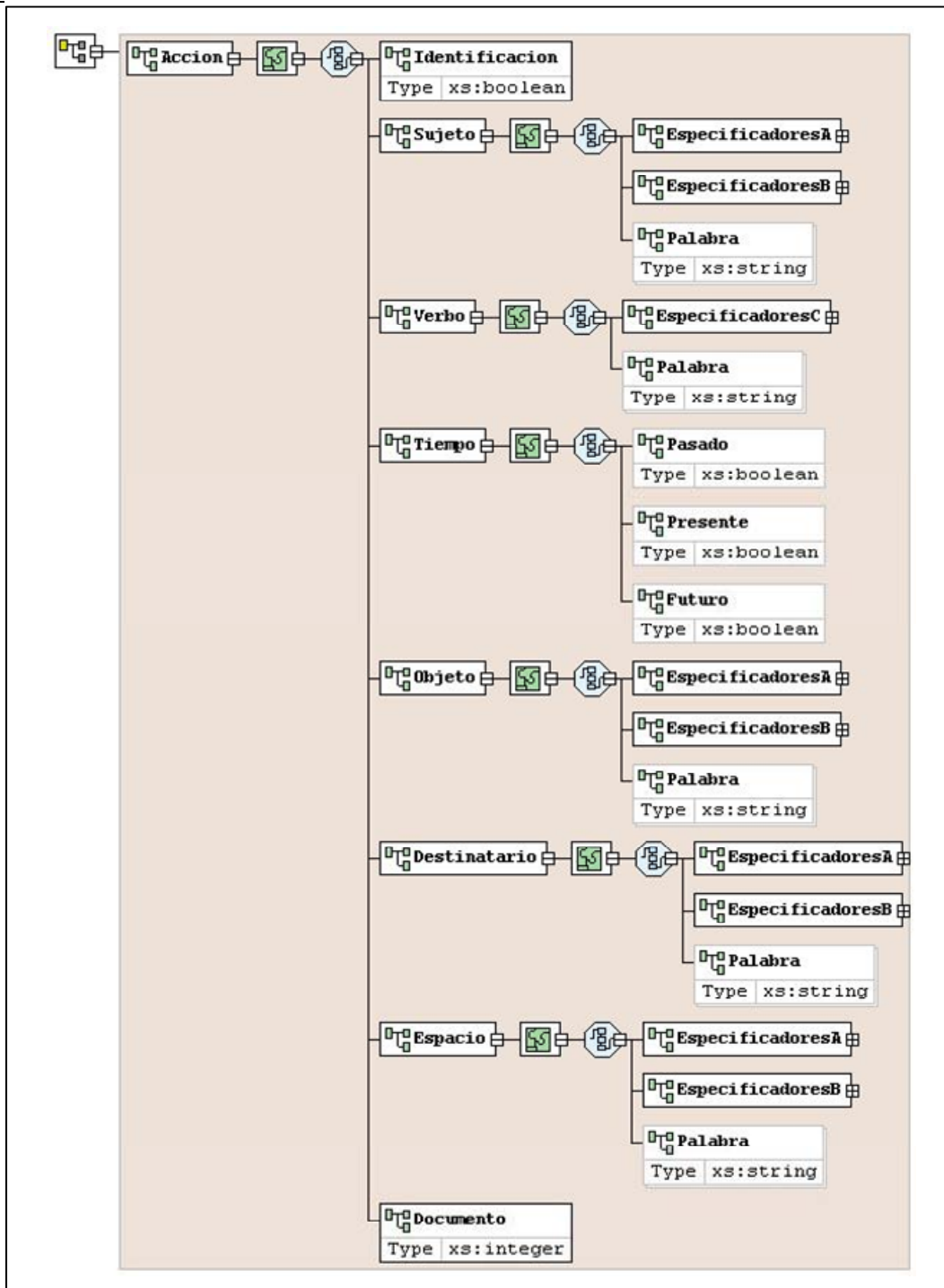


Figura 4.13 Representación gráfica del XML Schema de una estructura de tipo "Acción".

Cuando el usuario en la aplicación de mensajería instantánea de VILA_1 llena todos los conceptos correctos y pincha el botón “Enviar”, el sistema genera un fichero de tipo XML como el siguiente:

```
<?xml version="1.0"?>
<Accion xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/accion.xsd">
  <Identificacion>true</Identificacion>
  <Sujeto>
    <EspecificadoresA>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
      <Determinacion>>false</Determinacion>
    </EspecificadoresA>
    <EspecificadoresB>
      <Singular>>false</Singular>
      <Plural>>false</Plural>
      <MenosDe>>false</MenosDe>
      <MasDe>>false</MasDe>
      <MasOMenos>>false</MasOMenos>
      <Ninguno>>false</Ninguno>
      <Poco>>false</Poco>
      <Bastante>>false</Bastante>
      <Mucho>>false</Mucho>
      <Todo>>false</Todo>
    </EspecificadoresB>
    <Palabra>Ramón</Palabra>
  </Sujeto>
  <Verbo>
    <EspecificadoresC>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
    </EspecificadoresC>
    <Palabra>ir</Palabra>
  </Verbo>
  <Tiempo>
    <Pasado>>false</Pasado>
    <Presente>>false</Presente>
    <Futuro>>true</Futuro>
  </Tiempo>
</Accion>
```

```
<Palabra>jueves</Palabra>
</Tiempo>
<Objeto>
  <EspecificadoresA>
    <Negacion>>false</Negacion>
    <Interrogacion>>false</Interrogacion>
    <Determinacion>>false</Determinacion>
  </EspecificadoresA>
  <EspecificadoresB>
    <Singular>>false</Singular>
    <Plural>>false</Plural>
    <MenosDe>>false</MenosDe>
    <MasDe>>false</MasDe>
    <MasOMenos>>false</MasOMenos>
    <Ninguno>>false</Ninguno>
    <Poco>>false</Poco>
    <Bastante>>false</Bastante>
    <Mucho>>false</Mucho>
    <Todo>>false</Todo>
  </EspecificadoresB>
  <Palabra>    </Palabra>
</Objeto>
<Destinatario>
  <EspecificadoresA>
    <Negacion>>false</Negacion>
    <Interrogacion>>false</Interrogacion>
    <Determinacion>>false</Determinacion>
  </EspecificadoresA>
  <EspecificadoresB>
    <Singular>>false</Singular>
    <Plural>>false</Plural>
    <MenosDe>>false</MenosDe>
    <MasDe>>false</MasDe>
    <MasOMenos>>false</MasOMenos>
    <Ninguno>>false</Ninguno>
    <Poco>>false</Poco>
    <Bastante>>false</Bastante>
    <Mucho>>false</Mucho>
    <Todo>>false</Todo>
  </EspecificadoresB>
```

```

    <Palabra>    </Palabra>
  </Destinatario>
  <Espacio>
    <EspecificadoresA>
      <Negacion>false</Negacion>
      <Interrogacion>false</Interrogacion>
      <Determinacion>false</Determinacion>
    </EspecificadoresA>
    <EspecificadoresB>
      <Singular>false</Singular>
      <Plural>false</Plural>
      <MenosDe>false</MenosDe>
      <MasDe>false</MasDe>
      <MasOMenos>false</MasOMenos>
      <Ninguno>false</Ninguno>
      <Poco>false</Poco>
      <Bastante>false</Bastante>
      <Mucho>false</Mucho>
      <Todo>false</Todo>
    </EspecificadoresB>
    <Palabra>Universidad</Palabra>
  </Espacio>
</Accion>

```

Figura 4.14 XML del archivo “Acción” del ejemplo.

A continuación, este fichero envía al módulo de traducción, el sistema analiza los elementos, y según el idioma de destino, coge sus traducciones correspondientes en el vocabulario de la base de datos. Por ejemplo el “*jueves*” traduce a “*星期四*”, “*Universidad*” traduce a “*大学*”, “*ir*” traduce a “*去*”, y “*Ramón*” es un nombre, no hace falta traducirlo, y guarda como “*Ramón*”. Después del procesamiento de traducción, el sistema genera un fichero nuevo de tipo XML en chino:

```

<?xml version="1.0"?>
<动作 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/accion.xsd">
  <指定 >true</指定>
  <名词>
    <说明 A>

```

```
<否定>false</否定>
<疑问>false</疑问>
<特指>false</特指>
</说明 A>
<说明 B>
  <单数>false</单数>
  <复数>false</复数>
  <少于>false</少于>
  <多于>false</多于>
  <大约>false</大约>
  <没有>false</没有>
  <少>false</少>
  <足够>false</足够>
  <多>false</多>
  <全部>false</全部>
</说明 B>
<单词>Ramón</单词>
</名词>
<动词>
  <说明 C>
    <否定>false</否定>
    <疑问>false</疑问>
  </说明 C>
  <单词>去</单词>
</动词>
<时间>
  <过去>false</过去>
  <现在>false</现在>
  <将来>true</将来>
<单词>星期四</单词>
</时间>
<宾语>
  <说明 A>
    <否定>false</否定>
    <疑问>false</疑问>
    <特指>false</特指>
  </说明 A>
  <说明 B>
    <单数>false</单数>
    <复数>false</复数>
    <少于>false</少于>
```



```

    <多于>false</多于>
    <大约>false</大约>
    <没有>false</没有>
    <少>false</少>
    <足够>false</足够>
    <多>false</多>
    <全部>false</全部>
</说明 B>
<单词>    </单词>
</宾语>
<目标>
    <说明 A>
        <否定>false</否定>
        <疑问>false</疑问>
        <特指>false</特指>
    </说明 A>
    <说明 B>
        <单数>false</单数>
        <复数>false</复数>
        <少于>false</少于>
        <多于>false</多于>
        <大约>false</大约>
        <没有>false</没有>
        <少>false</少>
        <足够>false</足够>
        <多>false</多>
        <全部>false</全部>
    </说明 B>
    <单词>    </单词>
</目标>
<空间>
    <说明 A>
        <否定>false</否定>
        <疑问>false</疑问>
        <特指>false</特指>
    </说明 A>
    <说明 B>
        <单数>false</单数>
        <复数>false</复数>
        <少于>false</少于>
        <多于>false</多于>

```

```

<大约>false</大约>
<没有>false</没有>
<少>false</少>
<足够>false</足够>
<多>false</多>
<全部>false</全部>
</说明 B>
<单词>大学</单词>
</空间>
</动作>
    
```

Figura 4.15 XML del archivo "Acción" del ejemplo en chino.

Así, la frase de español ya ha traducido a chino:

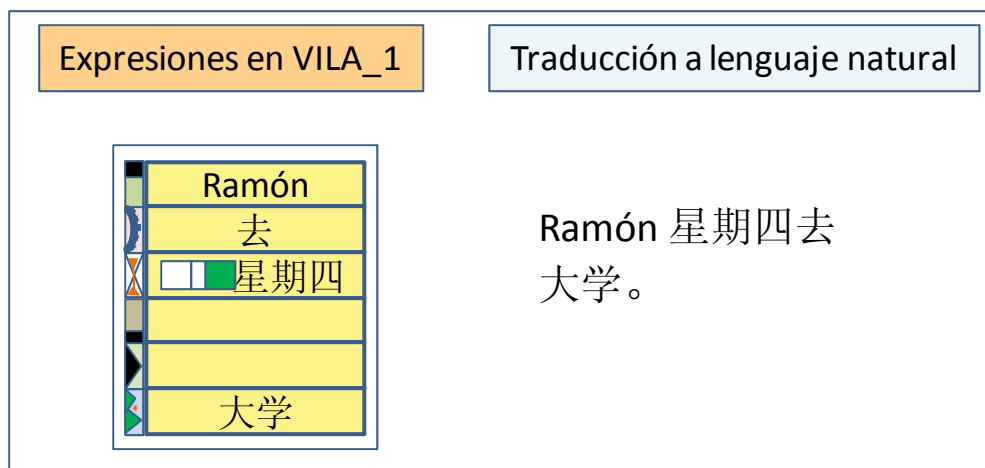


Figura 4.16 La frase china en VILA_1

Entonces, el procesamiento de traducción, en realidad es a través de tratar los archivos XML en el servidor y la base de datos. Es que sobre la características del lenguaje VILA_1, sólo necesita traducir las palabras claves, y otros elementos de la frase, por ejemplo, el conjuga del verbo o el géneros y números del objeto, hay un serie de símbolos especiales a presentarlos. Los usuarios se pueden entender las frases con una forma más audiovisual. Y también según esta característica, no debería considerar algunas gramáticas del lenguaje natural, por ejemplo la orden de los elementos de la frase, unas conjugan especiales, así es más fácil a realizar las traducciones y también ha evitado muchos problemas.

Vamos a ver otro ejemplo más complejo. Una frase que describe sobre China:

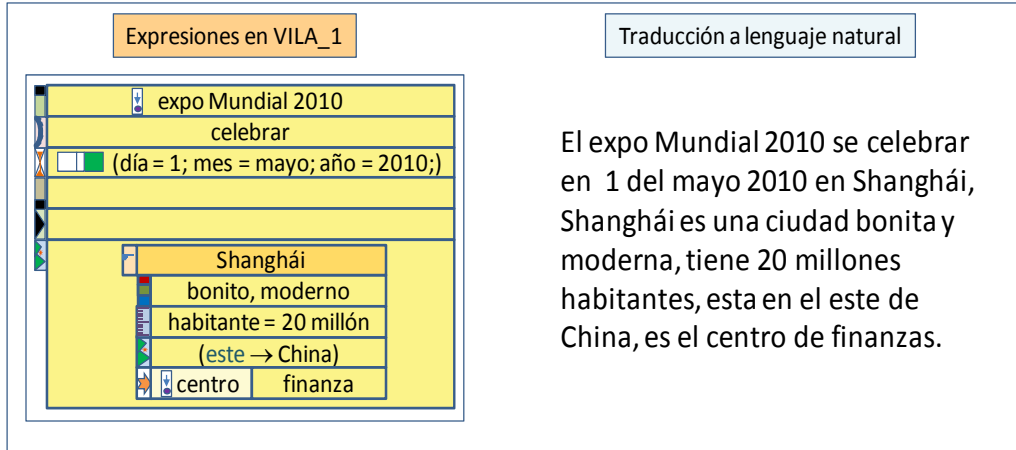


Figura 4.17 La frase española compleja en VILA_1

Esta frase está compuesta por cinco tipos de estructuras VILA_1 diferentes: “Acción”, “Adjetivo”, “Valor”, “Espacio” y “Relación”, que se detallan a continuación.

