CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO

cenidet

TRADUCTOR DE LENGUAJE NATURAL ESPAÑOL A SQL PARA UN SISTEMA DE CONSULTAS A BASES DE DATOS

T E S I S
QUE PARA OBTENER EL GRADO DE:
DOCTOR EN CIENCIAS EN CIENCIAS
DE LA COMPUTACIÓN

PRESENTA:

JUAN JAVIER GONZÁLEZ BARBOSA

DIRECTOR DE TESIS:
DR. RODOLFO A. PAZOS RANGEL

CODIRECTORES:
DR. ALEXANDER GELBUCKH
DR. JOAQUÍN PÉREZ ORTEGA

Índice de Contenido

Lista de Tablas

Lista de Figuras

RESUMEN

Este trabajo presenta el desarrollo de un modulo traductor para una Interfaz de Lenguaje Natural a Bases de Datos (ILNBDs) que permite traducir consultas en lenguaje natural a SQL. La técnica utilizada por el traductor es portable a cualquier dominio, ya que se mantiene independiente de la información contenida en la base de datos, Además, la creación automática de un diccionario de dominio utilizando los metadatos de la base de datos, evita las tediosas configuraciones que tendría que hacer el usuario para configurar la interfaz a un dominio dado. La consulta del usuario recibe un preprocesamiento en donde cada palabra es etiquetada con información morfosintactica y semántica, la primera es obtenida de un etiquetador gramatical y la segunda de la información proporcionada por los metadatos. La consulta es representada por un grafo y traducida a SQL tomando en cuenta la información obtenida por los diferentes procesos. A diferencia de la mayoría de las ILNBDs existentes, la técnica de traducción utiliza las operaciones de unión e intersección de la teoría de conjuntos para dar tratamiento a la preposición "de" y a la conjunción "y", las cuales forman parte de las palabras invariables de la oración en el idioma español. Los resultados experimentales realizados con dos bases de datos de dominios diferentes muestran que el tratamiento de palabras invariables mejora la precisión de la traducción al responder sin columnas extras la consulta del usuario. Con la ILNBD desarrollada, el 86% de las consultas fueron traducidas correctamente. Además, el proceso de configurar la interfaz de un dominio a otro tomó solamente diez minutos

1 INTRODUCCIÓN

Desde el nacimiento de la computadora electrónica, la posibilidad de comunicarse con ésta a través de lenguaje escrito o hablado ha sido el sueño del hombre. En la actualidad se han logrado grandes avances en este campo. La mayoría de las aplicaciones emplean técnicas que se utilizan en el procesamiento de lenguaje natural, las cuales proporcionan a las computadoras la posibilidad de entender el texto o el habla, utilizando diferentes métodos para lograr una comprensión del lenguaje. Para alcanzar esto es necesario diseñar un sistema de traducción, aplicar los métodos necesarios para comprender una oración, y que el sistema sea capaz de traducirla.

Actualmente grandes compañías están invirtiendo en el desarrollo de interfaces de lenguaje natural.

En julio de 2000 Oracle Corporation se asoció con Camelot Ventures, con la finalidad de financiar el proyecto Native Minds, que es el proveedor de soluciones líder en la creación de servicios para lenguaje natural. Native Minds ha desarrollado una suite de productos y servicios para crear representantes virtuales para el área de ventas, servicio y soporte, llamados vReps. Éstos atienden y dan

soporte a los usuarios en línea de la misma forma en que lo haría un representante de ventas, utilizando una conversación personal, rápida e informativa, todo esto en inglés simple; con lo cual se reduce el costo de soporte real de 24 horas los 7 días de la semana, a una fracción del costo de una llamada telefónica [1].

Bill Gates hizo el siguiente comentario: "El lenguaje natural, ya sea escrito en oraciones o en el reconocimiento del habla, está destinado a jugar un papel central en los sistemas operativos y en las aplicaciones del futuro. Podemos hacer que las computadoras sean más poderosas y suficientemente simples como para responder a comandos claros en inglés" [2]. Actualmente Microsoft incluye una interfaz de lenguaje natural llamada Microsoft English Query (MSEQ) en su software SQL Server.

La compañía InQuira desarrolló un software que permite a los clientes escribir en el teclado preguntas explícitas en lenguaje natural en los cibersitios de empresas, y recibir la información que necesitan, ya sea una lista de automóviles para venta de un precio determinado máximo, o un tema más técnico acerca de si dos programas son compatibles. La compañía con sede en California ya tiene un puñado de clientes que han adoptado el servicio, entre ellos Bank of America Corp., Sun Microsystems Inc. y BEA Systems Inc [3].

A la fecha el desarrollo con relación al procesamiento de lenguaje natural en español está enfocado en su mayor parte a aplicaciones diferentes a las de bases de datos, como por ejemplo la interpretación de textos [4], correctores ortográficos [5] y herramientas para el procesamiento del español [6, 7]. Algunos de los proyectos más importantes relacionados con interfaces de lenguaje natural a bases de datos en español los describen Pazos, Gelbukh y Zárate en [8].

El español es la tercera lengua por número de habitantes. Es idioma oficial para 332 millones de personas y la utilizan como lengua no oficial 23 millones más [9].

México es el país hispanohablante más poblado: existen 98 millones de personas en México y 35 millones en EUA [10].

Este trabajo de investigación se enfocará en el módulo de traducción de una interfaz de lenguaje natural en español para acceder a bases de datos. Se realizará un análisis de las técnicas de traducción empleadas por las Interfaces de Lenguaje Natural para Bases de Datos (ILNBDs) existentes, y se propondrá una técnica de traducción que mejore el desempeño o porcentaje de consultas realizadas correctamente.

1.1 MOTIVACIONES

De la revisión documental realizada, así como de las pruebas efectuadas a las ILNBDs disponibles en Internet junto con las pruebas hechas por otros investigadores, se observa que **ninguna de las ILNBDs fue diseñada para ser portada fácilmente a bases de datos diferentes de aquélla para la cual fue diseñada originalmente, ni el diccionario de datos y la base de conocimiento de la ILNBD fueron diseñados para reuso y compartimiento, que son los objetivos que se persiguen en este proyecto.** Por lo tanto es factible mejorar el módulo de traducción de consultas con la finalidad de lograr la portabilidad del dominio en una ILNBD.

1.2 DESCRIPCIÓN DEL PROBLEMA

Hasta la fecha, las interfaces de lenguaje natural para bases de datos no garantizan la traducción satisfactoria de la consulta en lenguaje natural o la obtención de una respuesta confiable [11]. Además, es importante mencionar que en lo que respecta a la independencia del dominio todavía hay mucho trabajo que realizar, ya que el proceso de configuración a diferentes dominios y bases de

datos aún es una meta lejana, debido a que depende de los diccionarios, los mecanismos de traducción y el análisis de sinónimos, antónimos y preposiciones entre otras cosas.
En todos los trabajos revisados se observa que la portabilidad del dominio es un problema que está lejos de ser resuelto, esto se debe principalmente a los siguientes factores:

- El lenguaje natural siempre será más expresivo que un lenguaje formal como SQL. Además, para realizar una consulta en lenguaje natural a una base de datos es necesario conocer si la consulta es factible de ser contestada.

- La mayoría de las interfaces existentes basan su funcionamiento en la búsqueda de palabras clave dentro de la oración de entrada que están relacionadas con una acción a realizar. La desventaja principal de utilizar palabras clave, es que estas se deben crear y configurar para cada dominio por un administrador de la base de datos o por alguien que conozca la estructura de la base de datos.

Es importante señalar que al sólo evaluar las palabras clave que aparecen dentro de la consulta, se puede pasar por alto información importante que sea necesaria para una mejor interpretación de la consulta.

Es por eso que se propone lo siguiente: dada una estructura de datos que contiene una oración en español, junto con las categorías sintácticas de cada palabra de la oración e información adicional, **traducir la oración en español a una expresión en SQL** semánticamente equivalente.

1.3 OBJETIVO DE LA TESIS

Desarrollar un módulo de traducción para una ILNBD en español, que mejore el desempeño o porcentaje de consultas contestadas correctamente con respecto a las técnicas hasta ahora utilizadas, y que permita al módulo de traducción ser independiente de la base de datos.

1.4 HIPÓTESIS

H1. Se puede implementar un traductor de lenguaje natural en español que tenga mayor portabilidad de dominio que los hasta ahora conocidos.

H2. Se puede implementar un traductor que responda un mayor porcentaje de consultas.

1.5 BENEFICIOS Y JUSTIFICACIONES

Es importante proponer el desarrollo de metodologías para crear una ILNBD que ofrezca independencia del dominio de la base de datos debido a las siguientes razones:

- La necesidad de ILNBDs se ha venido incrementando a medida de que más usuarios inexpertos requieren acceder a la información que se encuentra en bases de datos a través de sus computadoras, PDAs y teléfonos celulares.
- Las ILNBDs sólo se consideran útiles si la traducción de la consulta en lenguaje natural se realiza correctamente [11].
- Los mecanismos de traducción de las consultas en lenguaje natural a lenguaje formal y los diccionarios utilizados por las interfaces de lenguaje natural son altamente dependientes del dominio, además de que éstos son configurados de manera manual [11].

- Hasta la fecha, las ILNBDs sólo responden el 80% de las consultas realizadas por los usuarios [11], y son altamente dependientes del dominio y la base de datos.
- Internet contiene una gran cantidad de información que carece de una estructura uniforme que dificulta su acceso, por lo que una solución sería convertir a la Web en una base de datos virtual, y acceder a ésta a través de lenguaje natural [12].
- El uso de interfaces de lenguaje natural puede ser aplicado en interfaces para aplicaciones electrodomésticas como televisores, microondas, teléfonos, VCR, etc. [13].

1.6 APORTACIONES

- Mayor cobertura de consultas que pueden ser traducidas.
- La interfaz será de fácil configuración.
- Las oraciones de entrada serán en el idioma Español.

1.7 ALCANCES DE LA INVESTIGACIÓN

El producto esperado de este trabajo de investigación consiste en los siguientes elementos:

a) Un módulo computacional que tenga como entrada un archivo con los componentes de la oración de entrada etiquetados. La salida del módulo es la consulta en SQL.

b) Un diccionario de sinónimos general.

c) Un módulo de software para obtener los metadatos de la base de datos.

d) Un diccionario de dominio.