La Primera estructura VILA_1 de tipo “Acción”, ya explicada bien en el ejemplo anterior, aquí solo presenta la gráfica del XML Schema y el archivo XML:

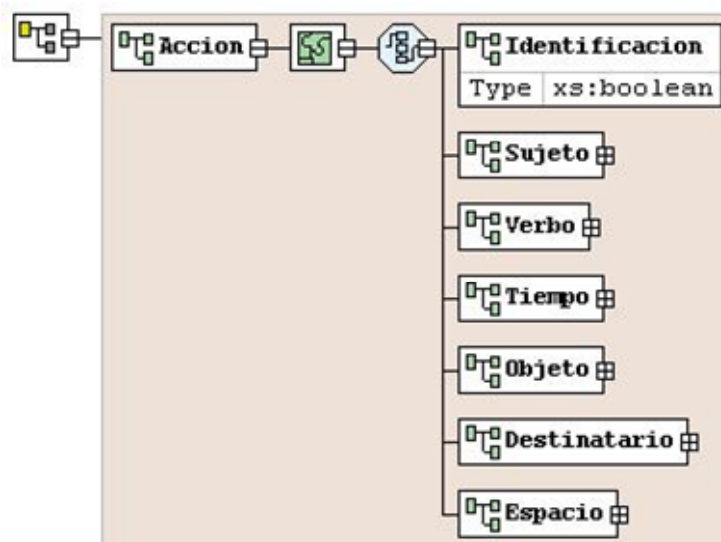


Figura 4.18 Representación gráfica del XML Schema de una estructura de tipo “Acción”

```

<?xml version="1.0"?>
<Accion xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/accion.xsd">
  <Identificacion>>false</Identificacion>
  <Sujeto>
    <EspecificadoresA>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
      <Determinacion>true</Determinacion>
    </EspecificadoresA>
    <EspecificadoresB>
      <Singular>>false</Singular>
      <Plural>>false</Plural>
      <MenosDe>>false</MenosDe>
      <MasDe>>false</MasDe>
      <MasOMenos>>false</MasOMenos>
      <Ninguno>>false</Ninguno>
      <Poco>>false</Poco>
      <Bastante>>false</Bastante>
      <Mucho>>false</Mucho>
      <Todo>>false</Todo>
    </EspecificadoresB>
    <Palabra>expo Mundial 2010</Palabra>
  </Sujeto>
  <Verbo>
    <EspecificadoresC>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
    </EspecificadoresC>
    <Palabra>celebrar</Palabra>
  </Verbo>
  <Tiempo>
    <Pasado>>false</Pasado>
    <Presente>>false</Presente>
    <Futuro>true</Futuro>
    <Dia>1</Dia>
    <Mes>mayo</Mes>
    <Año>2010</Añoa>
  </Tiempo>
  <Objeto>

```

```

<EspecificadoresA>
  <Negacion>>false</Negacion>
  <Interrogacion>>false</Interrogacion>
  <Determinacion>>false</Determinacion>
</EspecificadoresA>
<EspecificadoresB>
  <Singular>>false</Singular>
  <Plural>>false</Plural>
  <MenosDe>>false</MenosDe>
  <MasDe>>false</MasDe>
  <MasOMenos>>false</MasOMenos>
  <Ninguno>>false</Ninguno>
  <Poco>>false</Poco>
  <Bastante>>false</Bastante>
  <Mucho>>false</Mucho>
  <Todo>>false</Todo>
</EspecificadoresB>
<Palabra>    </Palabra>
</Objeto>
<Destinatario>
  <EspecificadoresA>
    <Negacion>>false</Negacion>
    <Interrogacion>>false</Interrogacion>
    <Determinacion>>false</Determinacion>
  </EspecificadoresA>
  <EspecificadoresB>
    <Singular>>false</Singular>
    <Plural>>false</Plural>
    <MenosDe>>false</MenosDe>
    <MasDe>>false</MasDe>
    <MasOMenos>>false</MasOMenos>
    <Ninguno>>false</Ninguno>
    <Poco>>false</Poco>
    <Bastante>>false</Bastante>
    <Mucho>>false</Mucho>
    <Todo>>false</Todo>
  </EspecificadoresB>
  <Palabra>    </Palabra>
</Destinatario>
<Espacio>

```

```

<EspecificadoresA>
  <Negacion>>false</Negacion>
  <Interrogacion>>false</Interrogacion>
  <Determinacion>>false</Determinacion>
</EspecificadoresA>
<EspecificadoresB>
  <Singular>>false</Singular>
  <Plural>>false</Plural>
  <MenosDe>>false</MenosDe>
  <MasDe>>false</MasDe>
  <MasOMenos>>false</MasOMenos>
  <Ninguno>>false</Ninguno>
  <Poco>>false</Poco>
  <Bastante>>false</Bastante>
  <Mucho>>false</Mucho>
  <Todo>>false</Todo>
</EspecificadoresB>
<Palabra>Shanghái</Palabra>
</Espacio>
</Accion>

```

Figura 4.19 XML del archivo “Acción” del ejemplo.

Una estructura VILA_1 de tipo “Adjetivo” sirven para darle valor cualitativo a magnitudes y otros elementos de valoración, y está formada por los campos de tipo “Identificación”, tipo “Identificador”, y tipo “Adjetivo”.

El campo de tipo “Identificador” contiene los siguientes elementos:

- “EspecificadoresA” (*Complex*). Almacena los especificadores como “Negación” (*booleano*), “Interrogación” (*booleano*) y “Determinación” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- “EspecificadoresB” (*Complex*). Almacena los especificadores como “Singular” (*booleano*), “Plural” (*booleano*), “MenosDe” (*booleano*), “MasDe” (*booleano*), “MasOMenos” (*booleano*), “Ninguno” (*booleano*), “Poco” (*booleano*), “Bastante” (*booleano*), “Mucho” (*booleano*), “Todo” (*booleano*), si un

especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).

- “Palabra” (*string*). Almacena los identificadores.

El campo de tipo “Adjetivo” está compuesto por los siguientes elementos:

- “EspecificadoresD” (*Complex*). Almacena los especificadores como “Negación” (*booleano*) y “Interrogación” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- “EspecificadoresE” (*Complex*). Almacena los especificadores como “Poco” (*booleano*), “Bastante” (*booleano*) y “Muy” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- “Palabra” (*string*). Almacena los adjetivos.

En la (Figura 4.30) se puede observar la representación gráfica del XML Schema de la estructura de “Adjetivo”, en la (Figura 4.31) el XML que se genera en el ejemplo con la misma.

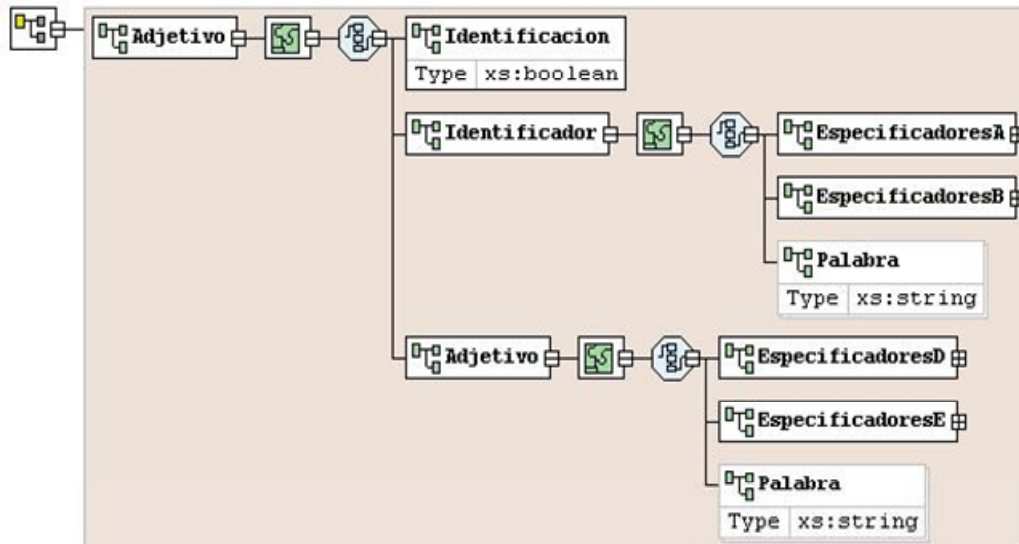


Figura 4.20 Representación gráfica del XML Schema de una estructura de tipo “Adjetivo”

```

<?xml version="1.0"?>
<Adjetivo xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/adjetivo.xsd"
  <Identificacion>true</Identificacion>
  <Identificador>
    <EspecificadoresA>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
      <Determinacion>>false</Determinacion>
    </EspecificadoresA>
    <EspecificadoresB>
      <Singular>>false</Singular>
      <Plural>>false</Plural>
      <MenosDe>>false</MenosDe>
      <MasDe>>false</MasDe>
      <MasOMenos>>false</MasOMenos>
      <Ninguno>>false</Ninguno>
      <Poco>>false</Poco>
      <Bastante>>false</Bastante>
      <Mucho>>false</Mucho>
      <Todo>>false</Todo>
    </EspecificadoresB>
    <Palabra>Shanghái</Palabra>
  </Identificador>
</Adjetivo>
  <EspecificadoresD>
    <Negacion>>false</Negacion>
    <Interrogacion>>false</Interrogacion>
  </EspecificadoresD>
  <EspecificadoresE>
    <Poco>>false</Poco>
    <Bastante>>false</Bastante>
    <Muy>>false</Muu>
  </EspecificadoresE>
  <Palabra>bonito</Palabra>
  <Palabra>moderno</Palabra>
</Adjetivo>
</Adjetivo>

```

Figura 4.21 XML del archivo “Adjetivo” del ejemplo.

Una estructura VILA_1 de tipo "Valor" puede especificar la característica y asignarle un valor, y está formada por los campos de tipo "Identificación", tipo "Identificador", y tipo "Valor".

El campo de tipo "Valor" contiene los siguientes elementos:

- "Característica" (*string*).
- "EspecificadoreC" (*Complex*). Almacena los especificadores como "Igua" (*booleano*), "MenosDe" (*booleano*), "MasDe" (*booleano*), "MasOMenos" (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- "Numero" (*integer*).
- "Palabra" (*string*).

En la (Figura 4.30) se puede observar la representación gráfica del XML Schema de la estructura de "Adjetivo", en la (Figura 4.31) el XML que se genera en el ejemplo con la misma.

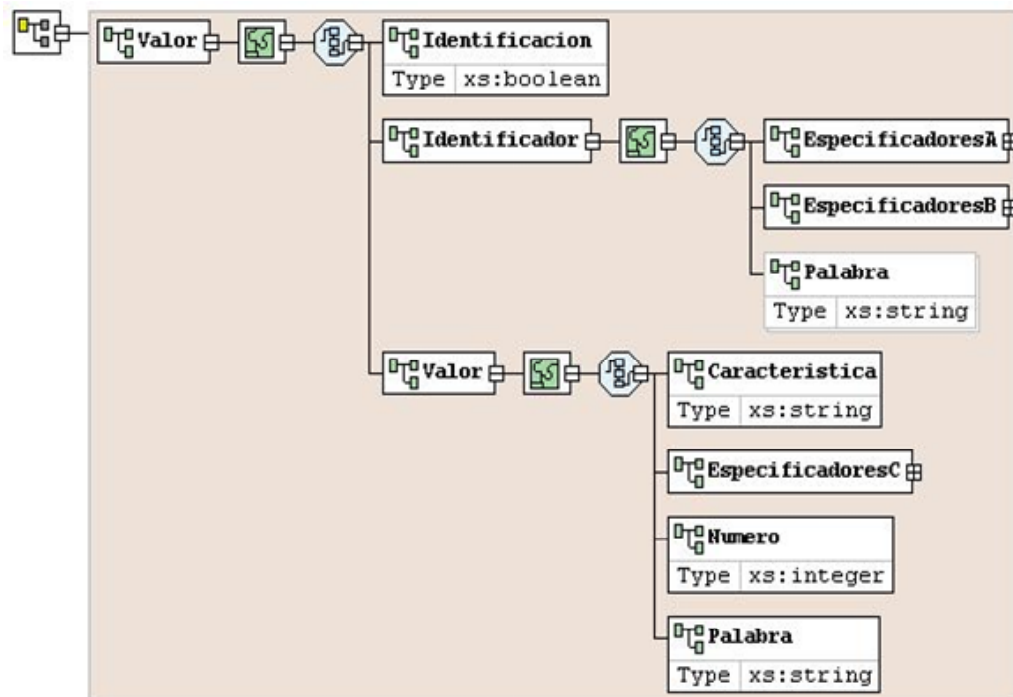


Figura 4.22 Representación gráfica del XML Schema de una estructura de tipo "Valor"

```

<?xml version="1.0"?>
<Valor xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/valor.xsd">
  <Identificacion>true</Identificacion>
  <Identificador>
    <EspecificadoresA>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
      <Determinacion>>false</Determinacion>
    </EspecificadoresA>
    <EspecificadoresB>
      <Singular>>false</Singular>
      <Plural> false </Plural>
      <MenosDe> false </MenosDe>
      <MasDe> false </MasDe>
      <MasOMenos> false </MasOMenos>
      <Ninguno> false </Ninguno>
      <Poco> false </Poco>
      <Bastante> false </Bastante>
      <Mucho> false </Mucho>
      <Todo> false </Todo>
    </EspecificadoresB>
    <Palabra>Shanghái</Palabra>
  </Identificador>
  <Valor>
    <Caracteristica>habitante</Caracteristica>
    <EspecificadoresC>
      <Igua>true</Igua>
      <MenosDe>>false</MenosDe>
      <MasDe>>false</MasDe>
      <MasOMenos>>false</MasOMenos>
    </EspecificadoresC>
    <Numero>20</Numero>
    <Palabra>millón</Palabra>
  </Valor>
</Valor>

```

Figura 4.23 XML del archivo “Valor” del ejemplo.

Una estructura VILA_1 de tipo “Espacio” puede especificarse el espacio físico o virtual en el que se encuentra ubicada la entidad o el concepto que se está describiendo o donde sucede algo, y está formada

por los campos de tipo “Identificación”, tipo “Identificador”, tipo “Espacio”, y tipo “Medida”.

El campo de tipo “Espacio” contiene los siguientes elementos:

- “TerminoRelativo” (*string*). Almacena los puntos de partida para especificar el espacio.
- “Identificador” (*Complex*). Almacena tanto los conceptos como las entidades.

El campo de tipo “Medida” está compuesto por los siguientes elementos:

- “Palabra” (*string*). Almacena los espacios figuras geométricas o términos.
- “Dato” (*string*). Almacena los datos de identificación.

En la (Figura 4.32) se puede observar la representación gráfica del XML Schema de la estructura de “Espacio”, en la (Figura 4.33) el XML que se genera en el ejemplo con la misma.

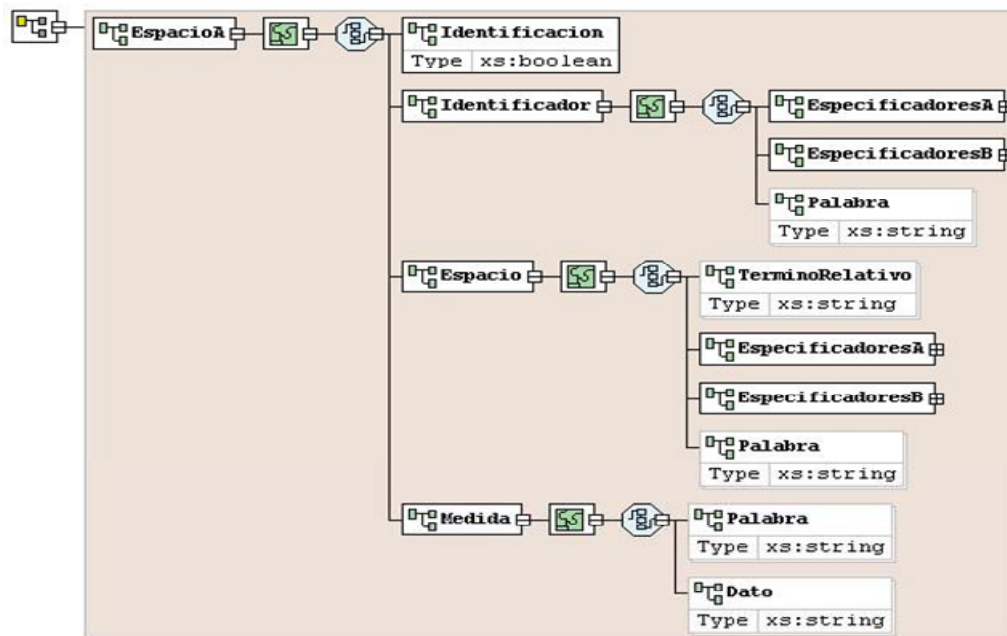


Figura 4.24 Representación gráfica del XML Schema de una estructura de tipo “Espacio”

```

<?xml version="1.0"?>
<Espacio xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/espacio.xsd">
  <Identificacion>true</Identificacion>
  <Identificador>
    <EspecificadoresA>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
      <Determinacion>>false</Determinacion>
    </EspecificadoresA>
    <EspecificadoresB>
      <Singular>>false</Singular>
      <Plural>>false</Plural>
      <MenosDe>>false</MenosDe>
      <MasDe>>false</MasDe>
      <MasOMenos>>false</MasOMenos>
      <Ninguno>>false</Ninguno>
      <Poco>>false</Poco>
      <Bastante>>false</Bastante>
      <Mucho>>false</Mucho>
      <Todo>>false</Todo>
    </EspecificadoresB>
    <Palabra>Shanghái</Palabra>
  </Identificador>
  <Espacio>
    <TerminoRelativo>este</TerminoRelativo>
  <Identificador>
    <EspecificadoresA>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
      <Determinacion>>false</Determinacion>
    </EspecificadoresA>
    <EspecificadoresB>
      <Singular>>false</Singular>
      <Plural>>false</Plural>
      <MenosDe>>false</MenosDe>
      <MasDe>>false</MasDe>
      <MasOMenos>>false</MasOMenos>
      <Ninguno>>false</Ninguno>
      <Poco>>false</Poco>
      <Bastante>>false</Bastante>
      <Mucho>>false</Mucho>
      <Todo>>false</Todo>
  </Identificador>
  </Espacio>
</Espacio>

```

```

    </EspecificadoresB>
    <Palabra>China</Palabra>
  </Identificador>
  </Espacio>
  <Medida>
    <Palabra>      </Palabra>
    <Dato>        </Dato>
  </Medida>
</Espacio>

```

Figura 4.25 XML del archivo “Espacio” del ejemplo.