Los límites de este trabajo son los siguientes:

- No habrá diálogo con el usuario.
- El sistema no recordará frases anteriores.
- El tipo de consultas a manejar tendrá las siguientes características:
  - Con columnas y tablas explícitas.
  - Con columnas implícitas y tablas explícitas.
  - Con columnas explícitas y tablas implícitas.
  - Con columnas y tablas implícitas.
  - Consultas con condición.
- Debido a que en SQL una consulta se puede expresar de diferentes maneras, no se considerará el problema de transformar una consulta a su equivalente óptimo.
- Solamente se traducirán consultas expresadas en español.
- No se dará información que no esté explícita en la BD, ya que tendría que ver con Bases de Datos Deductivas.
- No se tratarán consultas temporales.

2 TÉCNICAS DE TRADUCCIÓN

En este capítulo se describen las técnicas de traducción de consultas utilizadas por otras ILNBD. Además, se realiza una revisión de los trabajos relacionados con esta investigación, mostrando las técnicas de traducción y los métodos de evaluación utilizados por éstos.

2.1 TÉCNICAS DE TRADUCCIÓN UTILIZADAS POR LAS ILNBDS

Se pueden distinguir dos tipos de técnicas para la traducción de consultas en lenguaje natural:

- *Técnicas basadas en la sintaxis*: la consulta es analizada para construir su estructura sintáctica y después es traducida a un lenguaje de consultas a bases de datos. Ejemplos de este enfoque son los sistemas de reconocimiento de patrones y los sistemas basados en la sintaxis.

En los sistemas basados en patrones, se utilizan palabras claves definidas en cierta secuencia para la construcción de éstos, y a cada patrón se le asigna una

acción lógica con la cual responder. La principal ventaja del enfoque de reconocimiento de patrones es su simplicidad: no se necesitan módulos de interpretación ni de análisis sintáctico, y los sistemas son fáciles de implementar. Sin embargo, esta técnica resulta ser sumamente dependiente del dominio [7, 14].

La técnica de traducción basada en la sintaxis analiza sintácticamente la consulta del usuario generando un árbol sintáctico y usan gramáticas que asocian el árbol con consultas a bases de datos [7]. Sin embargo, el mapeo de reglas resulta complejo y tedioso de manipular, disminuyendo la portabilidad del lenguaje y del dominio [15].

- *Técnicas orientadas a la semántica*: la consulta se convierte en predicados de un lenguaje de consultas a bases de datos, sin construir su estructura sintáctica. Los sistemas gramática-semántica, sistemas de lenguaje de representación intermedia y los patrones léxicos semánticos son técnicas orientadas a la semántica.

Los sistemas de gramática semántica obtienen un árbol de la consulta de usuario, y a partir de éste se construye la consulta en un lenguaje de bases de datos, pero a diferencia de la técnica basada en la sintaxis, las categorías de la gramática corresponden a conceptos semánticos. Las gramáticas semánticas son útiles para desarrollar interfaces para dominios específicos y limitados, pero no tienen una portabilidad eficiente para nuevos dominios debido a que la información semántica es dependiente del dominio, lo cual puede tomarle meses al programador por cada base de datos [8, 15].

La técnica de lenguaje de representación intermedia utiliza reglas sintácticas y semánticas, estas reglas se encuentran en módulos independientes. Pero, a pesar de tener una arquitectura más completa requiere que toda la información sea

configurada previamente, lo cual la hace tediosa y difícil de utilizar cuando se desea cambiar de un dominio a otro.

Debido a las limitaciones de las técnicas convencionales, actualmente han surgido nuevas propuestas que aún se encuentran en fase experimental. Estos proyectos realizan un procesamiento lingüístico basado en patrones léxico-semánticos (PLS) y gramáticas multinivel para representar la estructura de la consulta y traducirla a SQL [16, 17]; mientras que otros utilizan grafos semánticos [18].

Hanmin Jung y Gary Geunbae Lee [15] proponen la aplicación de categorías semánticas y gramáticas basadas en patrones para la traducción de una consulta en lenguaje natural inglés o coreano a SQL. Esta técnica es una hibridación de los sistemas de reconocimiento de patrones y lenguajes de representación intermedia, donde los patrones cubren las relaciones léxico-semánticas y las gramáticas multinivel sirven para la representación intermedia [16, 17], pero como las clases semánticas son definidas por el usuario deben ser programadas para cada dominio.

Frank Meng y Wesley W. Chu [18] de la Universidad de California proponen la traducción de consultas del lenguaje natural inglés a SQL mediante un modelado semántico. La información de la base de datos es representada mediante un grafo, el cual se construye a partir del esquema de la base de datos e incluye relaciones definidas por el usuario.

La interfaz acepta la entrada en lenguaje natural del usuario y extrae la información necesaria. El proceso de extracción se realiza buscando palabras en la consulta que coincidan con las palabras del grafo semántico de la base de datos. Además se apoya en vectores de n-gramas para resolver ambigüedades. La consulta en SQL se construye a partir de las palabras de la consulta que fueron

localizadas en el grafo. La principal desventaja de esta técnica es la construcción semi automática del grafo.

El Departamento de Ciencias de la Computación de la Universidad de Concordia en Canadá [12, 13] propone la traducción de consultas en inglés a consultas en SQL utilizando plantillas semánticas. La consulta del usuario es etiquetada y analizada sintácticamente, y a partir de la información obtenida se realiza un análisis semántico usando plantillas semánticas.

La base de conocimiento semántico está compuesta por reglas y conjuntos semánticos para todas las tablas y nombres de relaciones en la base de datos. Una persona deberá relacionar las tablas y campos con una palabra en inglés, y a partir de ésta el sistema obtendrá una lista de sinónimos para esa palabra. Las *reglas semánticas* son reglas definidas por el usuario que permiten activar una acción o referir a cierta información cuando éstas aparecen en la consulta del usuario.

Como se puede observar, la principal desventaja de esta propuesta es la necesidad de crear plantillas para cada dominio; además de que la cobertura de la ILNBD estará limitada al número de plantillas implementadas.

La tabla 2.1 muestra las tecnicas usadas y su portabilidad en el dominio en las decadas pasadas.

Tabla 2.1 Técnicas de traducción utilizadas por las ILNBD's.

70´s	80´s	90´s
Reconocimiento de patrones. Arquitectura basada en sintaxis.	Gramáticas-semánticas. Predicados lógicos. Arquitectura de lenguaje de representación intermedio.	Uso de técnicas tradicionales Nuevas propuestas: grafos semánticos. patrones lexico-semánticos. gramáticas multinivel.
Dependencia total del dominio y la base de datos.	Diccionarios dependientes del dominio. El análisis semántico se basaba en la búsqueda de palabras clave.	Uso de diccionarios y ontologías para diferentes bases de datos de configuración manual.

2.2 REVISIÓN DE TRABAJOS RELACIONADOS

A pesar de que los primeros trabajos sobre interfaces de lenguaje natural surgieron en 1947, es hasta finales de los sesentas y principios de los setentas que se desarrollan las primeras interfaces de lenguaje natural para bases de datos, debido a que las bases de datos no alcanzaron su madurez hasta la aparición del modelo relacional de Codd [19].

Uno de los proyectos más sobresalientes en los años setentas fue LUNAR [20, 21], el cual manejaba una base de datos de estructura particular con información acerca de rocas lunares. Esta interfaz tenía una arquitectura basada en la sintaxis y predicados de primer orden, debido a lo cual era muy difícil de configurar para

utilizarse con nuevos dominios. LUNAR ayudaba al usuario a formular su consulta a través de diálogos. RENDEZVOUS [22] utilizaba técnicas semejantes a las de LUNAR.

A finales de los setentas aparecieron LADDER [23], PLANES [24] y PHILIQA1 [25], los cuales utilizaban gramáticas semánticas, una técnica que relacionaba el procesamiento sintáctico y semántico. La interfaz PHILIQA1 se enfocó principalmente en problemas semánticos y, al igual que LADDER, podía ser usada con grandes bases de datos y diferentes manejadores, pero debía ser configurada manualmente. Mientras que PLANES ayudaba al usuario mediante diálogos a formular sus consultas.

Las gramáticas semánticas utilizadas por los sistemas en los años setentas dificultaban la portabilidad del dominio, por lo que los investigadores abandonaron este tipo de técnicas y durante la década de los ochentas se enfocaron en la portabilidad del dominio [20].

A principios de los ochentas aparece CHAT-80 [26]. Este sistema traducía consultas en inglés a expresiones en Prolog, las cuales eran evaluadas sobre una base de datos de Prolog. Para realizar la traducción utilizaba predicados lógicos. Otras interfaces que también utilizaban predicados lógicos son NATLIN [27] y TELI [28]. NATLIN fue implementada en Prolog pero únicamente genera las consultas y, así como TELI, realiza un análisis sintáctico y semántico de la oración.

A mediados de los ochentas aparecen JANUS [29], DATALOG [30, 31], LDC [32, 33], TQA [34, 35], EUFID [36] y TEAM [37, 38]; y todavía algunas de las interfaces como JANUS Y EUFID utilizan la técnica de gramáticas semánticas. JANUS fue uno de los pocos sistemas en su tiempo que soportaba consultas temporales donde el usuario podía realizar consultas en tiempo presente, pasado y futuro.

En lo que respecta a la portabilidad de las interfaces, TEAM establece la importancia de dividir la información dependiente del dominio de la información que pudiera ser manejada de manera general, con la finalidad de proporcionar independencia a la interfaz del dominio que se esté utilizando, aunque no realizó grandes avances al respecto debido a que la técnica de traducción utilizada resultaba sumamente dependiente del dominio. TEAM tenía tres componentes principales: un componente de adquisición, un módulo basado en diálogos y un componente para acceso a datos. Además, presentaba los principios de la arquitectura de lenguaje de representación intermedio, ya que la consulta era convertida a una forma lógica y posteriormente a una representación formal.