Una estructura VILA_1 de tipo “Relación” permite expresar explícitamente el vínculo que existe entre dos elementos, sean entidades o conceptos, y está formada por los campos de tipo “Identificación”, tipo “Elemento1”, tipo “Vínculo”, y tipo “Elemento2”.

El campo de tipo “Elemento1” contiene los siguientes elementos:

- “EspecificadoresA” (*Complex*). Almacena los especificadores como “Negación” (*booleano*), “Interrogación” (*booleano*) y “Determinación” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- “EspecificadoresB” (*Complex*). Almacena los especificadores como “Singular” (*booleano*), “Plural” (*booleano*), “MenosDe” (*booleano*), “MasDe” (*booleano*), “MasOMenos” (*booleano*), “Ninguno” (*booleano*), “Poco” (*booleano*), “Bastante” (*booleano*), “Mucho” (*booleano*), “Todo” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- “Palabra” (*string*). Almacena los identificadores.

El campo de tipo “Vínculo” está compuesto por los siguientes elementos:

- “EspecificadoresA” (*Complex*). Almacena los especificadores como “Negación” (*booleano*), “Interrogación” (*booleano*) y

“Determinación” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).

- “Palabra” (*string*). Almacena los identificadores.

El campo de tipo “Elemento2” contiene los siguientes elementos:

- “EspecificadoresA” (*Complex*). Almacena las especificadores como “Negación” (*booleano*), “Interrogación” (*booleano*) y “Determinación” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- “EspecificadoresB” (*Complex*). Almacena las especificadores como “Singular” (*booleano*), “Plural” (*booleano*), “MenosDe” (*booleano*), “MasDe” (*booleano*), “MasOMenos” (*booleano*), “Ninguno” (*booleano*), “Poco” (*booleano*), “Bastante” (*booleano*), “Mucho” (*booleano*), “Todo” (*booleano*), si un especificador está insertado de manera aseverativa (*true*), o de manera negativa (*false*).
- “Palabra” (*string*). Almacena los identificadores.

En la (Figura 4.24) se puede observar la representación gráfica del XML Schema de la estructura de “Espacio”, en la (Figura 4.25) el XML que se genera en el ejemplo con la misma.

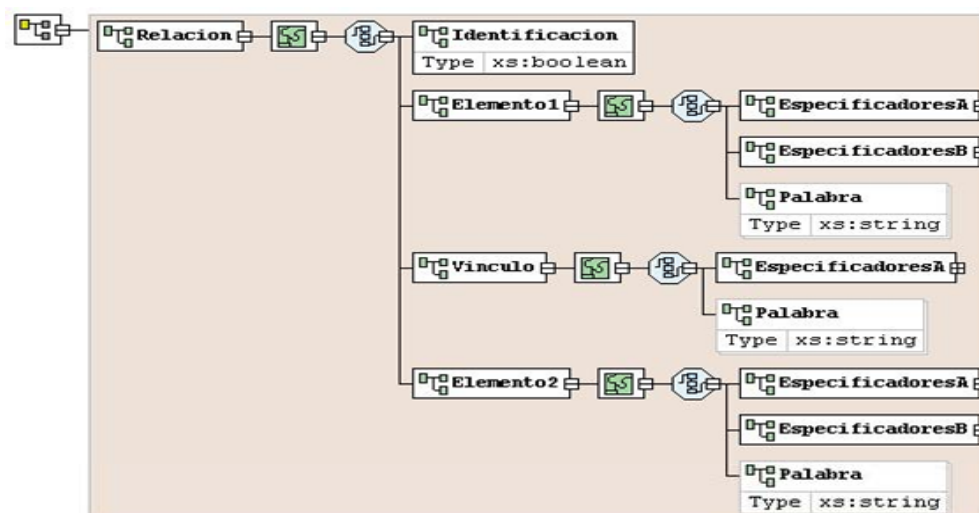


Figura 4.26 Representación gráfica del XML Schema de una estructura de tipo “Relación”

```

<?xml version="1.0"?>
<Relacion xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/relacion.xsd">
  <Identificacion>true</Identificacion>
  <Elemento1>
    <EspecificadoresA>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
      <Determinacion>>false</Determinacion>
    </EspecificadoresA>
    <EspecificadoresB>
      <Singular>>false</Singular>
      <Plural>>false</Plural>
      <MenosDe>>false</MenosDe>
      <MasDe>>false</MasDe>
      <MasOMenos>>false</MasOMenos>
      <Ninguno>>false</Ninguno>
      <Poco>>false</Poco>
      <Bastante>>false</Bastante>
      <Mucho>>false</Mucho>
      <Todo>>false</Todo>
    </EspecificadoresB>
    <Palabra>Shanghái</Palabra>
  </Elemento1>
  <Vinculo>
    <EspecificadoresA>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
      <Determinacion>true</Determinacion>
    </EspecificadoresA>
    <Palabra>centro</Palabra>
  </Vinculo>
  <Elemento2>
    <EspecificadoresA>
      <Negacion>>false</Negacion>
      <Interrogacion>>false</Interrogacion>
      <Determinacion>>false</Determinacion>
    </EspecificadoresA>
    <EspecificadoresB>
      <Singular>>false</Singular>
      <Plural>>false</Plural>
      <MenosDe>>false</MenosDe>
      <MasDe>>false</MasDe>
      <MasOMenos>>false</MasOMenos>

```

```

    <Ninguno>>false</Ninguno>
    <Poco>>false</Poco>
    <Bastante>>false</Bastante>
    <Mucho>>false</Mucho>
    <Todo>>false</Todo>
  </EspecificadoresB>
  <Palabra>finanza</Palabra>
</Elemento2>
</Relacion>

```

Figura 4.27 XML del archivo “Relación” del ejemplo.

```

<?xml version="1.0"?>
<动作 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/accion.xsd">
  <指定>>false</指定>
  <名词>
    <说明 A>
      <否定>>false</否定>
      <疑问>>false</疑问>
      <特指>>false</特指>
    </说明 A>
    <说明 B>
      <单数>>false</单数>
      <复数>>false</复数>
      <少于>>false</少于>
      <多于>>false</多于>
      <大约>>false</大约>
      <没有>>false</没有>
      <少>>false</少>
      <足够>>false</足够>
      <多>>false</多>
      <全部>>false</全部>
    </说明 B>
    <单词>2010 世博会</单词>
  </名词>
  <动词>
    <说明 C>
      <否定>>false</否定>
      <疑问>>false</疑问>
    </说明 C>
    <单词>举行</单词>

```



```

</动词>
<时间>
  <过去>false</过去>
  <现在>false</现在>
  <将来>true</将来>
<日>1</日>
  <月>五月</月>
  <年>2010</年>
</时间>
<宾语>
  <说明 A>
    <否定>false</否定>
    <疑问>false</疑问>
    <特指>false</特指>
  </说明 A>
  <说明 B>
    <单数>false</单数>
    <复数>false</复数>
    <少于>false</少于>
    <多于>false</多于>
    <大约>false</大约>
    <没有>false</没有>
    <少>false</少>
    <足够>false</足够>
    <多>false</多>
    <全部> false</全部>
  </说明 B>
  <单词>    </单词>
</宾语>
<目标>
  <说明 A>
    <否定>false</否定>
    <疑问>false</疑问>
    <特指>false</特指>
  </说明 A>
  <说明 B>
    <单数>false</单数>
    <复数>false</复数>
    <少于>false</少于>
    <多于>false</多于>
    <大约>false</大约>

```

```

    <没有>false</没有>
    <少>false</少>
    <足够>false</足够>
    <多>false</多>
    <全部>false</全部>
  </说明 B>
  <单词>    </单词>
</目标>
<空间>
  <说明 A>
    <否定>false</否定>
    <疑问>false</疑问>
    <特指>false</特指>
  </说明 A>
  <说明 B>
    <单数>false</单数>
    <复数>false</复数>
    <少于>false</少于>
    <多于>false</多于>
    <大约>false</大约>
    <没有>false</没有>
    <少>false</少>
    <足够>false</足够>
    <多>false</多>
    <全部>false</全部>
  </说明 B>
  <单词>上海</单词>
</空间>
</动作>

```

Figura 4.28 XML del archivo “Acción” del ejemplo en chino.

```

<?xml version="1.0"?>
<形容词 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/adjetivo.xsd"
  <指定>true</指定>
  <本体>
    <说明 A>
      <否定>false</否定>
      <疑问>false</疑问>
      <特指>false</特指>

```

```

</说明 A>
<说明 B>
  <单数>false</单数>
  <复数>false</复数>
  <少于>false</少于>
  <多于>false</多于>
  <大约>false</大约>
  <没有>false</没有>
  <少>false</少>
  <足够>false</足够>
  <多>false</多>
  <全部>false</全部>
</说明 B>
<单词>上海</单词>
</本体>
<形容词>
  <说明 D>
    <否定>false</否定>
    <疑问>false</疑问>
  </说明 D>
  <说明 E>
    <少>false</少>
    <足够>false</足够>
    <非常>false</非常>
  </说明 E>
  <单词>漂亮</单词>
  <单词>现代</单词>
</形容词>
</形容词>

```

Figura 4.29 XML del archivo “Adjetivo” del ejemplo en chino.

```

<?xml version="1.0"?>
<价值 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/valor.xsd"
  <指定>true</指定>
  <本体>
    <说明 A>
      <否定>false</否定>
      <疑问>false</疑问>
      <特指>false</特指>

```

```

</说明 A>
<说明 B>
  <单数>false</单数>
  <复数>false</复数>
  <少于>false</少于>
  <多于>false</多于>
  <大约>false</大约>
  <没有>false</没有>
  <少>false</少>
  <足够>false</足够>
  <多>false</多>
  <全部>false</全部>
</说明 B>
<单词>上海</单词>
</本体>
<价值>
<特点>人口</特点>
  <说明 C>
    <等于>true</等于>
    <少于>false</少于>
    <多于>false</多于>
    <大约>false</大约>
  </说明 C>
  <数值>20</数值>
  <单词>百万</单词>
</价值>
</价值>

```

Figura 4.30 XML del archivo “Valor” del ejemplo en chino.

```

<?xml version="1.0"?>
<空间 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/espacio.xsd"
  <指定 >true</指定>
  <本体>
    <说明 A>
      <否定>false</否定>
      <疑问>false</疑问>
      <特指>false</特指>
    </说明 A>
    <说明 B>

```

<单数>false</单数>
 <复数>false</复数>
 <少于>false</少于>
 <多于>false</多于>
 <大约>false</大约>
 <没有>false</没有>
 <少>false</少>
 <足够>false</足够>
 <多>false</多>
 <全部>false</全部>
 </说明 B>
 <单词>上海</单词>
 </本体>
 <空间>
 <相关术语>东方</相关术语>
 <本体>
 <说明 A>
 <否定>false</否定>
 <疑问>false</疑问>
 <特指>false</特指>
 </说明 A>
 <说明 B>
 <单数>false</单数>
 <复数>false</复数>
 <少于>false</少于>
 <多于>false</多于>
 <大约>false</大约>
 <没有>false</没有>
 <少>false</少>
 <足够>false</足够>
 <多>false</多>
 <全部>false</全部>
 </说明 B>
 <单词>中国</单词>
 </本体>
 </空间>
 <计量>
 <单词> </单词>
 <数据> </数据>

```

</计量>
</空间>

```

Figura 4.31 XML del archivo “Espacio” del ejemplo en chino.

```

<?xml version="1.0"?>
<关系 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:///e:/XML%20Bueno/relacion.xsd"
  <指定>true</指定>
  <元素 1>
    <说明 A>
      <否定>>false</否定>
      <疑问>>false</疑问>
      <特指>>false</特指>
    </说明 A>
    <说明 B>
      <单数>>false</单数>
      <复数>>false</复数>
      <少于>>false</少于>
      <多于>>false</多于>
      <大约>>false</大约>
      <没有>>false</没有>
      <少>>false</少>
      <足够>>false</足够>
      <多>>false</多>
      <全部>>false</全部>
    </说明 B>
    <单词>上海</单词>
  </元素 1>
  <联系>
    <说明 A>
      <否定>>false</否定>
      <疑问>>false</疑问>
      <特指>true</特指>
    </说明 A>
    <单词>中心</单词>
  </联系>
  <元素 2>
    <说明 A>
      <否定>>false</否定>

```


CAPÍTULO 5

Validación empírica. Probación y Evaluación de la base de datos

5.1 Introducción

A la hora de realizar la validación de las estructuras diseñadas para dar soporte al lenguaje VILA_1, se ha decidido llevar a cabo una serie de experimentos basados en la compresión de expresiones de este lenguaje visual, en que se basa la base de datos diseñada en este trabajo.

Cabe destacar que no se ha realizado una validación al uso, de una base de datos relacional, partiendo de las métricas definidas en diferentes fuentes, propias de la ingeniería del software. Estas métricas son interesantes a la hora de establecer una serie de pautas que evalúen la calidad del software, pero no proporcionan la información necesaria sobre la validez del diseño implementado para las acciones de traducción automática.

Debido a que este trabajo se centra en la capacidad de un lenguaje visual tal como VILA_1, para la comunicación entre usuarios que manejan chino y español, se ha diseñado una metodología de prueba basada en una serie de expresiones definidas por las estructuras diseñadas para VILA_1.

5.2 Metodología

La metodología de la validación se ha desglosado en una serie de etapas que a continuación de describen.

1. Selección de la población.
2. Entrenamiento de la población
3. Diseño de cuestionarios.
4. Resolución de cuestionarios por parte de los usuarios, corrección y tabulación de datos.
5. Análisis de resultados y conclusiones.

5.2.1 Selección de población

Para poder asegurar el correcto proceso de validación se debe partir de una serie de requisitos de los usuarios tales como:

- Los participantes de nacionalidad china tienen que entender los dos idiomas, chino y español. Estos requisitos son básicos, porque la traducción lleva implícita la lengua china y las tareas a realizar a la hora de resolver formularios se expresan en español.
- Los participantes tienen que tener los conocimientos sobre lingüística. Ya que la división de dominios del vocabulario y las tablas de la base de datos se refieren a los problemas del lenguaje natural.
- Conocimientos básicos y elementales del lenguaje VILA_1 que serán adquiridos en la etapa posterior.