A finales de los ochentas emergen las primeras ILNBDs comerciales, entre las cuales se encuentran ASK [39] y BBN´s Parlance [40]. ASK permitió a los usuarios enseñar a la interfaz nuevas palabras y conceptos en cualquier punto durante la interacción, mientras que BBN´s Parlance tenía módulos morfológicos que permitían al sistema determinar las diferentes formas de las palabras.

Hasta antes de la década de los noventas, la mayoría de las ILNBDs utilizaban como lenguaje natural el idioma inglés, pero a partir de esta fecha importantes investigaciones sobre las interfaces para el idioma español comienzan a surgir. Entre éstas se encuentran ILNES [41], Paso PC-315 [42], NATLIN [27, 43] y SISCO [44].

En el Instituto Tecnológico de Monterrey Campus Cuernavaca se creó el sistema ILNES, el cual consta de cinco componentes principales: un diccionario de símbolos, un diccionario de sinónimos, un diccionario de datos, un modelador del contexto de la base de datos, y un intérprete de lenguaje natural. Para obtener la consulta en SQL el interpretador utilizaba patrones que representan el conocimiento de la interfaz.

NATLIN fue desarrollada en la Universidad de Essex en Inglaterra, pero no podía utilizar una base de datos comercial porque generaba la consulta en Prolog; esto la limitaba notablemente, ya que no podía probarse con datos más complejos. Por tal motivo se continuó este proyecto en México, en la Universidad de las Américas, Puebla en donde adaptaron el sistema NATLIN para la generación de la consulta en SQL.

Paso PC315 de la Universidad Politécnica de Madrid y SISCO de la Universidad Politécnica de Alicante están basados en gramáticas de tipo lógico-modular.

En los años noventas se consolidan las ILNDBs con el surgimiento de MASQUE/SQL [53]. Esta interfaz está basada en la arquitectura de Lenguaje de Representación Intermedio y cuenta con un lexicón que tiene una lista de las posibles formas de las palabras que podrían ser usadas en la consulta del usuario, así como expresiones lógicas que describen el significado de cada palabra. El lenguaje intermedio que utiliza es LQL (Logical Query Language), y para obtener la consulta en SQL realiza mapeos a la base de datos y diccionarios mediante predicados lógicos. Puede manejar diferentes dominios, pero es necesario que éstos sean construidos mediante un editor de configuración que es semiautomático.

También durante la década de los noventas aparecen algunas interfaces comerciales, las cuales eran un poco más sofisticadas en comparación con las interfaces de la década anterior, pero todavía presentaban notables desventajas. A continuación se presentan las más importantes.

En el Laboratorio de Lenguaje Natural de la Universidad de Simon Fraser, nace SystemX [45]. Este sistema presenta un diseño modular, ya que posee un módulo

para el tratamiento de la consulta en lenguaje natural y otro para el análisis de la base de datos.

LOQUI [46], sistema desarrollado en Prolog por BIM, obtenía una consulta lógica y reglas expresadas en Prolog. La consulta lógica era tratada por el interpretador de Prolog y la información de la base de datos era tratada como hechos de Prolog.

INTELLECT [47], comercializado por IBM, era configurado para interactuar con la base de datos a través de un sistema experto, de esta manera se podía agregar una especie de razonamiento entre la ILNDB y la base de datos.

NATURAL LANGUAGE [48], de Natural Language Inc., tenía diccionarios que contenían una lista de las palabras más comunes. Los diccionarios podían ser administrados sólo por las personas que configuraban el sistema incluyendo la terminología del dominio específico.

Q&A [49] de Symantec era un sistema basado en menús, donde el usuario no podía escribir directamente sus consultas, por lo que éstas tenían que ser construidas escogiendo posibles palabras o frases desde menús.

SQ-HAL [50], de la Universidad de Monash, traduce consultas en lenguaje natural a consultas simples en SQL y está desarrollado en Perl. Posee un analizador sintáctico descendente recursivo y define una gramática propia (reglas de producción). Otros sistemas importantes son EXPEDIA Hotels [51] de Elf Software, Microsoft English Query [52] de Microsoft y LANGUAGE ACCESS [40] de IBM.

Entre 1997 y 1999 aparecen los primeros ILNBDs multilingües, destacando EDITE [14], VILIB [57] y AVENTINUS [58]. EDITE, al igual que MASQUE/SQL, está

desarrollado siguiendo la arquitectura de Lenguaje de Representación Intermedio, utilizando LIL (Logic Interface Language) como lenguaje intermedio.

Otras ILNBDs que utilizan la arquitectura de lenguaje intermedio son KID [54] y TAMIC-P [55]. KID, que maneja el idioma Coreano, utiliza OQL (Object Query Language) como lenguaje intermedio y patrones sintácticos para realizar la traducción, los cuales se construyen manualmente.

TAMIC-P es un sistema para el idioma alemán que asevera que se puede obtener una vista unificada de los datos del dominio y la jerarquía de conceptos a partir de las bases de datos y sus relaciones. Para realizar la traducción se auxilia de mapeos a las bases de datos para tener una consulta previa en lenguaje intermedio OQL y CPL.

En los últimos años han surgido importantes proyectos como InBase [56], basado en la arquitectura de Lenguaje de Representación Intermedio, y TABLEUX [16], interfaz para el idioma español, que utiliza autómatas y mapeos a la base de datos.

Todos los proyectos anteriormente mencionados enfocan su atención en el análisis léxico, sintáctico y semántico de la consulta en lenguaje natural; y en lo que respecta a la traducción de la oración de entrada, en una representación formal que pueda ser manipulada por la computadora con éxito. Las técnicas tradicionales que incorporan desde reconocimiento de patrones hasta predicados lógicos limitan a la interfaz a dominios y a bases de datos predefinidas por el administrador de la ILNBD. Por tal motivo se han comenzado a realizar nuevos proyectos que incorporan las ventajas de las técnicas utilizadas por los sistemas convencionales, pero con nuevas propuestas. A continuación se describen algunos de éstos.

Nchiql [17] es una ILNBD para el idioma chino. Este sistema traduce las consultas en lenguaje natural a SQL extrayendo información de la base de conocimiento del dominio, tal como entidades de la base de datos, atributos y relaciones. Para realizar la traducción se apoya en árboles semánticos.

PRECISE [11] fue desarrollado en la Universidad de Washington, y para realizar la traducción de una consulta en lenguaje natural, primero determina si es posible tratarla semánticamente. Para saber esto, mapea los elementos léxicos de la oración a los elementos de la base de datos a través de restricciones semánticas impuestas, reduciendo el problema a una comparación de grafos, el cual es resuelto con un algoritmo de flujo máximo. La desventaja principal es que el grafo semántico es diseñado manualmente por el usuario a través de un editor.

En la Universidad de California se está desarrollando una ILNBD que, para realizar la traducción, utiliza un modelado semántico que consiste en un grafo semántico de la base de datos, el cual representa la información almacenada de ésta. La interfaz acepta del usuario la consulta en lenguaje natural y extrae la información necesaria. El proceso de extracción es realizado usando palabras clave obtenidas del grafo semántico de la base de datos. Pero debido a que las palabras clave pueden tener diferentes significados sin un dominio dado, es necesario evitar ambigüedad de significados de las palabras clave usando aproximaciones estadísticas que involucran la comparación de vectores de n-gramas. Esta interfaz permite utilizar diferentes dominios, pero está limitada a oraciones de entrada cortas y simples, además de depender de una ontología limitada que soporta al dominio [18].

Por último, en Corea, se ha desarrollado una interfaz que utiliza patrones léxico-semánticos y gramáticas multinivel [59].

La tabla 2.2 muestra el desarrollo de las interfaces en las décadas pasadas.

Tabla 2.2 Desarrollo de las ILNBDs.

70´s	80´s	90´s
• LUNAR • LADDER • PLANES • PHILIQA1	• CHAT-80 • NATLIN • TELI • JANUS • EUFID • DATALOG • LDC • TQA • TEAM • ASK • BBN´S • PARLANCE	• ILNES • PASO PC-315 • NATLIN • SISCO • MASQUE/SQL • SYSTEM X • LOQUI • INTELLECT • NATURAL LANGUAGE • Q&A • SQ-HAL • EXPEDIA HOTELS • MS ENGLISH QUERY • LANGUAGE ACCESS • EDITE • VILIB • AVENTINUS • KID • INBASE • TAMIC-P • TABLEUX • NCHIQL • PRECISE

2.3 METODOLOGÍAS UTILIZADAS EN LA EVALUACIÓN DE ILNBDs

Al desarrollar una ILNBD, sus creadores deben probar la efectividad de su interfaz. Para ello los investigadores han recurrido a diferentes métodos para determinar el éxito de su sistema bajo diferentes criterios.

Según Ogden y Bernick [60] algunas metodologías importantes que han sido utilizadas en la evaluación de ILNBDs son: las simulaciones del Mago de Oz, la selección y entrenamiento de usuarios, generación y presentación de tareas y la evaluación del porcentaje de éxito. Sin embargo, [61] se han considerado otros métodos como: el uso de conjuntos de prueba, estudios de campo y tareas de pluma y papel.

Algunas interfaces existentes, o los prototipos de éstas, ya han sido sometidas a procesos de evaluación. A continuación se presenta una descripción de las pruebas que se han aplicado para estas evaluaciones.

Tabla 2.3 ILNBDs dependientes del dominio.

Interfaz	**Descripción**	**Método**
LADDER	Base de datos naval.	Selección y entrenamiento de usuarios para una simulación.
PLANES	BD de registros de mantenimiento y vuelos de la fuerza aérea de la marina.	Generación y presentación de tareas.
Automatic Advisor	Información de cursos de ingeniería ofrecidos por una universidad.	Generación y presentación de tareas.

TQA	Registros de cada parcela de tierra en la ciudad.	Estudios de campo.
INTELLECT	Problemas relacionados a trabajos.	Estudios de campo.

Las tablas 2.3 y 2.4 muestran un resumen de las interfaces, clasificadas de acuerdo a la portabilidad del dominio. Las interfaces dependientes del dominio fueron probadas en su mayoría por grupos de usuarios en evaluaciones de laboratorio o estudios de campo. En los experimentos destaca la importancia del entrenamiento, mediante el cual se le da a conocer a los usuarios el dominio de la información y la cobertura conceptual y funcional, lo que permite obtener mejores resultados. Para el caso de las interfaces independientes del dominio, se observa el uso de porcentajes de éxito como única medida de evaluación.