Según los requisitos, se han seleccionado 20 alumnos chinos y 20 alumno españoles de la Universidad de León. Cabe destacar que no se ha escogido ningún elemento de población relacionado con este trabajo de investigación.

Los miembros chinos de la población tienen como idioma nativo el chino, estos han estado establecidos en España de uno a tres años y han cursado cursos regulares de lenguaje español durante uno o dos años. Los participantes españoles tienen como idioma nativo el español mientras que tienen un conocimiento nulo del chino. Ambos grupos de participantes tienen suficiente experiencia en el uso de Internet y nuevas tecnologías.

Participantes Chinos				
Nombre	Sexo	Edad	Estudio de español(años)	Experiencia de internet (año)
Participante 1	Hombre	22	2	2
Participante 2	Hombre	22	3	2
Participante 3	Hombre	23	2	4
Participante 4	Mujer	23	2	4
Participante 5	Mujer	24	2	3
Participante 6	Mujer	23	2	5
Participante 7	Mujer	21	3	2
Participante 8	Mujer	22	3	3
Participante 9	Hombre	25	2	4
Participante 10	Hombre	20	2	1
Participante 11	Hombre	18	3	1
Participante 12	Mujer	17	4	1
Participante 13	Mujer	25	4	5
Participante 14	Hombre	25	3	6
Participante 15	Mujer	22	2	3
Participante 16	Mujer	23	3	2
Participante 17	Hombre	25	3	6
Participante 18	Hombre	21	3	2
Participante 19	Mujer	20	3	1
Participante 20	Mujer	24	4	3

Tabla 5.1 Participantes chinos.

Participantes Españoles				
Nombre	Sexo	Edad	Estudio de chino (años)	Experiencia de internet (año)
Participante 1	Hombre	25	0	3
Participante 2	Hombre	22	0	1
Participante 3	Mujer	22	0	1
Participante 4	Mujer	22	0	2
Participante 5	Mujer	20	0	2
Participante 6	Hombre	19	0	1
Participante 7	Mujer	19	0	1
Participante 8	Hombre	18	0	1
Participante 9	Hombre	21	0	1
Participante 10	Hombre	24	0	3
Participante 11	Hombre	30	0	5
Participante 12	Mujer	29	0	3
Participante 13	Mujer	25	0	2
Participante 14	Mujer	25	0	3
Participante 15	Mujer	21	0	1
Participante 16	Mujer	26	0	2
Participante 17	Mujer	24	0	2
Participante 18	Mujer	20	0	1
Participante 19	Hombre	20	0	2
Participante 20	Hombre	19	0	1

Tabla 5.2 Participantes españoles.

5.2.2 Enteramiento de la población

Con el fin de comprender el lenguaje VILA_1 y demostrar su unidad se han realizado un pequeño semanario sobre la interpretación de sentencias VILA_1. Durante esta sesión se han explicado brevemente las estructuras y la interpretación de las mismas, necesarias a hora de responder los cuestionarios.

No se ha dedicado el seminario a formar expertos en VILA_1 si no que se ha realizado una exposición concisa de cómo leerlo y

comprenderlo para así poder realmente analizar la potencia, claridad y expresividad del lenguaje y si el proceso de traducción es el correcto.

5.2.3 Diseño de cuestionarios

A la hora de diseñar los cuestionarios se han llevado a cabo tres acciones diferentes:

- Establecimiento de las expresiones comunes a ambos procesos de traducción. Chino-Español. Español-Chino.
- Traducción automática a VILA_1 de cada una de las expresiones del lenguaje origen al lenguaje VILA_1 de salida.

VILA_1 chino a VILA_1 español.

VILA_1 español VILA_1 chino.

- Establecimiento de 5 posibles respuestas por cada expresión. Habrá una única respuesta válida por cada expresión, de manera que esta será el significado correcto de la expresión VILA_1 traducida con respecto a la expresión origen.

Con respecto a la primera acción se han definido una serie de expresiones que en algunos casos pueden contener problemas de ambigüedad para demostrar como la capacidad de VILA_1 para resolver este tipo de problemas.

A continuación se presentan 10 ejemplos de expresiones (5 chinas y 5 españolas) para ser traducidas y que se encuentran incluidas en los cuestionarios. En total se han diseñado 40 expresiones, por motivos de legibilidad se muestran únicamente 10 frases junto con sus 5 opciones de respuesta asociadas, entre las que se encuentra la correcta, que lógicamente es la expresión origen y otras con un significado similar con el fin de medir eficazmente la validez de la traducción automática.

Español – chino

1. *Dos bocadillos cuestan 3 euros.*



Figura 5.1 La expresión de formato VILA_1 español a chino

- A. 2 个夹肉面包花费3 欧元。
- B. 2 个夹肉面包和 3 欧元。
- C. 2 个夹肉面包不要 3 欧元。
- D. 2 个夹肉面包送 3 欧元。

2. *León es una ciudad tranquila y bonita.*



Figura 5.2 La expresión de formato VILA_1 español a chino

- A. León 很美丽，但不宁静。
- B. León 不美丽，但是宁静。
- C. León 是一个城市。
- D. *León 是一个美丽宁静的城市*

3. *VILA_1 es un lenguaje visual.*

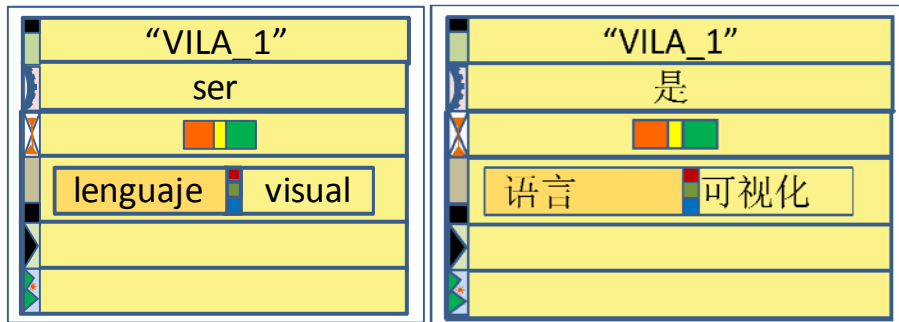


Figura 5.3 La expresión de formato VILA_1 español a chino

- A. VILA_1 是一种语言。
- B. VILA_1 是可视化的。
- C. *VILA_1 是一种可视化的语言。*
- D. 可以看看 VILA_1 是一种语言。

4. *Paella es una rica comida de España.*



Figura 5.4 La expresión de formato VILA_1 español a chino

- A. *海鲜饭是美味的西班牙食物。*
- B. 海鲜饭是一种很贵的西班牙食物。
- C. 西班牙美味的食物是海鲜饭。
- D. 美味的西班牙是海鲜饭食物。

5. *La playa de Andalucía es famosa siempre hay mucha gente.*



Figura 5.5 La expresión de formato VILA_1 español a chino

- A. 在 Andalucía 的海滩经常有很多人著名的人。
- B. 在 Andalucía 著名的海滩经常有很多人。
- C. 很多人在 Andalucía 的海滩，经常很著名。
- D. Andalucía 的海滩因为经常有很多人而很著名。

Chino – español.

- 1. 我是中国学生。



Figura 5.6 La expresión de formato VILA_1 chino a español

- A. ***Yo soy alumno de china.***
- B. Yo soy alumno en china.
- C. Yo soy alumno gusta china.
- D. Yo soy alumno conozco china.

2. 我很喜欢西班牙。



Figura 5.7 La expresión de formato VILA_1 chino a español

- A. Me gusta ir a España más veces.
- B. Me gusta vivir en España mucho tiempo.
- C. ***Me gusta España mucho.***
- D. No me gusta España.

3. 北京是中国的首都，在中国的北方。

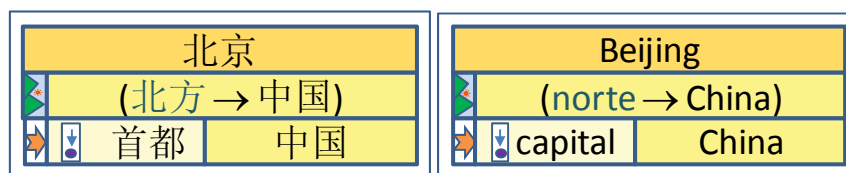


Figura 5.8 La expresión de formato VILA_1 chino a español

- A. El norte de Beijing es la capital de China.
- B. ***Beijing es la capital de China, está al norte de China.***
- C. La capital está norte de China, Beijing es de China.
- D. El norte de China es la capital de Beijing.

4. 黑色的狗跑得很快。

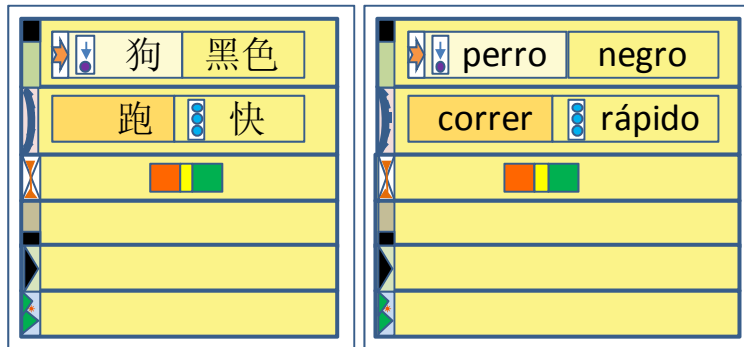


Figura 5.9 La expresión de formato VILA_1 chino a español

- A. En la noche el perro corre muy rápido.
- B. El perro no corre muy rápido.
- C. El perro negro corre muy rápido.**
- D. Corre rápido para coger el perro negro.

5. 星期五我们将去海滩

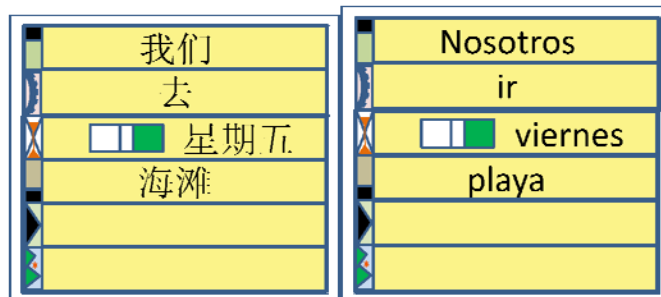


Figura 5.10 La expresión de formato VILA_1 chino a español

- A. Viernes nosotros fuimos a la playa.
- B. Viernes nosotros iremos a la playa.**
- C. Viernes nosotros iremos a la come paella.
- D. Viernes nosotros fuimos

5.2.4 Resolución de cuestionarios por parte de los usuarios

A la hora de resolver los cuestionarios se llevan a cabo las siguientes acciones según el idioma origen y destino:

Para traducción de chino a español:

Resolución del cuestionario por parte del usuario español a partir de la proposición: “Escoja entre las 5 expresiones en español la que más se ajuste al significado de la expresión VILA_1 propuesta”. Esta última traducida por el sistema a partir de la expresión VILA_1 origen.

Para traducción español a chino:

Resolución del cuestionario por parte del usuario chino a partir de la proposición: “Escoja entre las 5 expresiones en chino la que más se ajuste al significado de la expresión VILA_1 propuesta”. Esta última traducido por el sistema a partir de la expresión VILA_1 origen.

A continuación se presentan una serie de gráficos donde se pueden observar los resultados de comprensión de cada expresión donde se valida si el proceso de traducción ha sido correcto.

5.2.5 Análisis de resultados y conclusiones

A continuación se presentan una serie de gráficos donde se pueden observar los resultados de comprensión de cada expresión donde se valida si el proceso de traducción ha sido correcto, desde el punto de vista de los usuarios.

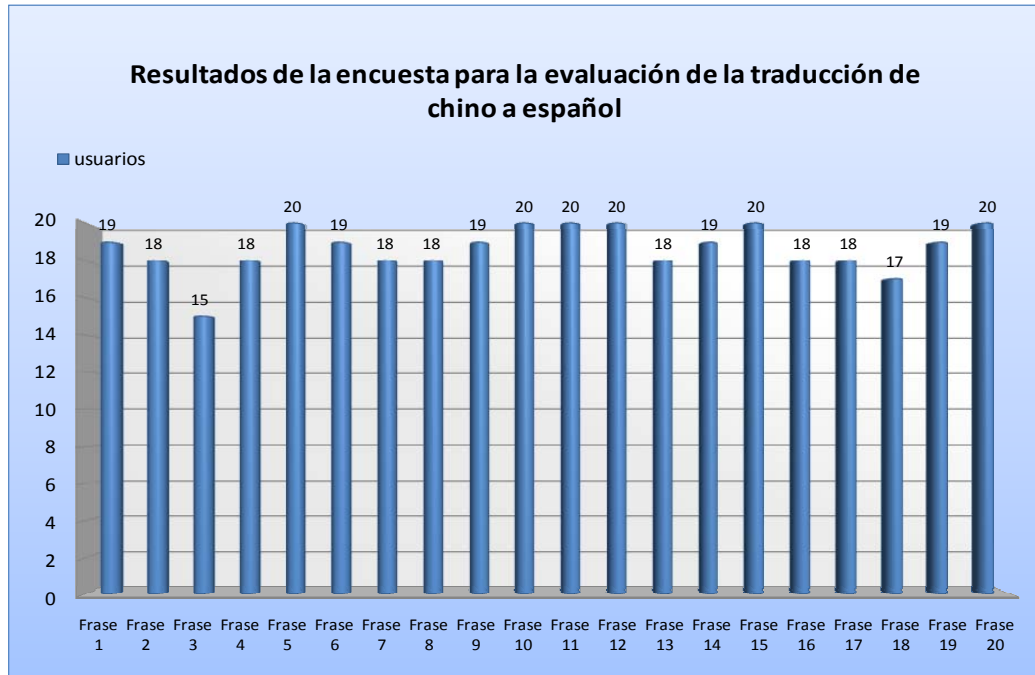


Figura 5.11 Gráfico de resultados de la evaluación de la traducción de chino a español

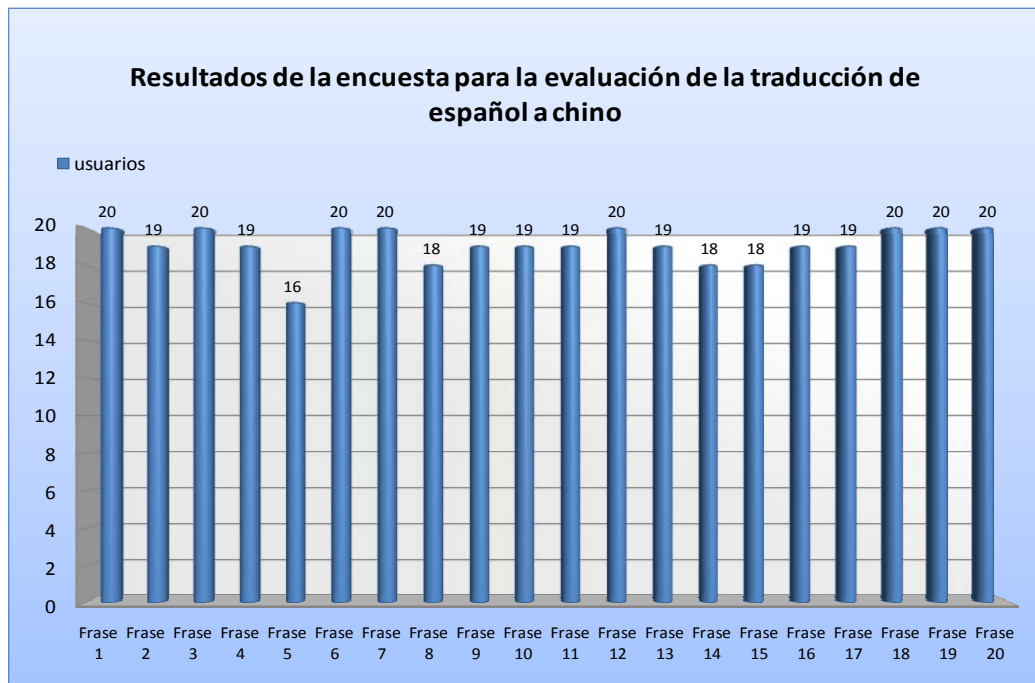


Figura 5.12 Gráfico de resultados de la evaluación de la traducción de español a chino

Como se puede observar, los usuarios son capaces de entender la mayor parte expresiones únicamente con el seminario de dos horas dedicado en la segunda etapa. La media de usuarios sobre 20 que han sido capaces de comprender la expresión traducida de manera automática por el sistema es de 93.25% de chino a español y 95.5% de español a chino.

Haciendo los cálculos por separado entre usuarios españoles y usuarios chinos, resulta más favorables los resultados del segundo grupo, ya que estos están más adaptados a un nivel mayor de abstracción que suponen las estructuras lingüísticas asociados a una lengua foránea.

Cabe destacar que existen una serie de expresiones como son la 3“*北京是中国的首都, 在中国的北方*” de la traducción chino español y la 5 “*En la playa Andalucía es famosa siempre hay mucha gente*” de la traducción de español a chino, que resultan más difíciles de entender por parte del usuario debido a su compleja estructura VILA_1, aunque su traducción ha sido comprobada realizando esta acción manualmente.

Sin embargo se puede asegurar que con un tiempo de formación mayor dedicado a la etapa 2.2, estas estructuras serán asequibles para la mayoría de los usuarios, a la hora de ser comprendidas.

En un futuro y gracias a los resultados extraídos de esta fase de validación, sería conveniente realizar un seminario más de especialización en VILA_1 amplio y así conformar una batería de expresiones más complejas.

CAPÍTULO 6

Conclusiones finales y trabajos futuros

6.1 Conclusiones

En esta tesis doctoral se ha llegado a las siguientes conclusiones:

- Se ha comprobado que el lenguaje VILA_1 es útil para la comunicación entre personas de habla china y de habla española.
- Se ha comprobado que la sintaxis gráfica de VILA_1 reduce extraordinariamente la dificultad propia de la traducción entre dos idiomas tan diferentes como el chino y el español.
- Se ha comprobado que la no correspondencia exacta entre la semántica de un buen número de palabras de ambos idiomas genera problemas e introduce dificultad en la comprensión de las frases.
- Se ha comprobado que la introducción de otros idiomas en el sistema es sencilla, con lo que su potencia comunicativa crecerá extraordinariamente.

6.2 Trabajos futuros

Sobre las ideas y experiencias obtenidas en esta tesis doctoral, hay varios trabajos que se pueden proponer. Por ejemplo se citan los siguientes:

- Añadir más idiomas a la base de datos. En un futuro, se pueden añadir otros idiomas de uso común.
- Añadir algunas frases comunes a la base de datos. En la vida cotidiana hay muchos proverbios o dichos, que a la gente le gusta usar para expresarse mejor. Pero es muy difícil clasificar sus elementos. Entonces, existe la posibilidad de establecer una estructura especial para meter estos proverbios o dichos.
- Investigar los problemas que puedan surgir entre diferentes tipos de idiomas (SOV frente a SVO).

Abreviaturas

-

ALPAC – Automatic Language Processing Advisory Committee.
ANSI – American National Standards Institute.
API – Application Programming Interface.
BDOO – Bases de Datos Orientadas a Objetos.
DBMS – Data Base Management System.
EURAMIS – European Advanced Multilingual Information System.
HTTP – Hypertext Transfer Protocol.
IA – Inteligencia Artificial.
IBM – International Business Machines.
IMS – Information Management System.
IDMS – Integrated Data Management System.
IDS – Integrated DataStore.
ISO – International Organization for Standardization.
ISST – Italian Syntactic-Semantic Treebank.
JDBC – Java Database Connectivity.
LDE – Lenguaje de Dominio Específico.
LVDE – Lenguaje Visual de Dominio Específico.
MDR – Machine Dictionary Readable.
ODBC – Open Database Connectivity.
ORMs – Object-relational mapping tool.
RAE – Real Academia Española.
RDBMS – Relational database management system.
SGBD – Sistemas de gestión de bases de datos.
SGBDOO – Sistemas de gestión de bases de datos orientadas a objetos.
SGBDR – Sistemas de gestión de bases de datos relacionales.
SQL – Structured Query Language.
UTF-8 – 8 Bit Unicode Transformation Format.
VILA – Visual Language.
VILA_1 – Visual Language 1.
WSDL – Web Services Description Language.
XHTML – Extensible Hypertext Markup Language.
XML – Extensible Markup Language.
XML Schema – Extensible Markup Language Schema.

Referencias

Nota: Todos los enlaces a los documentos electrónicos que aparecen en las siguientes referencias han sido comprobados con fecha 12 de abril de 2010

1. **Alonso Álvarez, Ángel.** *Introducción a VILA_1. El lenguaje de la accesibilidad. Hacia un mundo sin barreras lingüísticas.* León : Instituto de Automática y Fabricación (Universidad de León), 2009.
2. **Alonso Álvarez, Ángel and Fernández Moro, Mari Paz.** *Manual de Técnicas de Estudio.* León : Everest, 1991. ISBN: 84-241-2716-1.
3. *Una propuesta de lenguaje exclusivamente gráfica para la comunicación humana.* **Alonso Álvarez, Ángel.** 14, 1992, Comunicación y Lenguaje.
4. *Hacia un lenguaje gráfico en la enseñanza.* **Alonso Álvarez, Ángel, Foces Morán, Jose María and Fernández Moro, Mari Paz.** 18, 1993, Comunicación y Lenguaje.
5. *Transparencias-resumen en la enseñanza de electrónica.* **Alonso, Ángel and Ferrero, Miguel.** Madrid : Universidad Politécnica de Madrid, 1994. Actas del primer congreso sobre tecnologías aplicadas a la enseñanza de la electrónica.
6. *Un lenguaje visual en la enseñanza técnica.* **Ferrero, Miguel and Alonso, Ángel.** Las Palmas de Gran Canaria : Universidad de Las Palmas, 1995. Actas de las 6ª Jornadas de Tecnología Electrónica.
7. *Estructura gráfica de la electrónica mediante sistemas multimedia.* **Alonso, Ángel, et al.** Sevilla : Universidad de Sevilla, 1996. Actas del 2º congreso de tecnologías aplicadas a la enseñanza de la electrónica.
8. *Estructuración gráfica de la electrónica mediante sistemas multimedia.* **Alonso, Ángel, et al.** 3, Chile : s.n., 1998, Información Tecnológica, Vol. 9.
9. **Sierra, R.** *Tesis doctorales y trabajos de investigación científica.* Madrid : Paraninfo, 1994.

-
10. **Rozenberg, G. and Salomaa, A.** *Handbook of Formal Language Theory*. s.l. : Springer Verlag, 1997. Vols. 1-3.
 11. **RAE.** *Diccionario de la Lenguaje Española*. Vigésima segunda edición. Madrid : Real Academia Española, 2001.
 12. **Harrison, M. A.** *Introduction to Formal Language Theory*. Reading : Addison-Wesley, 1977.
 13. **Manning, Christopher D. and Schütze, Hinrich.** *Foundations of Statistical Natural Language Processing*. Cambridge : The MIT Press, 1999.
 14. **Berns, M. and Matsuda, K.** Applied Linguistics: Overview and History. *Encyclopedia of Language & Linguistics*. s.l. : Elsevier, 2006, Vol. II, pp. 394-405.
 15. **Mitkov, Ruslan.** *The Oxford Handbook of Computational Linguistics*. Oxford : Oxford University Press, 2003.
 16. **Chomsky, Noam.** *Syntactic Structures*. Paris : The Hague/Paris: Mouton, 1957.
 17. **Hutchins, John.** ALPAC: the (in)famous report. [book auth.] Sergei Nirenburg, H. L. Somers and Yorick Wilks. *Readings in machine translation*. Cambridge : The MIT Press, 2003, pp. 131-135.
 18. **Woods, William.** *Transition Networks for Natural Language Analysis*. Cambridge, Mass. : Harvard Computer Laboratory Report, 1969.
 19. **Anderson, John M.** *The grammar of case: towards a localistic theory*. London : Cambridge University Press, 1976.
 20. **Woods, William.** *Semantics for a Question-Answering System*. Cambridge : Harvard University, 1967.
 21. **Winograd, T.** *Understanding natural language*. Orlando : Academic Press, 1972.
 22. *Progress in natural language understanding--- an application in lunar geology.* **Woods, W.** s.l. : AFIPS Conference Proceedings, 1973. Proceedings of the 1973 National Computer Conference. pp. 441-450.
 23. *Developing a Natural Language Interface to ComplexData.* **Hendrix, G.G., et al.** 1978, ACM Transactions on Database Systems, pp. 105-147.
 24. *Panorama de la lingüística computacional en Europa.* **Martí Antonín, María Antonia.** 1999, Revista española de lingüística aplicada, pp. 11-24.
 25. **Dorr, Bonnie J., Jordan, Pamela W. and Benoit, John W.** A Survey of Current Paradigms in Machine Translation. *Advances in Computers*. s.l. : Elsevier, 1999, Vol. 49, pp. 1-68.
 26. **Sperberg-McQueen, C. M. and Burnard, Lou.** *Guidelines for Electronic Text Encoding and Interchange*. s.l. : ACH, ACL, ALLC, 1993.

-
27. EAGLES. *Expert Advisory Group on Language Engineering Standards*. [Online] 1996. <http://www.ilc.cnr.it/EAGLES96/home.html>.
28. *Intervals*. **Date, C. J., Darwen, Hugh and Lorentzos, Nikos A.** 2003, *Temporal Data & the Relational Model*, pp. 77-88.
29. **Gibbon, Dafydd, Mertins, Inge and Moore, Roger K.** Computational Linguistics. *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. s.l. : MIT Press, 2001, Vol. 27, pp. 149-150.
30. *SpeechDat Experiences in Creating Large Multilingual*. **Draxler, Christoph, Heuvel, Henk van den and Tropf, Herbert S.** 1998, Proc. of the First International Conference on Language Resources and Evaluation, pp. 361–366.
31. **Vossen, P.** *EuroWordNet: A multilingual database with lexical semantic networks*. Dordrecht : Kluwer, 1998.
32. *Panorama de la investigación en lingüística informática*. **Gómez Guinovart, J., et al.** 1999, *Revista Española de Lingüística Aplicada*, pp. 11-24.
33. Babylon. [Online] 2010. [Cited: abril 1, 2010.] <http://www.babylon.com/>.
34. SYSTRAN. [Online] 2010. [Cited: abril 01, 2010.] <http://www.systran.co.uk/>.
35. PROMT. [En línea] 2010. [Citado el: 1 de abril de 2010.] <http://www.promt.com/>.
36. **Blatt, A.** The EURAMIS Project. [book auth.] A. LAUER, et al. *Übersetzungswissenschaft im Umbruch*. Tübingen : Festschrift für Wolfram Wilss, 1996, pp. 131-134.
37. *Overview of AlethGen*. **Coch, Josó.** Herstmonceux : INLG-96, 1996. Proceedings of the International Workshop on.
38. *The organization and use of information: Contributions of information science, computational linguistics and artificial intelligence*. **Walker, Donald E.** 2007, *Journal of the American Society for Information Science*, pp. 347 - 363.
39. *Corpus Linguistics*. **Hunston, S.** 2006, *Encyclopedia of Language & Linguistics*, pp. 234-248.
40. **Abeillé, Anne.** *Treebanks: building and using parsed corpora*. s.l. : Springer, 2003.
41. **Biber, Douglas, Conrad, Susan and Reppen, Randi.** *Corpus Linguistics: Investigating Language Structure and Use*. s.l. : Cambridge University Press, 2004.
42. *Principles for encoding machine readable dictionaries*. **Jean Véronis, Nancy Ide, Warwick-Armstrong, Susan and Calzolari, Nicoletta.** Finland : University of Tampere, 1992. Fifth Euralex International Congress,.
43. **Marriott, Kim and Meyer, Bernd E.** *Visual language theory*. s.l. : Springer, 1998.


-
44. **Rudolf, Arnheim.** *Visual Thinking*. Berkeley : University of California Press, 1969.
 45. **Narayanan, N. Hari and Hübscher, Roland.** Visual Language Theory: Towards a Human-Computer Interaction Perspective. [book auth.] Kim Marriott and Bernd Meyer. *Visual Language Theory*. New York : Springer, 1998, pp. 87-128.
 46. *Visual Language and Converging Technologies in the Next 10-15 Years (and Beyond)*. **Horn, Robert E.** 2001. National Science Foundation Conference on Converging Technologies (Nano-Bio-Info-Cogno) for Improving Human Performance.
 47. **Frank, M. R. and Foley, J. D.** *A pure reasoning engine for programming by example*. Graphics, Visualization and Usability Center, Georgia Institute of Technology. Atlanta (Estados Unidos) : s.n., 1994. Technical Report GIT-GVU-94-11.
 48. **Tufte, E. R.** *Envisioning Information*. Cheshire : Graphics Press, 1990.
 49. *A principled taxonomy of software visualization*. **Price, B. A., Baecker, R. M. and Small, I. S.** 1993, Journal of Visual Language and Computing, pp. 211-266.
 50. *Why a diagram is (sometimes) worth ten thousand words*. **Larkin, J. H. and Simon, H. A.** 1987, Cognitive Science, pp. 65-99.
 51. **CACM.** Special section on educational technology. *Communications of the ACM*. Abril 1996.
 52. **Cox, Philip T.** Visual Programming Languages. [book auth.] Benjamin W. Wah. *Encyclopedia of Computer Science and Engineering*. s.l. : Wiley-Interscience, 2009.
 53. **Burnett, Margaret M.** Visual Programming. [book auth.] J. G. Webster. *Encyclopedia of Electrical and Electronics Engineering*. New York : Wiley, 1999.
 54. **Antonietti, A.** Why does mental visualization facilitate problem solving? [book auth.] R. H. Logie and M. Denis. *Mental Images in Human Cognition*. Amsterdam : Elsevier Science Publishers, 1991, pp. 211-227.
 55. **Horn, Robert E.** *Visual Language: Global Communication for the 21st Century*. Primera edición. s.l. : Macrovu Inc., 1999.
 56. *Domain-specific language design requires feature descriptions*. **van Deursen, A. and Klint, P.** 2002, Journal of Computing and Information Technology, pp. 1-17.
 57. *A preliminary study on various implementation approaches of domain-specific language*. **Kosar, Tomaž, et al.** 2008, Information and Software Technology, pp. 390-405.
 58. *A domain-specific visual language for domain model evolution*. **Sprinkle, Jonathan and Karsai, G. Gabor.** 2004, Journal of Visual Languages & Computing, pp. 291-307.
 59. *Database as a genre of new media* . **Manovich, Lev.** 2000, AI & Society, pp. 176-183.