Tabla 2.4 ILNBDs independientes del dominio.

Interfaz	**Descripción**	**Método**
Sistema Basado en Plantillas	BD de la biblioteca Cindi. En idioma inglés. 15 tablas.	Selección y entrenamiento de usuarios. Porcentaje de éxito.
Respuesta a preguntas de diversos lenguajes	BD productos de audio-video Idiomas inglés y coreano. BD de comparación de precios. Idioma coreano.	Conjuntos de prueba. Evaluación del porcentaje de éxito.
PRECISE	BD restaurantes, trabajos y geografía. BD ATIS.	Precisión de la interfaz y rechazo. Porcentaje de respuestas.

La tabla 2.5 muestra una comparación de la efectividad de varias interfaces de lenguaje natural en el dominio de la base de datos ATIS. Se puede observar que las mejores ILNBDs dependientes del dominio tienen un **96.2 de porcentaje de**

acierto en las respuestas y las mejores ILNBDs independientes del dominio tienen un **94 de porcentaje de acierto en las respuestas**.

Tabla 2.5 Comparación de efectividad de ILNBDs en el dominio ATIS

PRECISE	PRECISE-1	AT&T	CMU	MIT	SRI	BBN	UNISYS	MITRE	HEY
94.0%	89.1%	**96.2%**	**96.2%**	95.5%	93%	90.6%	76.4%	69.4%	92.5%

Sin embargo, para que la interfaz Precise pueda alcanzar el 94% de éxito es necesario reentrenar el analizador sintáctico, la tabla 2.6 muestra los resultados de un experimento realizado por los desarrolladores de Precise para medir la efectividad de las configuraciones de la interfaz Precise [62]. En donde la columna **PRECISE** almacena el porcentaje de consultas contestadas correctamente cuando esta interfaz obtiene varias posibles respuestas para una misma consulta. **PRECISE-1** es forzado a contestar una sola respuesta. Esto muestra que aunque se dice que la interfaz es independiente del dominio, ésta requiere de una configuración manual realizada por un experto humano para alcanzar los porcentajes descritos.

Tabla 2.6 Porcentajes de efectividad de PRECISE con sus distintas configuraciones

Configuración del sistema	PRECISE	PRECISE -1
Analizador sintáctico original	61.9%	60.3 %
Analizador sintáctico entrenado	92.4%	88.2%
Analizador sintáctico correcto	95.8%	91.9%

3 METODOLOGÍA DE SOLUCIÓN

El objetivo de esta investigación es diseñar una técnica de traducción que permita convertir una consulta en lenguaje natural a SQL satisfactoriamente y que pueda ser reutilizada sin importar el dominio, es decir, que el administrador de la ILNBD o el usuario no tengan que hacer modificaciones al código o realizar tediosas configuraciones para que ésta funcione en un dominio diferente. Para lograr esta independencia, se consideran dos puntos importantes:

1. La generación automática de un subconjunto del lenguaje, denominado dominio, a partir de la base de datos dada.
2. El diseño de una técnica de traducción cuyos procesos sean independientes de la información de la base de datos.

En la siguiente figura se muestran los principales aspectos que soportan el diseño de la técnica de traducción propuesta, los cuales se analizan en las siguientes secciones.

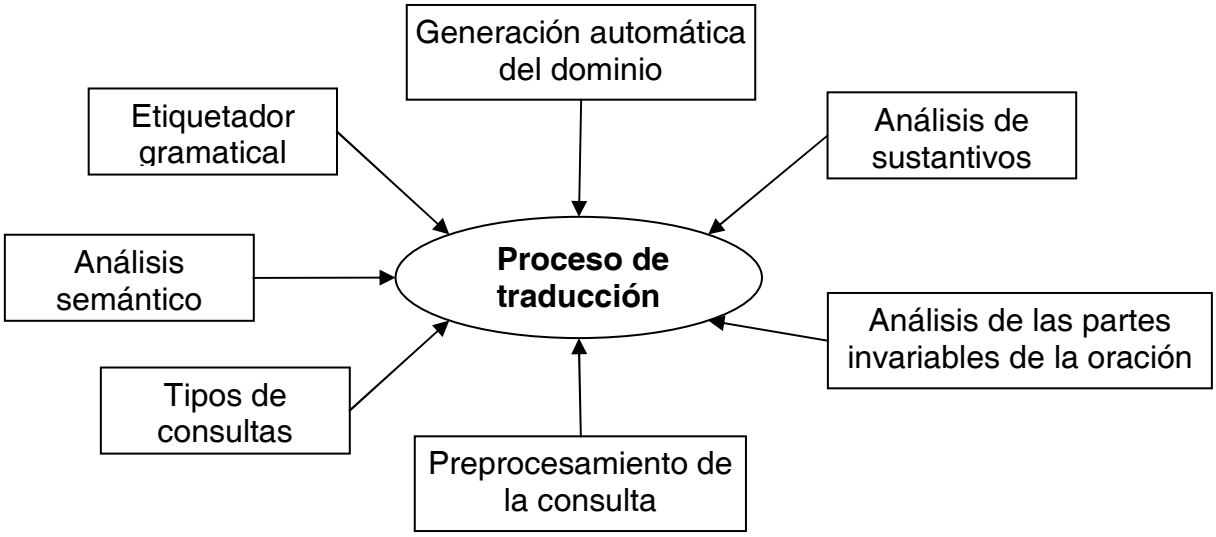


Figura 3.1 Aspectos considerados en el proceso de traducción.

3.1 GENERACIÓN AUTOMÁTICA DEL DOMINIO DE LA ILNBD

El dominio se encuentra estrechamente relacionado con el concepto de base de datos. Una base de datos es un conjunto de datos que modelan los objetos de una parte del mundo real. Estos conjuntos están relacionados mediante una determinada estructura lógica y sirven de soporte a una aplicación informática [63].

La generación automática del dominio evita las configuraciones tediosas para el usuario o administrador, y se logra explotando los metadatos y diferentes tipos de diccionarios y/o ontologías.

Para la generación automática del dominio es necesario un buen diseño de la base de datos, el cual incluye la creación de metadatos técnicos adecuados.Los metadatos técnicos se registran e integran en la fase de diseño y modelado de cualquier aplicación de cómputo o base de datos moderna, para quedar registrados en lo que se conoce como *diccionario de datos*.

Si bien se han presentado diferentes propuestas, hasta la fecha ninguna está cercana a alcanzar la meta general de que una ILNBD sea independiente de la base de datos y precisa [13]. Esto se debe principalmente a que el uso del lenguaje natural es demasiado variado como para poder representarlo completamente [64]. Las ILNBDs existentes sólo trabajan con un subconjunto del idioma, al que frecuentemente se le denomina dominio, restringiendo la interfaz a procesar sólo las consultas que se encuentren dentro de éste.

Un lenguaje es un conjunto de palabras y métodos para combinar palabras, que es usado y entendido por un extenso grupo de personas, pero que es complejo y difícil de caracterizar en su totalidad. Una base de datos es una abstracción del mundo real. Con el propósito de modelar estas entidades, es importante el uso de lexicones que contengan la mayoría o todas las palabras utilizadas en el idioma, además de diccionarios (sinónimos, antónimos, etc.) u ontologías que aporten información para el análisis posterior.

Hasta la fecha, los diccionarios utilizados por las ILNBDs desarrolladas han sido creados manualmente o semiautomáticamente [8]. En este trabajo se construye el diccionario de dominio para la interfaz de forma automática a partir de un diccionario de sinónimos, un etiquetador gramatical y un diccionario de metadatos.

Diccionario de Sinónimos: Este diccionario se creó utilizando una enciclopedia digital. Actualmente contiene aproximadamente 20,000 palabras con sus sinónimos y antónimos. Este diccionario puede utilizarse de manera inmediata en cualquier dominio, debido a que es un diccionario de sinónimos general. Además, cuenta con una interfaz que permite agregar sinónimos y antónimos.

Diccionario de Metadatos: La meta información de la base de datos puede ser usada como un recurso para una mejor interpretación de la consulta en un

dominio limitado [65]. Los metadatos tienen gran relevancia, tanto para el usuario como para el productor de la información. Pero a pesar de que los metadatos describen a los datos en términos técnicos de contexto, contenido, disponibilidad y vigencia, éstos han sido tratados como un asunto menor puramente documental y estático [66]. Este diccionario contiene información de la base de datos como el número de tablas, el total de columnas, la ubicación. Además almacena información específica para cada tabla como el nombre de cada columna junto con su descripción y el tipo de dato, así como también cuáles son las llaves primarias y las llaves foráneas que contiene, por lo que se pueden conocer las relaciones entre las tablas. Es importante señalar que en esta propuesta se decidió utilizar las descripciones de las columnas y tablas para la creación del diccionario de dominio, debido a que éstas contienen información referente a sus nombres, ya que por lo regular, los nombres de columnas y tablas no se consideran significativos y la mayoría de las veces se encuentran abreviados [66].

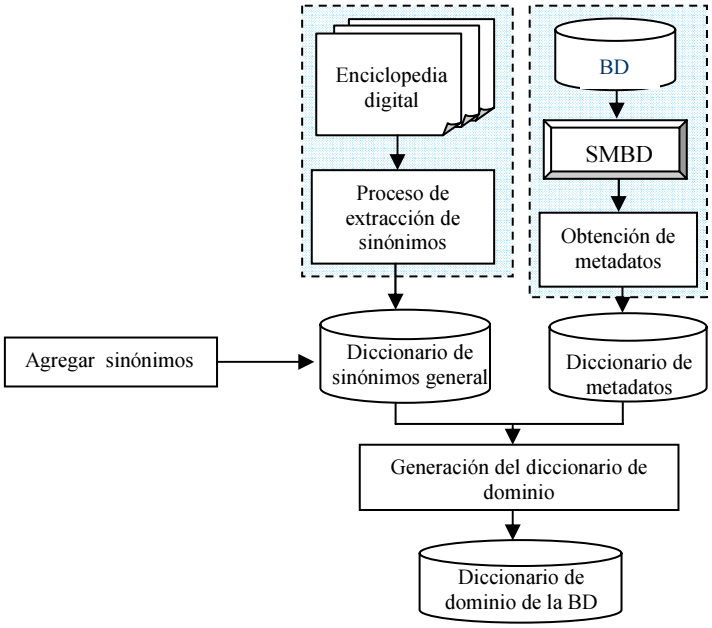


Figura 3.2 Construcción del Diccionario de Dominio

Diccionario de Dominio: Para construir el diccionario de dominio, se toma primero la información referente a la descripción de cada columna, contenida en el diccionario de metadatos. Mediante un etiquetador gramatical se obtiene el lema y la categoría sintáctica de cada palabra de la descripción. A continuación cada nombre de columna se relaciona con los sustantivos que aparecen en su descripción y con los sinónimos de cada sustantivo. Como se puede observar, es más sencillo dar descripciones significativas a columnas y tablas, para que a partir de éstas la interfaz se configure automáticamente, que configurar manualmente los diccionarios y módulos de la interfaz, para que ésta reconozca y relacione cada tabla y columna con alguna palabra del diccionario de dominio.