-
60. **Orallo, José Hernández.** *La Disciplina de los Sistemas de Bases de Datos.* Valencia : s.n., 2002.
61. **Dai Xinyu, Yin Cunyan, Chen Jiajun and Zheng Guoliang.** *Machine Translation: Past, Present, future.* 2004.
62. *Machine translation: boundaries and practice in the late '90s.* **Lewis, T.** London : s.n., 2002.
63. *The Current Situation and Problems in Machine Translation.* **Feng, Zhiwei.** 2003.
64. **Arnheim, R.** *Visual Thinking.* Berkeley : University of California Press, 1969.
65. **Saint-Martin, F.** *Semiotics of Visual Language.* Bloomington : Indiana University Press, 1990.
66. **Group, Miniwatts Marketing.** Internet World Stats - Web Site Directory. *INTERNET WORLD USERS BY LANGUAGE.* [Online] 2001 - 2010. [Cited: apr 25, 2010.] <http://www.internetworldstats.com/stats7.htm>.
67. **Bray, Tim, et al.** Extensible Markup Language (XML) 1.0 (Fifth Edition). [Online] 2008. <http://www.w3.org/TR/REC-xml/>.
68. **UDDI.** The UDDI Technical White Paper. [Online] 2000. <http://uddi.xml.org/>.
69. Extensible Markup Language (XML). *W3C.* [Online] Junio 1, 2009. [Cited: Junio 1, 2009.] <http://www.w3.org/xml/>.
70. **Baeza-Yates, R and Ribeiro-Neto, B.** *Modern Information Retrieval.* s.l. : Addison Wesley - Longman Publishing co., 1999.
71. *Entropy: architecture and performance of an enterprise desktop grid system.* **Chien, Andrew, et al.** s.l. : Elsevier, 2003, Journal of Parallel and Distributed Computing, pp. 597-610.
72. *Three models for the description of language.* **Chomsky, Noam.** 1956, IEEE Transactions on Information Theory, pp. 113-124.
73. **Newcomer, Eric.** *Understanding Web services: XML, WSDL, SOAP, and UDDI.* s.l. : Addison-Wesley, 2002.
74. *Identification of ambiguous queries in web search.* **Song, Ruihua, et al.** s.l. : Elsevier, 2009, Information Processing & Management, pp. 216-229.
75. **ISO.** The iso 9126 standard. [Online] 1991. <http://www.issco.unige.ch/en/research/projects/ewg96/node13.html>.
76. **Sun Microsystems, Inc.** Java™ Remote Method Invocation Specification. [Online] 2006. <http://java.sun.com/javase/6/docs/platform/rmi/spec/rmi-title.html>.
77. **Mitra, Nilo.** SOAP Versión 1.2 Parte 0: Fundamentos. *W3C.* [Online] Junio 24, 2003. [Cited: Junio 1, 2009.] <http://www.w3c.es/Traducciones/es/TR/2003/REC-soap12-part0-20030624/>.

-
78. **Von Riegen, Claus, et al.** UDDI Version 2.03 Data Structure Reference. *UDDI*. [Online] Julio 19, 2002. [Cited: Junio 1, 2009.] <http://uddi.org/pubs/DataStructure-V2.03-Published-20020719.htm>.
79. **Srinivasan, R.** *RPC: Remote Procedure Call Protocol Specification Version 2*. s.l. : DDN Network Information Center, 1995.
80. **Kerievsky, J.** *Refactoring to patterns*. s.l. : Addison-Wesley, 2004.
81. **Kohavi, Z., Hamming, R. W . and Feigenbaum, E. A.** *Switching and Finite Automata Theory: Computer Science Series*. s.l. : McGraw-Hill Higher Education, 1990.
82. **Salomaa, A and Sneddon, IN.** *Theory of automata*. s.l. : Pergamon Press Reprint, 1969.
83. **Myhill, J.** *Linear bounded automata*. Ohio : Wright Air Development Division, 1960.
84. *On computable numbers, with an application to the Entscheidungsproblem.* **Turing, A. M.** 1936, Proc. of London Mathematical Society 2, no. 42 and no. 43, pp. 230-236,544-546.
85. *Investigating the limits of SOAP performance for scientific computing.* **Chiu, K., Govindaraju, M and Bramley, R.** 2002, High Performance Distributed Computing, 2002. HPDC-11 2002. Proceedings. 11th IEEE International Symposium on, pp. 246 - 254.
86. **Merriam-Webster.** *Merriam-Webster's Collegiate Dictionary* . s.l. : Merriam-Webster, 2003.
87. *A chomsky hierarchy of isotonic array grammars and languages.* **Cook, Curtis. R and Wang, Patrick Shen-Pei.** 1978, Computer Graphics and Image Processing, pp. 144-152.
88. *Backus Normal form vs. Backus Naur form.* **Knuth, Donald E.** s.l. : ACM, 1964, Communications of the ACM, pp. 735 - 736.
89. **Goldfarb, Charles F. and Rubinsky, Yuri.** *The SGML handbook*. s.l. : Oxford University Press, 1990.
90. **Pemberton, Steven.** XHTML 1.0: The Extensible HyperText Markup Language. [Online] 2000. [Cited: Julio 1, 2009.] <http://www.w3.org/TR/xhtml1>.
91. **Ion, Patrick and Miner, Robert.** Mathematical markup language (MathML). [Online] 1999. [Cited: Julio 1, 2009.] <http://www.w3.org/TR/REC-MathML/>.
92. **Lamport, Leslie.** *LATEX: A document preparation system: user's guide and reference manual*. Segunda edición. Boston : Addison-Wesley Longman Publishing Co., Inc, 1994.
93. **Fowler, M.** *Refactoring: Improving the design of existing code*. s.l. : Addison-Wesley, 2000.
94. **McGuinness, Deborah L. and Van Harmelen, Frank.** OWL Web Ontology Language. [Online] 10 2004. [Cited: Julio 1, 2009.] <http://www.w3.org/TR/owl-features/>.

-
95. **Houde, Stephanie and Hill, Charles.** What do Prototypes Prototype. [book auth.] M. G. Helander, T. K. Landauer and P. V. Prabhu. *Handbook of human-computer interaction*. New York : Elsevier Science Inc, 1997, pp. 367-381.
96. **Turner, J. Rodney.** Pilot Study. *Encyclopedia of Social Measurement*. s.l. : Academic Press, 2005, pp. 63-69.
97. **Suehring, Steve.** *Mysql bible*. New York : Wiley, 2002.
98. **Gosling, James, et al.** *Java Language Specification: The Java Series*. 2nd edition. Boston : Addison-Wesley Longman Publishing Co., Inc, 2000.
99. **Horstmann, Cay S. and Cornell, Gary.** *Core Java 2 Volument II Características avanzadas*. Madrid : Pearson Educación,S.A., 2006.
100. **ISO.** *Codes for the representation of languages Part 6: Alpha-4 representation for comprehensive coverage of language variation*. Switzerland : International Organization for Standardization, 2007.
101. —. *Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes*. Switzerland : International Organization for Standardization, 2006.
102. **Gordon, Wendy and Langmaid, Roy.** *Qualitative market research: A practitioner's and buyer's guida*. s.l. : Gower , 1988.
103. **Koffka, Kurt.** *Principles of Gestalt Psychology*. s.l. : Routledge, 1999.
104. *Recombinant Chinese Pinyin system for efficient processing of information in Chinese*. **Tong, Harry and Jin, Linda.** November 16, 1998, Computer Standards & Interfaces, Vol. 20, pp. 25-29.
105. **Brown, Allen, et al.** XML Schema: Formal Description. [Online] 2001. [Cited: agosto 31, 2009.] <http://www.w3.org/TR/xmlschema-formal/>.
106. Internet World Stats. [Online] 2009. [Cited: agosto 27, 2009.] <http://www.internetworldstats.com/stats7.htm>.
107. *Language translation module for instant messaging systems*. **Rutaks, A.** 2008. Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2008 Proceedings of International Conference on . pp. 630 - 631.
108. *An Exploratory Investigation into Instant Messaging Preferences in Two Distinct*. **Guo, Z., et al.** 2008. Professional Communication, IEEE Transactions on. pp. 396 - 415.
109. *qqqngaknlak*. **Dorro.B.J.** 1098, agagb, pp. 44-45.
110. *A feasibility study for Chinese-Spanish statistical*. **Banchs, R. E., et al.** 2006, Proc. of the 5th Int. Sym. on Chinese Spoken Language Processing.

-
111. **Banchs, Rafael E. and Li, Haizhou.** Exploring Spanish-morphology effects on Chinese–Spanish SMT. [book auth.] Iñaki Alegria, Lluís Màrquez and Kepa Sarasola. *Mixing Approaches to Machine Translation*. 2008, pp. 49-52.
 112. *System Combination for Machine Translation of Spoken and Written Language.* **Matusov, E., et al.** 2008, Transactions on Audio, Speech, and Language Processing , pp. 1222 - 1237.
 113. *Logical and semantic database integration.* **Kohler, J., et al.** 2000, Proceedings IEEE International Symposium on Bio-Informatics and Biomedical Engineering, pp. 77 - 80 .
 114. **Winskel, Glynn.** *The formal semantics of programming languages: an introduction.* s.l. : MIT Press, 1993.



- cdcol (1)
- information_schema (17)
- lenguajebit (1)
- mysql (23)
- phpmyadmin (8)
- soccer (1)
- test (0)
- vegazana (52)
- vila (6)
- webauth (1)

请选择数据库

localhost

服务器版本: 5.0.51a

- ▶ Protocol version: 10
- 服务器: localhost via TCP/IP
- ▶ 用户: root@localhost
- MySQL 字符集: **UTF-8 Unicode (utf8)**
- MySQL 连接校对: utf8_unicode_ci

创建一个新的数据库

整理

- 显示 MySQL 的运行信息
- 显示 MySQL 的系统变量
- 进程
- 字符集和整理
- Storage Engines
- Reload privileges
- 权限
- 数据库
- 导出
- Import

phpMyAdmin - 2.11.4

- ▶ MySQL client version: 5.0.51a
- ▶ Used PHP extensions: mysql
- Language: 中文 - Chinese simplified
- 主题 / 风格: Original
- ▶ Font size: 82%
- phpMyAdmin 文档
- phpMyAdmin wiki
- phpMyAdmin 官方网站
- ▶ [ChangeLog] [Subversion] [Lists]

您配置文件中的设定与 MySQL 默认权限账户对应(没有密码的 root)。您的 MySQL 服务器使用默认值运行当然没有问题, 不过这样的话, 被入侵的可能性会很大, 您真的应该先补上这个安全漏洞。

phpMyAdmin 服务器: localhost ▾ 数据库: vila

结构 SQL 搜索 查询 导出 Import Designer 操作 权限 删除

	表	操作	记录数	类型	整理	大小	多余
<input type="checkbox"/>	adjetivos		186	MyISAM	utf8_unicode_ci	12.3 KB	-
<input type="checkbox"/>	adverbios		209	MyISAM	utf8_unicode_ci	15.1 KB	-
<input type="checkbox"/>	lugar		248	MyISAM	utf8_unicode_ci	17.9 KB	-
<input type="checkbox"/>	numerales		12	MyISAM	utf8_unicode_ci	2.5 KB	-
<input type="checkbox"/>	sustantivos		2,492	MyISAM	latin1_swedish_ci	192.1 KB	-
<input type="checkbox"/>	tiempo		71	MyISAM	utf8_unicode_ci	5.1 KB	-
<input type="checkbox"/>	verbos		2,518	MyISAM	latin1_swedish_ci	151.6 KB	-
	7 个表	总计	5,736	MyISAM	utf8_unicode_ci	396.6 KB	0 字节

全选 / 全部不选 选中项:

打印预览 数据字典

在数据库 vila 中创建一个新表

名字: Number of fields:

Open new phpMyAdmin window

数据库

vila (7)

vila (7)

- adjetivos
- adverbios
- lugar
- numerales
- sustantivos
- tiempo
- verbos

服务器: localhost | 数据库: vila | 表: sustantivos

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除

字段	类型	整理	属性	Null	默认	额外	操作
<input type="checkbox"/> Id	int(11)			否		auto_increment	
<input type="checkbox"/> Espanol	text	utf8_unicode_ci		否			
<input type="checkbox"/> Dominio	text	utf8_unicode_ci		否			
<input type="checkbox"/> English	text	utf8_unicode_ci		否			
<input type="checkbox"/> Domain	text	utf8_unicode_ci		否			
<input type="checkbox"/> 中文	text	utf8_unicode_ci		否			
<input type="checkbox"/> 领域	text	utf8_unicode_ci		否			
<input type="checkbox"/> Pinyin	text	utf8_unicode_ci		否			

全选 / 全部不选 选中项:

打印预览 关系查看 规划表结构

添加 1 字段 于表结尾 于表开头 于 Id 之后 执行

索引:				已使用空间	
键名	类型	基数	操作	字段	用法
PRIMARY	PRIMARY	2492		Id	数据 169,028 字节
在第 1	列创建索引	执行			索引 27,648 字节
					统计 196,676 字节

行统计	
语句	值
格式	动态
整理	latin1_swedish_ci
行数	2,492
行长度 <input type="checkbox"/>	67
行大小 <input type="checkbox"/>	79 字节
下一个 Autoindex	
创建时间	2009 年 06 月 26 日 10:41
最后更新时间	2009 年 06 月 26 日 09:41

Open new phpMyAdmin window

phpMyAdmin

服务器: localhost | 数据库: vila | 表: sustantivos

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除

显示行 0 - 29 (2,492 总计, 查询花费 0.0298 秒)

SQL 查询:
 SELECT
 FROM 'sustantivos'
 LIMIT 0, 30

Profiling [编辑] [解释 SQL] [创建 PHP 代码] [刷新]

显示: 30 行, 开始行数: 30
 以 水平 模式显示, 并且在 100 个单元格后重复标题
 页号: 1

主键排序: 无

			Id	Espanol	Dominio	English	Domain	中文	领域	Pinyin
<input type="checkbox"/>			1	carne	alimentos, bebidas y agricultura	meat	food, drink and farming	肉	食物, 饮料与农业	rou
<input type="checkbox"/>			2	ternera	alimentos, bebidas y agricultura	calf	food, drink and farming	小牛肉	食物, 饮料与农业	xiaoniurou
<input type="checkbox"/>			3	cordero	alimentos, bebidas y agricultura	mutton	food, drink and farming	羊肉	食物, 饮料与农业	yangrou
<input type="checkbox"/>			5	solomillo	alimentos, bebidas y agricultura	steak	food, drink and farming	牛脊肉	食物, 饮料与农业	niujirou
<input type="checkbox"/>			6	bistec	alimentos, bebidas y agricultura	steak	food, drink and farming	牛排	食物, 饮料与农业	niupai
<input type="checkbox"/>			7	chuleta	alimentos, bebidas y agricultura	cutlet	food, drink and farming	连骨肉	食物, 饮料与农业	liangurou
<input type="checkbox"/>			8	estofado	alimentos, bebidas y agricultura	stew	food, drink and farming	炖肉	食物, 饮料与农业	dunrou
<input type="checkbox"/>			11	jamón	alimentos, bebidas y agricultura	ham	food, drink and farming	火腿	食物, 饮料与农业	huotui
<input type="checkbox"/>			12	tocino	alimentos, bebidas y agricultura	bacon	food, drink and farming	咸肉	食物, 饮料与农业	xianrou
<input type="checkbox"/>			13	salchicha	alimentos, bebidas y agricultura	sausage	food, drink and farming	香肠	食物, 饮料与农业	xiangchang
<input type="checkbox"/>			14	morcilla	alimentos, bebidas y agricultura	sausage	food, drink and farming	血肠	食物, 饮料与农业	xuechang
<input type="checkbox"/>			15	fiambres	alimentos, bebidas y agricultura	deli meats	food, drink and farming	冷盘	食物, 饮料与农业	lengpan
<input type="checkbox"/>			16	pollo	alimentos, bebidas y agricultura	chicken	food, drink and farming	鸡	食物, 饮料与农业	ji
<input type="checkbox"/>			17	pavo	alimentos, bebidas y agricultura	turkey	food, drink and farming	火鸡	食物, 饮料与农业	huoji
<input type="checkbox"/>			18	pato	alimentos, bebidas y agricultura	duck	food, drink and farming	鸭	食物, 饮料与农业	ya
<input type="checkbox"/>			19	pescado	alimentos, bebidas y agricultura	fish	food, drink and farming	鱼肉	食物, 饮料与农业	yurou

phpMyAdmin
服务器: localhost ▾ 数据库: vila ▾ 表: verbos

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除

字段	类型	整理	属性	Null	默认	额外	操作
<input type="checkbox"/> Id	int(11)			否		auto_increment	
<input type="checkbox"/> Espanol	text	utf8_unicode_ci		否			
<input type="checkbox"/> English	text	utf8_unicode_ci		否			
<input type="checkbox"/> 中文	text	utf8_unicode_ci		否			
<input type="checkbox"/> Pinyin	text	utf8_unicode_ci		否			