3.2 ETIQUETADOR GRAMATICAL

La función del etiquetador es la de proporcionar las etiquetas léxicas para cada palabra de las siguientes dos oraciones: las descripciones de las columnas y la consulta del usuario. El etiquetador hace uso de un lexicón que contiene un conjunto de palabras con su respectiva etiqueta gramatical y lema.

3.3 PREPROCESAMIENTO DE LA CONSULTA

Este preprocesamiento consiste en analizar cada palabra de la consulta en lenguaje natural para obtener su información léxica, sintáctica y semántica. La información léxica consiste en el lema de la palabra, y la información sintáctica se refiere a su categoría sintáctica (verbo, sustantivo, preposición, etc.). Esta información se obtiene mediante el etiquetador gramatical. La información semántica de cada palabra se obtiene del diccionario de dominio, de tal manera que cada palabra estará asociada a un conjunto de posibles tablas y/o columnas a las cuáles puede hacer referencia la consulta.

Tabla 3.1 Ejemplo del preprocesamiento de una consulta

Consulta: *cuáles son los nombres y las direcciones de los empleados*				
Palabra	**Lema**	**Información morfosintactica**	**Columnas**	**Tablas**
cuáles	cuál	interrogativa		
son	ser	verbo, indicativo, 3rd persona, plural		
los	el	plural, masculino		
nombres	nombre	plural, masculino **sustantivo**	Categories.CategoryName, Customers. CompanyName, Employees.FirstName, Orders.ShipName, ...	
y	y	**conjunción**		
direcciones	dirección	plural, femenino **sustantivo**	Employee.Address, Orders.ShipAddress, Suppliers.Address, Customers.Address	
de	de	**preposición**		
los	el	plural, masculino		
empleados	empleado	plural, masculino **sustantivo**	Employee.EmployeeID, Orders.EmployeeID	Employee.

Por lo tanto, si sólo mediante las palabras que conforman la consulta puede hacerse referencia a elementos de la base de datos, es importante hacer un análisis de éstas. Por lo cual a continuación se hará una revisión de algunos

elementos que componen la oración de los cuales hace uso la técnica de traducción propuesta.

3.4 SUSTANTIVOS DE LA ORACIÓN

Los sustantivos indican el nombre de una persona cosa o elemento y en una base de datos comúnmente hacen referencia al nombre de las columnas o tablas. Las palabras clave que utilizan algunas interfaces son sustantivos que hacen referencia a columnas o tablas de la BD. Es por esto la importancia que tiene el análisis de los sustantivos en la consulta.

3.5 PARTES INVARIABLES DE LA ORACIÓN

Una oración es una palabra o conjunto de palabras que se caracteriza por poseer un sentido completo. Algunas de estas palabras pueden mantenerse siempre iguales; es decir, no variar ni en género ni en número independientemente de lo que se diga (partes invariables de la oración), mientras que otras pueden variar en género y número (partes variables de la oración) [61].

En la investigación realizada se observó que el análisis de las partes invariables de la oración como apoyo en el proceso de traducción no ha sido explorado suficientemente, ya que la mayoría de las ILNBDs buscan palabras clave dentro de la oración [12, 45, 67] o centran el análisis en los sustantivos y verbos [18], eliminando parte de las palabras de la consulta como preposiciones, conjunciones o adverbios, que son necesarias para poder interpretar ésta correctamente.

Es importante resaltar que el análisis de las partes invariables de la oración (preposiciones, conjunciones y adverbios) puede aportar información importante al proceso de traducción, ya que su ausencia o presencia puede cambiar por completo el sentido de la oración.

Como se mencionó anteriormente, dentro de las partes invariables de la oración se encuentran las preposiciones y conjunciones. La función que éstas realizan puede variar de un idioma a otro. Dentro de la gramática española estas categorías de palabras también son conocidas como *elementos de relación* debido a que su misión es relacionar unos elementos con otros. A continuación se presenta una descripción de las preposiciones y conjunciones que se utilizan en el idioma español.

3.5.1 PREPOSICIONES

La preposición, como se mencionó anteriormente, es una partícula invariable que sirve para enlazar una palabra principal (núcleo sintáctico) con su complemento (vaso de vino, voy a Roma) [65].

Las preposiciones españolas son: a, ante, bajo, cabe, con, contra, de, desde, en, entre, hacia, hasta, para, por, según, sin, so, sobre y tras.

3.5.2 CONJUNCIONES

La conjunción ha sido definida tradicionalmente como la parte de la oración que sirve para unir dos o más elementos en una relación de igualdad [68]. Las conjunciones copulativas dan una idea de suma o acumulación y corresponden a la *e* y la *y*.

3.6 CLASIFICACIÓN DEL TIPO DE CONSULTAS

Para el ser humano hablar su idioma natal es un proceso que se da de manera natural, cuando este proceso se da entre dos o más personas se denomina comunicación verbal. Ésta puede ir acompañada de ademanes o gestos que

permiten omitir algunas palabras en la comunicación que se da entre dos personas; algunos fenómenos documentados de esto son la elipsis y la anáfora.
En esta sección se presenta una clasificación del tipo de consultas en relación con el tipo de información que el usuario expresa en su consulta. Esta clasificación permitirá conocer qué tipo de consultas realiza el usuario con mayor frecuencia.

Las consultas del usuario se clasifican según los siguientes términos:

Columnas explícitas y/o implícitas: Una columna es implícita cuando se hace referencia a ella mediante el contexto de la oración, en cambio una columna es explícita cuando se encuentra explícitamente en la consulta. Por ejemplo “Dame el teléfono de Erika Alarcón Ruiz”, *teléfono* seria una columna explícita y *nombre* una columna implicada por “Erika Alarcón Ruiz”.

Tabla explícitas y/o implícitas: Puede ser, al igual que las columnas, implícita y explícita. Por ejemplo:

Dame el teléfono de Erika Alarcón Ruiz.	Consulta con tabla implícita.
Dame el teléfono del *empleado* Erika Alarcón Ruiz.	Consulta con tabla explícita.

Condiciones: En una consulta el usuario puede incluir condiciones que deben de ser consideradas cuando se realiza la traducción de la consulta a SQL.
Ejemplos:

Dame los nombres de los empleados que viven en Cd. Madero.

Quién está contratado desde el 01/04/1992.

Con funciones SQL: Se refiere a las funciones propias del lenguaje SQL que el usuario puede utilizar en sus consultas, como por ejemplo las funciones de agregación o de resumen: AVG, COUNT, MIN, MAX, SUM.

Ejemplos:

Dame el nombre del empleado con el salario más alto.

Dame el promedio de las ventas del mes.

Difíciles de Resolver: Este término se les da a las consultas que requieren que sean planteadas de otra forma, debido a que no hay información suficiente para el procesamiento de la consulta, o que requieren de procesos de deducción o funciones que necesitan ser programadas.

Ejemplos:

Dame el nombre del empleado que nació después del empleado que nació el año de 1948.

Dame la información de Erika Alarcón Ruiz.

Dame los nombres de los empleados que nacieron en años bisiestos.

En la tabla 3.2 se muestra la clasificación de los tipos de consulta con base en los términos anteriormente descritos.

Tabla 3.2 Clasificación del tipo de consultas.

Grupo	Columnas	Tablas	Funciones SQL	Condiciones
A	1. Explícitas	Explícitas	No	No
	2. Implícitas	Explícitas	No	No
	3. Explícitas	Implícitas	No	No
	4. Implícitas	Implícitas	No	No
B	5. Explícitas	Explícitas	Sí	No
	6. Implícitas	Explícitas	Sí	No

	7. Explícitas	Implícitas	Sí	No
	8. Implícitas	Implícitas	Sí	No
C	9. Explícitas	Explícitas	No	Sí
	10. Implícitas	Explícitas	No	Sí
	11. Explícitas	Implícitas	No	Sí
	12. Implícitas	Implícitas	No	Sí
D	13. Explícitas	Explícitas	Sí	Sí
	14. Implícitas	Explícitas	Sí	Sí
	15. Explícitas	Implícitas	Sí	Sí
	16. Implícitas	Implícitas	Sí	Sí
E	17. Difíciles de contestar			

3.7 ANÁLISIS SEMÁNTICO

Una de las tareas clave de la interpretación semántica, consiste en considerar qué combinaciones de significados de palabras individuales son posibles a la hora de crear un significado coherente de la oración. Esto puede reducir el número de posibles significados para cada palabra de una oración determinada [27].

Analizar semánticamente una oración es todavía una tarea compleja [69], ya que tan sólo el hecho de definir los sentidos de las palabras en sí ya es muy difícil, debido a la polisemia; por ejemplo, "gato" es un felino y también una herramienta.

El análisis semántico de una consulta en lenguaje natural involucra, en la mayoría de los casos, la búsqueda de palabras clave dentro de la oración de entrada las cuales son evaluadas según un patrón establecido, apoyándose en múltiples mapeos a la base de datos. En algunas investigaciones se realiza el análisis semántico a partir de modelos semánticos probabilísticos [60, 70]. Estos modelos requieren de un corpus etiquetado con información semántica, además de que

este enfoque es subjetivo y requiere de un trabajo manual considerable. Su aplicación se limita a tareas específicas dentro de un dominio semántico restringido [64].

En esta investigación se propone el uso de un grafo semántico para la representación de la consulta del usuario. El grafo semántico funciona como una representación intermedia entre la consulta en lenguaje natural y su correspondiente consulta en SQL.

El análisis semántico se encuentra estrechamente relacionado con el proceso de traducción de la consulta. Éste necesita principalmente una estructura que permita almacenar los significados y relaciones de cada una de las palabras que conforman la oración, un diccionario que contenga el conocimiento general (palabras, significados, relaciones, sinónimos, antónimos, etc.) y algoritmos o técnicas que permitan obtener la información precisa para poder realizar la traducción a un lenguaje formal sin importar el dominio.

3.8 DISEÑO DE LA TÉCNICA DE TRADUCCIÓN

En los trabajos revisados en esta investigación, ninguna ILNBD hace mención acerca de la necesidad de condiciones para el diseño de la base de datos o para la formulación de las consultas. El presente trabajo requiere del cumplimiento de algunas condiciones en el diseño de la base de datos y en la formulación de consultas en lenguaje natural, con el objeto de generar automáticamente el dominio y realizar una correcta traducción de la consulta.

3.8.1. CONDICIONES DE LA INTERFAZ

C1. La base de datos debe ser relacional y puede ser modelada de acuerdo al modelo relacional o entidad-relación.

C2. Cada tabla tiene una llave primaria explícita.

C3. Cada columna de una tabla está asociada explícitamente a un dominio de información (el mismo tipo de dato).

C4. Las relaciones referenciales entre tablas están expresadas explícitamente a través de llaves foráneas.

C5. Cada tabla y cada columna tienen una descripción textual.

C6. Todas las tablas están en la segunda forma normal.

Las condiciones C1-C6 son básicas en un buen diseño de bases de datos relacional, además la información necesaria para las condiciones C2-C5 puede ser extraída de los metadatos de la base de datos. Una de las propiedades deseables en el diseño de una base de datos es la aciclicidad de su esquema, esto se obtiene mediante la normalización, ya que cuando el esquema de una base de datos es cíclico se presenta una mayor complejidad para obtener las relaciones de una consulta [71, 72].

Dado que las descripciones de las tablas y columnas son cruciales para una interpretación y traducción correctas de las consultas, se suponen las siguientes condiciones:

C7. Las descripciones de las tablas y de las columnas son léxica y sintácticamente correctas.

La formulación de descripciones léxica y sintácticamente correctas es responsabilidad del administrador de la base de datos al momento de diseñar la base de datos.

C8. Las descripciones de las tablas y de las columnas son frases significativas breves, las cuales consisten de al menos un sustantivo y opcionalmente algunas palabras importantes (sustantivos y adjetivos, excepto verbos) y algunas otras de menor importancia (artículos, preposiciones y conjunciones).

C9. La palabra más significativa de una descripción es un sustantivo.

Las condiciones C8 y C9 son importantes, ya que el proceso de traducción propuesto se basa principalmente en el análisis de sustantivos que aparecen en las descripciones.

C10. La descripción de cada tabla es diferente de la descripción de alguna otra tabla o columna.

C11. La descripción de cada columna de una tabla es diferente de la descripción de alguna otra columna de la misma tabla (las columnas en tablas diferentes pueden tener la misma descripción).

C12. La descripción de una columna significativa, aquélla que se refiere a un identificador o nombre, debe incluir el nombre de la tabla.

C13. La conjunción *y* se tratará siempre como una disyunción.

Las consultas que exprese el usuario en lenguaje natural deben tener las siguientes condiciones:

A1. Las consultas deben ser léxica y sintácticamente correctas.

El usuario debe considerar esta condición para obtener una respuesta satisfactoria.

A2. La información solicitada en la consulta a la base de datos debe estar antes de la condición.

A3. Las consultas pueden ser en forma interrogativa e imperativa.

A4. La consulta debe ser clara, no ambigua, es decir, debe tener información suficiente para responder correctamente.

Por ejemplo: Muéstrame el nombre de la compañía, en este caso el usuario no indica de cuál compañía requiere información (clientes, proveedores o fletadores).

A5. Si la consulta incluye un valor numérico o tipo fecha, deberá tener explícitamente uno o más sustantivos que hagan referencia a la columna que almacena el valor en la base de datos, ver ejemplo 3.6.

A6. Si la consulta contiene un valor compuesto por dos o más palabras, el cual corresponde a una columna de la base de datos, deberá estar escrito entre comillas dobles para poder encontrar su ubicación en la base de datos. Por ejemplo: nombres de personas, domicilios, etc. ("Juan Pérez Fernández").

A7. El formato aceptado para representar fechas es: dd/mm/aaaa.

A continuación se describe el proceso de traducción utilizado por la interfaz, tomando como base las condiciones anteriormente descritas.

3.8.2 PROCESO DE TRADUCCIÓN

El proceso de traducción se lleva a cabo en tres fases: (1) identificación de la frase select y where; (2) identificación de tablas y columnas, y (3) construcción del grafo relacional.

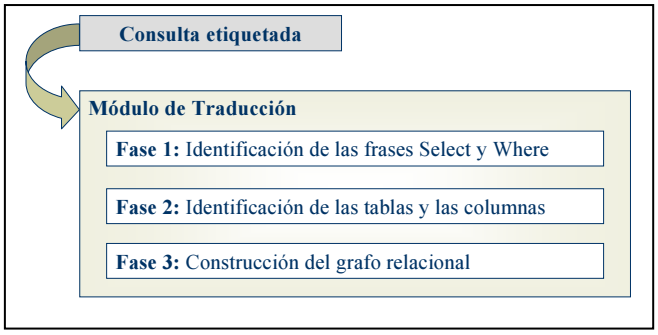


Figura 3.3 Módulo de traducción

Fase 1: Identificación de las frases select y where.

Una consulta en lenguaje natural está formada por los siguientes componentes:

- La parte que contiene los requerimientos, llamada frase select.
- La parte que contiene la condición, llamada frase where.

Para realizar el análisis de una consulta es importante separarla en sus componentes con el propósito de identificar correctamente cuál es la información requerida y cuál es la condición que esta información debe cumplir. Es posible que la frase select de una consulta anteceda su frase where, por ejemplo: "Dame los empleados con fecha de contratación 23/07/1996". También existe la posibilidad de encontrar una consulta que inicie con la frase where y que finalice con la frase select, por ejemplo: "De la ciudad de México, dame los clientes".

Para identificar las frases select y where la consulta se tiene que dividir en dos partes, por lo que se hace uso de las condiciones C8 y A2: la condición C8 dice: cada frase en la consulta incluye al menos un sustantivo (y posiblemente preposiciones, conjunciones, artículos, adjetivos, etc.). La condición A2 dice: la frase que define la cláusula select generalmente precede a la frase que define la cláusula where. Después de un estudio de los corpus obtenidos en este trabajo fueron detectados dos criterios para la separación de las frases de una consulta: la

presencia de dos sustantivos adyacentes y la presencia de valores. Estos dos criterios permiten obtener las dos frases de la consulta, dejando en cada una los elementos necesarios para su análisis.

Para realizar la separación de ambas frases es importante mencionar que solamente son tomados en cuenta los sustantivos (tabla y columna), la conjunción *y*, la preposición *de* y los valores de la oración, debido al diseño de esta técnica de traducción.

1. El primer criterio a considerar es la presencia de dos sustantivos adyacentes. La consulta se divide al separar ambos sustantivos.

Ejemplos:

Ejemplo: ¿Cuál es el apellido del empleado con fecha de contratación 23/07/1996?

Elementos de la consulta analizados por la interfaz						
apellido	de	**empleado**	**fecha**	de	contratación	23/07/1996
sustantivo (columna)	prep.	**sustantivo (tabla)**	**sustantivo (columna)**	prep.	sustantivo (columna)	valor

Para este caso la frase where se forma a partir del segundo sustantivo adyacente.

Separación de la consulta	
frase select	**frase where**
apellido de empleado	fecha de contratación 23/07/1996

Ejemplo: En que ciudad se encuentra la compañía “Exotic liquids”

Elementos de la consulta analizados por la interfaz		
ciudad	**compañía**	“Exotic liquids”
sustantivo (columna)	**sustantivo (columna)**	valor

La frase where se forma a partir del segundo sustantivo adyacente.

Separación de la consulta	
frase select	**frase where**
Ciudad	compañía “Exotic liquids”

2. Si no existen sustantivos adyacentes, se utiliza el segundo criterio, la presencia de valores. Se pueden presentar los siguientes casos:

- Cuando se encuentra un valor pero no hay sustantivo previo que haga referencia a tablas, la oración se divide desde el inicio de la oración hasta antes de donde se encuentra el valor.

Ejemplo: Dame el identificador de “Nancy Davolio”

Elementos de la consulta analizados por la interfaz		
Identificador	de	**Nancy Davolio**
Sustantivo (columna)	prep.	**valor**

En esta consulta no existen sustantivos adyacentes, entonces se busca el primer valor presente en la consulta y como no existen sustantivos que hagan referencia a tablas, la frase where se forma a partir del valor encontrado.

Separación de la consulta	
frase select	**frase where**
identificador	Nancy Davolio

- Cuando se encuentra un valor y existe un sustantivo previo que haga referencia a tablas, la oración se divide después de donde está el sustantivo que hace referencia a tablas.

Ejemplo: ¿Cuáles son los empleados de la ciudad de México?

Elementos de la consulta analizados por la interfaz				
empleados	de	ciudad	de	**México**
sustantivo (tabla)	prep.	sustantivo (columna)	prep.	**valor**

En esta consulta, al igual que la consulta anterior, no existen sustantivos adyacentes, entonces se busca el primer valor presente en la consulta y se retrocede en ésta hasta encontrar el último sustantivo que haga referencia a tablas antes del valor. Por lo tanto la consulta queda separada de la siguiente manera.

Separación de la consulta	
frase select	**frase where**
empleados	ciudad de Mexico

Fase 2: Identificación de tablas y columnas.