↑ 全选 / 全部不选 选中项:

打印预览 关系查看 规划表结构

添加 1 字段 于表结尾 于表开头 于 Id 之后

索引:

键名	类型	基数	操作	字段
PRIMARY	PRIMARY	2518		Id
在第 1	列创建索引		<input type="button" value="执行"/>	

已使用空间

类型	用法
数据	124,528 字节
索引	30,720 字节
统计	155,248 字节

行统计

语句	值
格式	动态
整理	latin1_swedish_ci
行数	2,518
行长度 <input type="button" value="表"/>	49
行大小 <input type="button" value="表"/>	62 字节
下一个 Autoindex	2,587
创建时间	2009 年 06 月 23 日 12:42
最后更新时间	2009 年 06 月 23 日 11:46

Open new phpMyAdmin window

phpMyAdmin 服务器: localhost 数据库: vila 表: verbos

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除

显示行 0 - 29 (2,518 总计, 查询花费 0.0095 秒)

SQL 查询:
 SELECT
 FROM `verbos`
 LIMIT 0, 30

Profiling [编辑] [解释 SQL] [创建 PHP 代码] [刷新]

显示: 30 行, 开始行数: 30
 以 水平 模式显示, 并且在 100 个单元格后重复标题

主键排序: 无

	←	→	Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>			51	abandonar	to abandon	放弃	fangqi
<input type="checkbox"/>			292	abanicar	to fan	煽动	shandong
<input type="checkbox"/>			294	abaratar	to lower the price of	降价	jiangjia
<input type="checkbox"/>			297	abarcar	to include	包括	baokuo
<input type="checkbox"/>			299	abarrotar	to fill	装满	zhuangman
<input type="checkbox"/>			302	abastecer	to supply	供给	gongji
<input type="checkbox"/>			303	abatir	to demolish	毁坏	huihuai
<input type="checkbox"/>			306	abdicar	to abdicate	退位	tuiwei
<input type="checkbox"/>			307	abismar	to humble	挫	cuo
<input type="checkbox"/>			311	ablandar	to soften	使变柔软	shibianrouuan
<input type="checkbox"/>			312	abochornar	to embarrass	使困窘	shikunjiong
<input type="checkbox"/>			313	abofetear	to slap	拍击	paiji
<input type="checkbox"/>			314	abogar	to advocate	提倡	tichang
<input type="checkbox"/>			315	abollar	to dent	使凹下	shiaoxia
<input type="checkbox"/>			320	abonar	to settle	安放	anfang
<input type="checkbox"/>			322	aborrecer	to hate	憎恨	zenghen
<input type="checkbox"/>			323	abortar	to have a miscarriage	失败	shibai
<input type="checkbox"/>			324	abotonar	to button	扣	kou
<input type="checkbox"/>			325	abrasar	to burn	烧焦	shaojiao
<input type="checkbox"/>			326	abrazar	to embrace	拥抱	yongbao
<input type="checkbox"/>			327	abreviar	to abbreviate	缩短	suoduan
<input type="checkbox"/>			328	abrigar	to shelter	掩蔽	yanbi
<input type="checkbox"/>			329	abrilantar	to polish	擦亮	cailiang
<input type="checkbox"/>			52	abrir	to open	打开	dakai
<input type="checkbox"/>			332	abrochar	to button	扣住	kouzhu
<input type="checkbox"/>			333	abrumar	to overwhelm	战胜	zhansheng
<input type="checkbox"/>			334	absolver	to absolve	免除	mianchu
<input type="checkbox"/>			335	absorber	to absorb	吸收	xishou
<input type="checkbox"/>			336	abstenerse	to abstain	放弃	fangqi
<input type="checkbox"/>			300	abstinarse	to be obstinate	固执	guzhi

全选 / 全部不选 选中项:

显示: 30 行, 开始行数: 30
 以 水平 模式显示, 并且在 100 个单元格后重复标题

Query results operations
 打印预览 打印预览 (全文显示) 导出 CREATE VIEW

将此 SQL 查询加入书签
 标签: 让每个用户都可以访问这个书签

将此 SQL 查询加入书签

Open new phpMyAdmin window

phpMyAdmin
服务器: localhost | 数据库: vila | 表: adjetivos

浏览
结构
SQL
搜索
插入
导出
Import
操作
清空
删除

字段	类型	整理	属性	Null	默认	额外	操作
<input type="checkbox"/> Id	int(11)			否		auto_increment	
<input type="checkbox"/> Espanol	text	utf8_unicode_ci		否			
<input type="checkbox"/> English	text	utf8_unicode_ci		否			
<input type="checkbox"/> 中文	text	utf8_unicode_ci		否			
<input type="checkbox"/> Pinyin	text	utf8_unicode_ci		否			

↑ 全选 / 全部不选 选中项:

打印预览
关系查看
规划表结构

添加 1 字段
 于表结尾
 于表开头
 于 Id
 之后
执行

索引:

键名	类型	基数	操作	字段
PRIMARY	PRIMARY	186		Id
在第 1	列创建索引		执行	

已使用空间

类型	用法
数据	8,500 字节
索引	4,096 字节
统计	12,596 字节

行统计

语句	值
格式	动态
整理	utf8_unicode_ci
行数	186
行长度 \emptyset	45
行大小 \emptyset	68 字节
下一个 Autoindex	187
创建时间	2009 年 06 月 23 日 12:41
最后更新时间	2009 年 06 月 23 日 12:05

Open new phpMyAdmin window

服务器: localhost | 数据库: vila | 表: adjetivos

浏览
结构
SQL
搜索
插入
导出
Import
操作
清空
删除

显示行 0 - 29 (186 总计, 查询花费 0.0004 秒)

数据库

vila (7)

- adjetivos
- adverbios
- lugar
- numerales
- sustantivos
- tiempo
- verbos

SQL 查询:

```
SELECT
FROM `adjetivos`
LIMIT 0, 30
```

Profiling [编辑] [解释 SQL] [创建 PHP 代码] [刷新]

显示: 30 行, 开始行数: 30

以 水平 模式显示, 并且在 100 个单元格后重复标题

页号: 1

主键排序: 无						
		Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>		1	absoluto	absolute	绝对的	jueduide
<input type="checkbox"/>		2	actual	present	现在的	xianzaide
<input type="checkbox"/>		3	amplio	ample	宽敞的	kuanchangde
<input type="checkbox"/>		4	anterior	previous	在前的	zaiqiande
<input type="checkbox"/>		5	bajo	low	低的	dide
<input type="checkbox"/>		6	bastante	enough	足够的	zugoude
<input type="checkbox"/>		7	bueno	good;well	好的	haode
<input type="checkbox"/>		8	capaz	able	有能力的	younenglide
<input type="checkbox"/>		9	central	central	中心的	zhongxinde
<input type="checkbox"/>		10	científico	scientific	科学的	kexuede
<input type="checkbox"/>		11	civil	civil	城市的	chengshide
<input type="checkbox"/>		12	claro	clearly	明亮的	mingliangde
<input type="checkbox"/>		13	completo	complete	完整的	wanzhengde
<input type="checkbox"/>		14	común	common	公共的	gonggongde
<input type="checkbox"/>		15	conjunto	joint	连接的	lianjiade
<input type="checkbox"/>		16	constante	constant	持久的	chijiude
<input type="checkbox"/>		17	contrario	opposite	相反的	xiangfande
<input type="checkbox"/>		18	corriente	current	最新的	zuixinde
<input type="checkbox"/>		19	cultural	cultural	文化的	wenhuede
<input type="checkbox"/>		20	definitivo	definitive	决定性的	juedingxingde
<input type="checkbox"/>		21	demás	rest of	其余的	qiyude
<input type="checkbox"/>		22	demasiado	too much	过分的	guofende
<input type="checkbox"/>		23	diferente	different	不同的	butongde
<input type="checkbox"/>		24	difícil	difficult	困难的	kunnande
<input type="checkbox"/>		25	directo	direct	真的	zhide
<input type="checkbox"/>		26	distinto	different	不同的	butongde
<input type="checkbox"/>		27	butongde	diverse	各种各样的	gezhonggeyangde
<input type="checkbox"/>		28	doble	double	双倍的	shuangbeide
<input type="checkbox"/>		29	económico	economic	经济的	jingjide
<input type="checkbox"/>		30	enfermo	ill	有病的	youbingde

全选 / 全部不选 选中项:

显示: 30 行, 开始行数: 30

以 水平 模式显示, 并且在 100 个单元格后重复标题

页号: 1

Query results operations

打印预览 打印预览 (全文显示) 导出 CREATE VIEW

将此 SQL 查询加入书签

标签: 让每个用户都可以访问这个书签

Open new phpMyAdmin window

数据库

vila (7)

- adjetivos
- adverbios
- lugar
- numerales
- sustantivos
- tiempo
- verbos

服务器: localhost | 数据库: vila | 表: adverbios

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除

字段	类型	整理	属性	Null	默认	额外	操作
<input type="checkbox"/> Id	int(11)			否		auto_increment	
<input type="checkbox"/> Espanol	text	utf8_unicode_ci		否			
<input type="checkbox"/> English	text	utf8_unicode_ci		否			
<input type="checkbox"/> 中文	text	utf8_unicode_ci		否			
<input type="checkbox"/> Pinyin	text	utf8_unicode_ci		否			

打印预览 关系查看 规划表结构

添加 1 字段 于表结尾 于表开头 于 Id 之后

索引: Id				已使用空间		行统计	
键名	类型	基数	操作	类型	用法	语句	值
PRIMARY	PRIMARY	209		数据	10,356 字节	格式	动态
在第 1		列创建索引	<input type="button" value="执行"/>	索引	5,120 字节	整理	utf8_unicode_ci
				统计	15,476 字节	行数	209
						行长度	49
						行大小	74 字节
						下一个 Autoindex	211
						创建时间	2009 年 06 月 23 日 12:41
						最后更新时间	2009 年 06 月 23 日 11:46

Open new phpMyAdmin window

phpMyAdmin
服务器: localhost ▾ 数据库: vila ▾ 表: adverbios

浏览
结构
SQL
搜索
插入
导出
Import
操作
清空
删除

显示行 0 - 29 (209 总计, 查询花费 0.0003 秒)

SQL 查询:

```
SELECT
FROM `adverbios`
LIMIT 0, 30
```

Profiling [编辑] [解释 SQL] [创建 PHP 代码] [刷新]

vila (7)

- adjetivos
- adverbios
- lugar
- numerales
- sustantivos
- tiempo
- verbos

显示: 30 行, 开始行数: 30
> >>
页号: 1

以 水平 模式显示, 并且在 100 个单元格后重复标题
> >>
页号: 1

主键排序: 无

			Espanol	English	中文	Pinyin
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1 abajo	below	在下	zaixia
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2 en absoluto	not at all	一点也不	yidianyebu
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3 acá	here	在这里	zaizheli
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4 acaso	perhaps	或许	huoxu
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5 actualmente	at present	现在	xianzai
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6 adelante	forward	向前地	xiangqiandi
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7 además	besides	此外	ciwai
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8 adentro	inside	在里面	zailimian
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9 adónde	where	什么地方	shenmedifang
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	10 adrede	on purpose	故意	guyi
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	11 afuera	outside	在外面	zaiwaimian
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	12 ahí	there	在那里	zainali
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	13 ahora	now	现在	xianzai
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	14 ahorita	right now	立刻	like
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	15 algo	somewhat	稍微	shaowei
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	16 alias	alias	别名为	biemingwei
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	17 allá	there	在那里	zainali
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	18 allí	there	在那里	zainali
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	19 alrededor	around	周围	zhouwei
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	20 anoche	lastnight	昨晚	zuowan
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	21 antaño	long ago	从前	congqian
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	22 anteanoche	the night before last	前天晚上	qiantianwanshang
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	23 anteayer	the day before yesterday	前天	qiantian
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	24 de antemano	beforehand	预先	yuxian
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	25 antes	before	先前	xianqian
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	26 antiguamente	formerly	从前	congqian
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	27 aposta	deliberately	故意地	guyidi
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	28 aprisa	hurriedly	仓促地	cangcudi
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	29 aquí	here	在这里	zaizheli
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	30 arriba	above	在上面	zaishangmian

全选 / 全部不选 选中项:
> >>
页号: 1

Query results operations

打印预览
 打印预览 (全文显示)
 导出
 CREATE VIEW

将此 SQL 查询加入书签

标签: 让每个用户都可以访问这个书签

Open new phpMyAdmin window



服务器: localhost ▾ 数据库: vila ▾ 表: numerales

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除



数据库

vila (7)

vila (7)

- adjetivos
- adverbios
- lugar
- numerales
- sustantivos
- tiempo
- verbos

字段	类型	整理	属性	Null	默认	额外	操作							
<input type="checkbox"/> Id	int(11)			否		auto_increment								
<input type="checkbox"/> Espanol	text	utf8_unicode_ci		否										
<input type="checkbox"/> English	text	utf8_unicode_ci		否										
<input type="checkbox"/> 中文	text	utf8_unicode_ci		否										
<input type="checkbox"/> Pinyin	text	utf8_unicode_ci		否										

全选 / 全部不选 选中项:

打印预览 关系查看 规划表结构

添加 1 字段 于表结尾 于表开头 于 Id 之后 执行

索引: 1				已使用空间		行统计	
键名	类型	基数	操作	类型	用法	语句	值
PRIMARY	PRIMARY	30		数据	1,300 字节	格式	动态
				索引	2,048 字节	整理	utf8_unicode_ci
在第 1	列创建索引		执行	统计	3,348 字节	行数	30
				行长度		行大小	43
				下一个 Autoindex			112 字节
				创建时间			31
				最后更新时间			2009 年 06 月 23 日 12:44
							2010 年 03 月 06 日 17:10

Open new phpMyAdmin window



服务器: localhost ▾ 数据库: vila ▾ 表: numerales



数据库
vila (7)

显示行 0 - 29 (30 总计, 查询花费 0.0003 秒)

- adjetivos
- adverbios
- lugar
- numerales
- sustantivos
- tiempo
- verbos

```
SQL 查询:
SELECT *
FROM `numerales`
LIMIT 0, 30
```

Profiling [编辑] [解释 SQL] [创建 PHP 代码] [刷新]