Generalmente cada sustantivo en la frase select o where puede hacer referencia a una o varias columnas o tablas de la base de datos (ver tabla 3.1), lo cual podría producir varias posibles traducciones de la consulta. Por lo tanto, para determinar con exactitud las columnas y tablas referidas, es necesario analizar la preposición *de* y la conjunción *y*, dado que casi siempre aparecen en las frases select o where [68]. El análisis de las preposiciones y conjunciones permite, considerando también el significado individual de los sustantivos, determinar el significado preciso de una frase select o where que involucre sustantivos relacionados por preposiciones y conjunciones. En la Universidad Politécnica de Madrid se realizó una investigación donde reportan las 10 palabras más frecuentes usadas en un diario local, figurando la preposición "de" en primer lugar y la conjunción "y" en sexto lugar [28]. Es por esta razón que la preposición *de* y la conjunción *y* serán abordadas en este trabajo. Éstas son representadas como operaciones utilizando la teoría de conjuntos, debido a la función que desempeñan dentro de la consulta.

La preposición *de* establece una estrecha relación entre una palabra y su complemento [73], de tal manera que, si existe una frase select o where que incluya dos sustantivos *p* y *q* unidos por una preposición *de*, entonces la frase se refiere a los elementos comunes (columnas) referidos por *p* y *q*. Formalmente, *S*(*p prep_de q*) = *S*(*p*) ∩ *S*(*q*), donde *S(x)* es el conjunto de columnas referidos por un sustantivo o el resultado de otra operación.

Sin embargo, cuando la preposición de relaciona nombres de columnas y tablas, entonces la preposición *de* implica la unión de la intersección entre las columnas de los dos conjuntos ($\{c_p \cap c_q\}$), la intersección de las columnas de un conjunto y las tablas del otro ($\{c_p \cap t_q\}$) y viceversa ($\{t_p \cap c_q\}$), y la unión de las tablas de ambos conjuntos ($\{t_p \cup t_q\}$).

$$S(p) \cap S(q) = \{\{c_p \cap c_q\} \cup \{c_p \cap t_q\} \cup \{t_p \cap c_q\} \cup \{t_p \cup t_q\}\}$$

Este tipo de operación sólo es valida entre sustantivos que representen columnas o tablas de la BD. En caso de que la preposición establezca una relación entre un conjunto de columnas o tablas y un valor, se tratará de manera distinta. Si la preposición *de* opera sobre una columna y un valor (v.q. salario de 20,000), se tomará como una asignación.

La conjunción *y* expresa una idea de suma o acumulación [73], de tal manera que, si existe una frase select o where que incluya dos sustantivos *p* y *q* unidos por una conjunción *y*, entonces la frase se refiere a todos los elementos referidos por *p* y *q*. Formalmente, $S(p\ conj\_y\ q) = S(p) \cup S(q)$. La conjunción *y* en una frase where es tratada como una operación booleana.

Cuando una consulta requiere de dos o más operaciones de unión o intersección, es necesario determinar el orden en el que se deben realizar las operaciones. Por tal motivo, para la evaluación de las preposiciones y conjunciones, se le asignan valores numéricos de la siguiente manera:

Nomenclatura usada: [n1 prep/conj n2]

- n1: indica el valor numérico del sustantivo previo a la preposición o conjunción (1 hace referencia a columnas, 2 hace referencia a tablas).
- prep/conj: indica una preposición *de*, una contracción *del*, una conjunción *y* o una conjunción *e*.
- n2: indica el valor numérico del sustantivo siguiente a la preposición o conjunción (1 hace referencia a columnas, 2 hace referencia a tablas).

Casos:

1. Se verifica el patrón [2 prep 2], si éste ocurre se cambia a [1 prep 2], sólo cuando el primer sustantivo es una columna que pertenece a la tabla a la cual hace referencia el segundo sustantivo.

2. Se le da un valor numérico de 1 a las preposiciones que cumplan [1 prep 1] o [2 prep 2].
3. Se le da un valor numérico de 2 a las conjunciones que cumplan [1 conj 1] o [2 conj 2].
4. Se le da un valor numérico de 3 a las preposiciones que cumplan [1 prep 2].
5. Se le da un valor numérico de 4 a cualquier otro caso.

Ahora sólo se procesan las preposiciones y conjunciones de acuerdo a su valor numérico, primero se procesan aquellas preposiciones y conjunciones que tengan valor numérico de 1, luego las que tengan un valor de 2 y así sucesivamente.

Los siguientes ejemplos muestran la asignación de la prioridad a las operaciones de unión e intersección de acuerdo con el caso que presenten.

Ejemplo 3.1 Dame el **identificador de empleados y clientes**
Esta consulta muestra el caso 4 ([1 prep 2] identificador de empleado) y el caso 3 ([2 conj 2] empleados y clientes).

Consulta a procesar	identificador	de	empleados	y	clientes
Valores numéricos	1	3	2	2	2
Prioridad		3		2	

Ejemplo 3.2: Dame la **dirección y el teléfono de** los **empleados y clientes**
Este ejemplo presenta el caso 3 ([1 conj 1] dirección y telefono; [2 prep 2] empleados y clientes) y el caso 4 ([1 prep 2] teléfono de empleados).

Consulta a procesar	dirección	y	teléfono	de	empleados	y	clientes
Valores numéricos	1	2	1	3	2	2	2
Prioridad		2		3		2	

Ejemplo 3.3: Dame la **dirección de** los **empleados y** el **teléfono de** los **clientes**
Esta consulta contiene el caso 4 ([1 prep 2] dirección de empleados; [1 prep 2] telefono de clientes) y el caso 5 ([2 conj 1] empleados y teléfono).

Consulta a procesar	dirección	de	empleados	y	teléfono	de	clientes
Valores numéricos	1	3	2	4	1	3	2
Prioridad		3		4		3	

Las siguientes dos consultas presentan el caso 1, donde es necesario cambiar el valor del primer operando, en caso de que éste sea una columna del segundo operando.

Ejemplo 3.4: Cuál es la **región de** los **empleados y** los **clientes**

Consulta a procesar	región	de	empleados	y	clientes
Valores numéricos	2		2		2
Se Cambia	1	3	2	1	2
Prioridad		3		1	

Ejemplo 3.5: Dame el **identificador y** la **región de** los **clientes**

Consulta a procesar	identificador	y	región	de	clientes
Valores numéricos	1		2		2
Se Cambia	1	2	1	3	2
Prioridad		2		3	

Ejemplo 3.6: Muéstrame la **fecha de nacimiento y** el **nombre del empleado**
En este ejemplo se pueden observar los conjuntos de columnas y tablas obtenidos con este tratamiento. Esta consulta incluye los casos 2 [1 prep 1], 3 [1 conj 1] y 4 [1 prep 2].

Consulta a procesar	fecha	de	nacimiento	y	nombre	de	empleado
Penalizaciones	1	1	1	2	1	3	2
Prioridad		1		2		3	

Los conjuntos para cada sustantivo son los siguientes:

Sustantivo	fecha	de	nacimiento
Campos	Employees.BirthDate Employees.HireDate Orders.OrderDate Orders.RequiredDate Orders.ShippedDate		Employees.BirthDate
Tablas	∅		∅

Sustantivo	nombre	empleado
Campos	Categories.CategoryName Customers. CompanyName Customers. ContactName Employees. FirstName Employees. LastName Orders. ShipName Products.ProductName Shippers.CompanyName Suppliers.CompanyName Suppliers.ContactName	Employees.EmployeeID EmployeeTerritories.EmployeeID Orders.EmployeeID
Tablas	∅	Employees

Al aplicar las operaciones sobre los conjuntos en el orden previamente establecido tenemos:

Operación 1: fecha de nacimiento

Operación 1	fecha	de	nacimiento
Intersección	Employees.BirthDate Employees.HireDate Orders.OrderDate Orders.RequiredDate Orders.ShippedDate		Employees.BirthDate
Resultado	Employees.BirthDate		

Operación 2: (fecha de nacimiento) y nombre

Operación 1	fecha de nacimiento	y	nombre
Unión	Employees.BirthDate		Categories.CategoryName Customers. CompanyName Customers. ContactName Employees. FirstName Employees. LastName Orders. ShipName Products.ProductName Shippers.CompanyName Suppliers.CompanyName Suppliers.ContactName
Resultado	Employees.BirthDate Categories.CategoryName Customers. CompanyName Customers. ContactName		

	Employees. FirstName Employees. LastName Orders. ShipName Products.ProductName Shippers.CompanyName Suppliers.CompanyName Suppliers.ContactName

Operación 3: (fecha de nacimiento y nombre) de empleados

Operación 1	fecha de nacimiento y nombre	de	empleado
Intersección	Employees.BirthDate Categories.CategoryName Customers. CompanyName Customers. ContactName Employees. FirstName Employees. LastName Orders. ShipName Products.ProductName Shippers.CompanyName Suppliers.CompanyName Suppliers.ContactName		**Columnas:** Employees.EmployeeID EmployeeTerritories. EmployeeID Orders.EmployeeID **Tablas:** Employees
Resultado final	**Columnas:** Employees.BirthDate Employees. FirstName Employees. LastName **Tablas:** Employees		

Mediante este tratamiento es posible determinar las columnas requeridas en la consulta, eliminando otras columnas con las que se asocian los sustantivos, las cuales son innecesarias.