显示: 30 行, 开始行数: 0
以 水平 模式显示, 并且在 100 个单元格后重复标题

主键排序: 无

		Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>		1	uno	one	一	yi
<input type="checkbox"/>		2	dos	two	二	er
<input type="checkbox"/>		3	tres	three	三	san
<input type="checkbox"/>		4	cuadro	four	四	si
<input type="checkbox"/>		5	cinco	five	五	wu
<input type="checkbox"/>		6	seis	six	六	liu
<input type="checkbox"/>		7	siete	seven	七	qi
<input type="checkbox"/>		8	ocho	eight	八	ba
<input type="checkbox"/>		9	nueve	nine	九	ju
<input type="checkbox"/>		10	diez	ten	十	shi
<input type="checkbox"/>		11	once	eleven	十一	shiyi
<input type="checkbox"/>		12	doce	twelve	十二	shier
<input type="checkbox"/>		13	trece	thirteen	十三	shisan
<input type="checkbox"/>		14	catorce	fourteen	十四	shisi
<input type="checkbox"/>		15	quince	fifteen	十五	shiwu
<input type="checkbox"/>		16	dieciseis	sixteen	十六	shiliu
<input type="checkbox"/>		17	diecisiete	seventeen	十七	shiqi
<input type="checkbox"/>		18	dieciocho	eighteen	十八	shiba
<input type="checkbox"/>		19	diecinueve	nineteen	十九	shijiu
<input type="checkbox"/>		20	veinte	twenty	二十	ershi
<input type="checkbox"/>		21	treinta	thirty	三十	sanshi
<input type="checkbox"/>		22	cuarenta	forty	四十	sishi
<input type="checkbox"/>		23	cincuenta	fifty	五十	wushi
<input type="checkbox"/>		24	sesenta	sixty	六十	liushi
<input type="checkbox"/>		25	setenta	seventy	七十	qishi
<input type="checkbox"/>		26	ochenta	eighty	八十	bashi
<input type="checkbox"/>		27	noventa	ninety	九十	jiushi
<input type="checkbox"/>		28	ciento	hundred	百	bai
<input type="checkbox"/>		29	mil	thousand	千	qian
<input type="checkbox"/>		30	millon	million	百万	baiwan

全选 / 全部不选 选中项:

显示: 30 行, 开始行数: 0
以 水平 模式显示, 并且在 100 个单元格后重复标题

Query results operations
 打印预览 打印预览 (全文显示) 导出 CREATE VIEW

将此 SQL 查询加入书签

标签: _____ 让每个用户都可以访问这个书签

Open new phpMyAdmin window

phpMyAdmin

服务器: localhost | 数据库: vila | 表: tiempo

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除

字段	类型	整理	属性	Null	默认	额外	操作
<input type="checkbox"/> Id	int(11)			否		auto_increment	
<input type="checkbox"/> Espanol	text	utf8_unicode_ci		否			
<input type="checkbox"/> English	text	utf8_unicode_ci		否			
<input type="checkbox"/> 中文	text	utf8_unicode_ci		否			
<input type="checkbox"/> Pinyin	text	utf8_unicode_ci		否			

全选 / 全部不选 选中项:

打印预览 关系查看 规划表结构

添加 1 字段 于表结尾 于表开头 于 Id 之后

索引: Id				已使用空间		行统计	
键名	类型	基数	操作	类型	用法	语句	值
PRIMARY	PRIMARY	71		数据	3,192 字节	格式	动态
在第 1	列创建索引	<input type="button" value="执行"/>		索引	2,048 字节	整理	utf8_unicode_ci
				统计	5,240 字节	行数	71
						行长度 <input type="text" value="e"/>	44
						行大小 <input type="text" value="e"/>	74 字节
						下一个	78
						Autoindex	
						创建时间	2010 年 03 月 05 日 12:27
						最后更新时间	2010 年 03 月 05 日 12:27

Open new phpMyAdmin window

phpMyAdmin 服务器: localhost 数据库: vila 表: tiempo

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除

显示行 0 - 29 (71 总计, 查询花费 0.0003 秒)

SQL 查询:

```
SELECT
FROM `tiempo`
LIMIT 0, 30
```

Profiling [编辑] [解释 SQL] [创建 PHP 代码] [刷新]

显示: 30 行, 开始行数: 30
 以 水平 模式显示, 并且在 100 个单元格后重复标题 页号: 1

主键排序: 无

		Id	Espanol	English	中文	Pinyin	
<input type="checkbox"/>			1	milenio	millennium	千年	qiannian
<input type="checkbox"/>			2	siglo	century	世纪	shiji
<input type="checkbox"/>			3	década	decade	十年	shinian
<input type="checkbox"/>			4	año	year	年	nian
<input type="checkbox"/>			5	mes	month	月	yue
<input type="checkbox"/>			6	día	day	日	ri
<input type="checkbox"/>			7	semana	week	星期	xingqi
<input type="checkbox"/>			8	fecha	date	日期	riqi
<input type="checkbox"/>			10	horario	schedule	日程表	richengbiao
<input type="checkbox"/>			11	almanaque	calendar	日历	rili
<input type="checkbox"/>			13	lunes	Monday	星期一	xingqiyi
<input type="checkbox"/>			14	martes	Tuesday	星期二	xingqier
<input type="checkbox"/>			15	miércoles	Wednesday	星期三	xingqisan
<input type="checkbox"/>			16	jueves	Thursday	星期四	xingqisi
<input type="checkbox"/>			17	viernes	Friday	星期五	xingqiwu
<input type="checkbox"/>			18	sábado	Saturday	星期六	xingqiliu
<input type="checkbox"/>			19	domingo	Sunday	星期日	xingqiri
<input type="checkbox"/>			20	ayer	yesterday	昨天	zuotian
<input type="checkbox"/>			21	hoy	today	今天	jintian
<input type="checkbox"/>			22	mañana	tomorrow	明天	mingtian
<input type="checkbox"/>			23	enero	January	一月	yi yue
<input type="checkbox"/>			24	febrero	February	二月	er yue
<input type="checkbox"/>			25	marzo	march	三月	sanyue
<input type="checkbox"/>			26	abril	april	四月	siyue
<input type="checkbox"/>			27	mayo	May	五月	wuyue
<input type="checkbox"/>			28	junio	June	六月	liuyue
<input type="checkbox"/>			29	julio	July	七月	qiyue
<input type="checkbox"/>			30	agosto	August	八月	bayue
<input type="checkbox"/>			31	septiembre	September	九月	jiuyue
<input type="checkbox"/>			32	octubre	October	十月	shiyue

全选 / 全部不选 选中项: | |

显示: 30 行, 开始行数: 30
 以 水平 模式显示, 并且在 100 个单元格后重复标题 页号: 1

Query results operations
 打印预览 打印预览 (全文显示) 导出 CREATE VIEW

将此 SQL 查询加入书签
 标签: 让每个用户都可以访问这个书签
 将此 SQL 查询加入书签

Open new phpMyAdmin window

phpMyAdmin

服务器: localhost | 数据库: vila | 表: lugar

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除

字段	类型	整理	属性	Null	默认	额外	操作
<input type="checkbox"/> Id	int(11)			否		auto_increment	
<input type="checkbox"/> Espanol	text	utf8_unicode_ci		否			
<input type="checkbox"/> English	text	utf8_unicode_ci		否			
<input type="checkbox"/> 中文	text	utf8_unicode_ci		否			
<input type="checkbox"/> Pinyin	text	utf8_unicode_ci		否			

全选 / 全部不选 选中项:

打印预览 关系查看 规划表结构

添加 1 字段 于表结尾 于表开头 于 Id 之后

索引: <input checked="" type="checkbox"/>				已使用空间		行统计	
键名	类型	基数	操作	类型	用法	语句	值
PRIMARY	PRIMARY	248		数据	13,220 字节	格式	动态
在第 1		列创建索引	<input type="button" value="执行"/>	索引	5,120 字节	整理	utf8_unicode_ci
				统计	18,340 字节	行数	248
						行长度 <input type="checkbox"/>	53
						行大小 <input type="checkbox"/>	74 字节
						下一个 Autoindex	250
						创建时间	2010 年 03 月 05 日 12:26
						最后更新时间	2010 年 03 月 05 日 12:31

Open new phpMyAdmin window

phpMyAdmin 服务器: localhost 数据库: vila 表: lugar

浏览 结构 SQL 搜索 插入 导出 Import 操作 清空 删除

显示行 0 - 29 (248 总计, 查询花费 0.0003 秒)

SQL 查询:

```
SELECT
FROM `lugar`
LIMIT 0, 30
```

Profiling [编辑] [解释 SQL] [创建 PHP 代码] [刷新]

显示: 30 行, 开始行数: 30
 以 水平 模式显示, 并且在 100 个单元格后重复标题

页号: 1

主键排序: 无

	←	→	Id	Espanol	English	中文	Pinyin
<input type="checkbox"/>			1	habitación	room	房间	fangjian
<input type="checkbox"/>			2	piso	flat	套间	taojian
<input type="checkbox"/>			3	casa	house	房屋	fangwu
<input type="checkbox"/>			4	chalet	villa	别墅	bieshu
<input type="checkbox"/>			5	edificio	building	建筑	jianzhu
<input type="checkbox"/>			6	calle	street	街	jie
<input type="checkbox"/>			7	barrios	neighborhood	区	qu
<input type="checkbox"/>			8	pueblo	village	村镇	cunzhen
<input type="checkbox"/>			9	ciudad	city	城市	chengshi
<input type="checkbox"/>			10	provincia	province	省	sheng
<input type="checkbox"/>			11	capital	capital	首都	shoudu
<input type="checkbox"/>			12	pais	country	国家	guojia
<input type="checkbox"/>			13	continente	continent	洲	zhou
<input type="checkbox"/>			14	Asia	asia	亚洲	yazhou
<input type="checkbox"/>			15	Europa	europa	欧洲	ouzhou
<input type="checkbox"/>			16	África	africa	非洲	feizhou
<input type="checkbox"/>			17	norteamericano	northAmerican	北美洲	beimeizhou
<input type="checkbox"/>			18	sudamericano	southAmerican	南美洲	nanmeizhou
<input type="checkbox"/>			19	Oceanía	oceania	大洋洲	dayangzhou
<input type="checkbox"/>			20	Antártida	antarctica	南极洲	nanjizhou
<input type="checkbox"/>			21	tierra	earth	地球	diqu
<input type="checkbox"/>			22	Vía Láctea	galaxy	银河系	yinhexi
<input type="checkbox"/>			23	universo	universe	宇宙	yuzhou
<input type="checkbox"/>			24	océano	ocean	大洋	dayang
<input type="checkbox"/>			25	isla	island	岛屿	daoyu
<input type="checkbox"/>			26	península	peninsula	半岛	bandao
<input type="checkbox"/>			27	América del Norte y Central	North America	北美洲	beimeizhou
<input type="checkbox"/>			28	Alaska	Alaska	阿拉斯加	alasia
<input type="checkbox"/>			29	Canadá	Canada	加拿大	jianada
<input type="checkbox"/>			30	Groenlandia	Greenland	格林兰	gelinglan

全选 / 全部不选 选中项:

显示: 30 行, 开始行数: 30
 以 水平 模式显示, 并且在 100 个单元格后重复标题

页号: 1

Query results operations
 打印预览 打印预览 (全文显示) 导出 CREATE VIEW

将此 SQL 查询加入书签
 标签: 让每个用户都可以访问这个书签

将此 SQL 查询加入书签

Open new phpMyAdmin window

phpMyAdmin 服务器: localhost 数据库: vegazana

结构 SQL 搜索 查询 导出 Import Designer 操作 权限 删除

数据库 vegazana (52)

vegazana (52)

- accionado
- accionador
- adjetivo
- adjetivos
- adverbio
- adverbios
- animales
- atributoscomparacion
- atributovaloracionabsoluta
- colores
- concepto
- conectorgenerico
- cuantificadoraccionado
- cuantificadoraccionador
- cuantificadoradjetivo
- cuantificadoradverbio
- cuantificadorconcepto
- cuantificadordeadjetivos
- cuantificadordeconceptos
- cuantificadordestinatario
- cuantificadorpredicado
- cuantificadorsujeto
- cuantificadorverboaccion
- cuantificadorverbosinaccion
- deportes
- destinatario
- espaciogeografico
- idgenerico
- idiomas
- lugarespaciounico
- nombrelugar
- nombretiempo
- objetos
- operadores
- otrostiempo
- países
- palabras
- plantas
- posicionlugar
- predicado
- profesion
- sujeto
- tamaño
- tiempo
- users
- valoracionabsoluta
- valorcomparacion
- valorvaloracionabsoluta
- verboaccion
- verbos
- verbosinaccion
- verbosinacción

表	操作	记录数	类型	整理	大小	多余
accionado		2,288	MyISAM	utf8_unicode_ci	129.4 KB	-
accionador		363	MyISAM	utf8_unicode_ci	22.5 KB	-
adjetivo		147	MyISAM	utf8_unicode_ci	10.8 KB	-
adjetivos		144	MyISAM	utf8_unicode_ci	10.7 KB	-
adverbio		207	MyISAM	utf8_unicode_ci	15.0 KB	-
adverbios		209	MyISAM	utf8_unicode_ci	17.7 KB	-
animales		249	MyISAM	latin1_swedish_ci	15.8 KB	-
atributoscomparacion		29	MyISAM	utf8_unicode_ci	3.3 KB	-
atributovaloracionabsoluta		0	MyISAM	utf8_unicode_ci	1.0 KB	-
colores		16	MyISAM	latin1_swedish_ci	2.6 KB	-
concepto		250	MyISAM	utf8_unicode_ci	15.8 KB	-
conectorgenerico		13	MyISAM	utf8_unicode_ci	2.9 KB	-
cuantificadoraccionado		15	MyISAM	utf8_unicode_ci	2.6 KB	-
cuantificadoraccionador		15	MyISAM	utf8_unicode_ci	2.6 KB	-
cuantificadoradjetivo		18	MyISAM	utf8_unicode_ci	2.7 KB	20 字节
cuantificadoradverbio		15	MyISAM	utf8_unicode_ci	2.6 KB	-
cuantificadorconcepto		22	MyISAM	utf8_unicode_ci	2.9 KB	-
cuantificadordeadjetivos		15	MyISAM	utf8_unicode_ci	2.6 KB	-
cuantificadordeconceptos		21	MyISAM	utf8_unicode_ci	2.9 KB	-
cuantificadordestinatario		21	MyISAM	utf8_unicode_ci	2.9 KB	-
cuantificadorpredicado		21	MyISAM	utf8_unicode_ci	2.9 KB	-
cuantificadorsujeto		21	MyISAM	utf8_unicode_ci	2.9 KB	-
cuantificadorverboaccion		15	MyISAM	utf8_unicode_ci	2.6 KB	-
cuantificadorverbosinaccion		15	MyISAM	utf8_unicode_ci	2.6 KB	-
deportes		136	MyISAM	latin1_swedish_ci	11.7 KB	-
destinatario		2,288	MyISAM	utf8_unicode_ci	129.4 KB	-
espaciogeografico		26	MyISAM	latin1_swedish_ci	3.1 KB	-
idgenerico		4,733	MyISAM	utf8_unicode_ci	268.2 KB	-
idiomas		51	MyISAM	utf8_unicode_ci	4.8 KB	-
lugarespaciounico		248	MyISAM	utf8_unicode_ci	17.9 KB	-
nombrelugar		246	MyISAM	utf8_unicode_ci	17.8 KB	-
nombretiempo		71	MyISAM	utf8_unicode_ci	5.1 KB	-
objetos		1,735	MyISAM	latin1_swedish_ci	153.2 KB	20 字节
operadores		15	MyISAM	utf8_unicode_ci	2.7 KB	-
otrostiempo		10	MyISAM	utf8_unicode_ci	2.4 KB	-
países		220	MyISAM	latin1_swedish_ci	17.0 KB	-
palabras		76	MyISAM	utf8_unicode_ci	5.1 KB	-
plantas		192	MyISAM	latin1_swedish_ci	12.6 KB	-
posicionlugar		21	MyISAM	utf8_unicode_ci	3.0 KB	-
predicado		2,288	MyISAM	utf8_unicode_ci	129.4 KB	-
profesion		114	MyISAM	latin1_swedish_ci	9.7 KB	-
sujeto		2,288	MyISAM	utf8_unicode_ci	129.4 KB	-
tamaño		28	MyISAM	latin1_swedish_ci	3.0 KB	-
tiempo		71	MyISAM	utf8_unicode_ci	5.1 KB	-