Continuando con el proceso para determinar las tablas y columnas y después de llevar a cabo el tratamiento de la conjunción *y*, la preposición *de* y los sustantivos presentes en la consulta de realizan los siguientes pasos:

1. Se realiza un proceso de marcación permanente de las columnas con objeto de comprobar que todas las columnas marcadas temporalmente en el tratamiento son necesarias para la traducción.
2. Si la consulta no contiene la preposición *de* o la conjunción *y*, se marcan temporalmente todas las columnas y tablas referidas por los sustantivos en la oración. Este proceso generalmente arroja un conjunto grande de tablas y columnas marcadas, por tal motivo éstas pasan por un proceso, que consiste en marcar permanentemente las columnas que corresponden a tablas referidas a través de un sustantivo en la consulta.
3. Si una palabra (usualmente un sustantivo) en la frase select se refiere solamente a una tabla (y no a alguna otra tabla o columna) entonces la tabla es marcada permanentemente y asociada a esta palabra. Si ésta se refiere a varias tablas, las tablas encontradas son marcadas temporalmente.
4. Posteriormente el nuevo conjunto de tablas y columnas marcadas pasan al tratamiento de la parte where de la oración. Una heurística relaciona cada valor presente con un sustantivo que haga referencia a una columna dentro de la frase where de la consulta a través de un operador. Si la columna correspondiente a cada valor no tiene un sustantivo que le haga referencia en la frase where, se realiza una búsqueda del valor en las tablas marcadas de la base de datos. De no encontrar dicho valor en ellas, se busca en el resto de la base de datos; para tal efecto, de acuerdo con la condición A5 solamente es necesario buscar valores de tipo cadena en la base de datos,

debido a que los valores numéricos y tipo fecha deben tener explícitamente un sustantivo asociado que se refiera a esas columnas. Este tratamiento obtiene un conjunto de columnas y tablas marcadas permanentemente, el cual corresponde a la parte de la condición de la consulta. Si aún existen tablas temporalmente marcadas correspondientes a la frase select, se marcarán permanentemente únicamente aquellas que hayan sido marcadas definitivamente para la frase where.

Fase 3: Construcción del grafo relacional: El proceso para construir el grafo relacional es el siguiente:

1. Considerando la condición C1, se construye un grafo no dirigido con base en el modelo relacional de la base de datos. Cada uno de sus nodos representa una tabla, y cada uno de sus arcos representa una relación referencial entre tablas. Se suponen relaciones binarias (que involucran dos tablas); esto no es una limitación seria, puesto que una relación que implica más de dos tablas (*T*1, *T*2, ..., *Tn*) puede ser sustituida siempre por una tabla auxiliar *Ta* con un enlace binario con las tablas *T*1, *T*2, ..., *Tn*.
2. Los nodos correspondientes a las tablas permanentemente marcadas en la Fase 2 son marcados. Posteriormente, por cada condición de selección simple en la frase where que implique una columna de una tabla (por ejemplo: *con órdenes para el barco Mercury*), el nodo correspondiente a la tabla se etiqueta con su correspondiente condición de selección simple. Finalmente, cada nodo es etiquetado con las columnas (de la tabla correspondiente) referidas en la frase select.
3. Por cada condición de selección simple en la frase where que involucre columnas de dos tablas, el arco incidente a los nodos que representan las tablas es marcado; si no existe un arco incidente, éste es añadido al grafo y marcado. Cada arco marcado en este paso es etiquetado con su correspondiente condición de selección simple.
4. Si todas las condiciones de selección son enunciadas explícitamente en la

consulta, entonces el subgrafo consistente de todos los nodos y arcos marcados debe ser un grafo conectado. A partir de este subgrafo se obtiene la traducción a una expresión en SQL.

5. Un subgrafo desconectado significa que existen condiciones de selección implícitas en la consulta o que la consulta está formulada incorrectamente. En el primer caso, la ILNBD tiene que determinar las condiciones de selección implícitas. Para esto se utiliza una búsqueda preferente por amplitud, la cual está basada en la siguiente suposición: todas las condiciones de selección implícitas se refieren a las reuniones naturales que involucran las tablas y las columnas que participan en una relación referencial. Por consiguiente, se construye un subgrafo conectado al agregar arcos no marcados al grafo desconectado, tal que el número de arcos no marcados sea el mínimo. De este subgrafo la traducción a una expresión en SQL es directa. Si no se puede construir un subgrafo conectado, entonces la consulta es reportada como incorrecta.

4 VALIDACIÓN DE LA METODOLOGÍA PROPUESTA

En este capítulo se ilustra el funcionamiento de la técnica de traducción mediante la resolución de ejemplos, y se presentan los experimentos efectuados en la investigación. Este capítulo detalla el proceso de traducción a partir del preprocesamiento de la consulta, la separación de las frases select y where, la selección de tablas y columnas, la construcción del grafo, para finalizar con la obtención de la consulta en SQL. Además se describen los tipos de experimentos aplicados a la interfaz.

4.1 CASOS DE PRUEBA

La realización de casos de prueba permite observar cómo se aplican los distintos procesos que generan la traducción al lenguaje SQL. En el siguiente apartado se ejemplifica el funcionamiento del algoritmo principal de la técnica de traducción.

4.1.1 PRUEBA GENERAL DEL ALGORITMO PROPUESTO

La resolución de ejemplos permite observar la manera en la que se lleva a cabo el proceso de traducción. A continuación se muestra la traducción de dos ejemplos de consultas en lenguaje natural:

Ejemplo 4.1: Dame el **nombre del empleado** con **fecha de contratación 23/07/1996**

La siguiente tabla contiene la salida del preprocesamiento de esta consulta:

Consulta en lenguaje natural			
Palabra	**Lema**	**Categoría**	**Valor**
dame	dar	VLII1P	0
el	el	ARTDMS	0
nombre	nombre	s	1
del	del	PDEL	0
empleado	empleado	s	2
con	con	PREP	0
fecha	fecha	s	1
de	de	PREP	0
contratacion	contratacion	s	1
23/07/1996	23/07/1996	a	0

La técnica de traducción de esta interfaz está basada en el procesamiento de los sustantivos, la conjunción *y,* la preposición *de* y los valores numéricos o alfanuméricos dados en la consulta, por lo que considerando únicamente estos elementos, la consulta queda de esta manera:

nombre del empleado fecha de contratación 23/07/1996

Ahora es necesario determinar las frases select y where de esta consulta. Se utiliza el primer criterio para separar la consulta, el cual verifica si existen dos sustantivos adyacentes: *empleado* y *fecha* en este caso. El resultado se puede observar la siguiente tabla:

Separación de la consulta	
Frase select:	nombre de empleado
Frase where:	fecha de contratación 23/07/1996

El siguiente paso es realizar el tratamiento de la preposición *de* y de la conjunción *y* para la frase select de esta consulta: *nombre de empleado*. En primer lugar, del diccionario de dominio, se obtienen los conjuntos de tablas y columnas asociadas con ambos sustantivos.

Sustantivo	nombre	de	Empleado
Columnas	Categories.CategoryName Customers.CompanyName Customers.ContactName Employees.FirstName Employees.LastName Shippers.ShipName Products.ProductName Suppliers.CompanyName Suppliers.ContactName		Employees. EmployeeID EmployeesTerritories.EmployeeID
Tablas	∅		Employees

Posteriormente se aplica una operación de intersección entre ellos.

Operación: intersección	nombre de empleado
Intersección columnas	∅
Intersección columnas nombre y tablas empleado	Employees.FirstName Employees.LastName
Intersección columnas empleado y tablas nombre	∅
Unión tablas	Employees

El resultado se puede ver en la siguiente tabla:

Análisis de la frase select		
frase select	**columnas referidas**	**tablas referidas**
nombre de empleado	Employees.FirstName Employees.LastName	Employees

La frase where también requiere de este tratamiento para obtener la columna correspondiente a *fecha de contratación*. Al conjunto relacionado con estos sustantivos se le aplica una operación de intersección:

Sustantivo	fecha	de	contratación
Columnas	Employees.HireDate Employees.BirthDate Orders.OrderDate Orders.RequiredDate Orders.ShippedDate		Employees.HireDate
Tablas	∅		∅

Operación: intersección	fecha de contratación
Intersección columnas	Employee.HireDate
Intersección columnas fecha y tablas contratación	∅
Intersección columnas contratación y tablas fecha	∅
Unión tablas	∅

El resultado de la selección de las tablas y columnas para la frase where, queda de la siguiente manera:

Análisis de la frase where		
frase where	**columnas referidas**	**tablas referidas**
fecha de contratación	Employees.HireDate	Ninguna
23/07/1996 (valor tipo fecha)		

Con la información obtenida en el análisis anterior, se procede a la construcción del grafo de la consulta. Para tal efecto se marcan las tablas encontradas y se etiquetan con sus correspondientes columnas marcadas. Para la condición se agrega además el operador y el valor correspondiente. El grafo desconectado se puede ver en la figura 4.1:

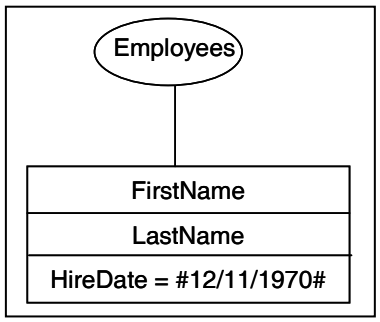


Figura 4.1 Grafo final de la consulta del ejemplo 4.1

Como la consulta es a una tabla no es necesario conectar el grafo, por consiguiente el paso restante es la traducción a SQL con base en el grafo resultante.

```
SELECT   Employees.FirstName,
         Employees.LastName
FROM     Employees
WHERE    Employees.HireDate = #1996/07/23#
```

Este ejemplo fue resuelto satisfactoriamente por el algoritmo de traducción. El siguiente ejemplo requiere de varias tablas, por lo que es un caso de relaciones implícitas.

Ejemplo 4.2: Dame el nombre de los clientes que compraron el producto "Zaanse koeken"

La fase del preprocesamiento de esta consulta arroja la siguiente tabla.

Consulta en lenguaje natural			
Palabra	**Lema**	**Categoría**	**Valor**
dame	dar	VLII1P	0
el	el	ARTDMS	0
nombre	nombre	s	1
de	de	PREP	0
los	el	ARTDMP	0
clientes	cliente	s	2
que	que	CQUE	0
compraron	comprar	V	0
el	el	ARTDMS	0
producto	producto	s	2
Zaanse koeken	Zaanse koeken	a	0

El sistema solamente utiliza los sustantivos (nombre, cliente y producto), la preposición de y el valor “Zaanse koeken”, quedando la consulta de la siguiente manera:

nombre de cliente producto “Zaanse koeken”

El siguiente paso es separar la consulta en sus componentes principales: la frase select y la frase where. Para esta consulta se aplica el primer criterio, el cual se

refiere a la presencia de dos sustantivos adyacentes: cliente y producto. Por consiguiente las frases select y where quedan formadas de la siguiente manera:

Separación de la consulta	
Frase select:	nombre de cliente
Frase where:	producto “Zaanse koeken”

Una vez que se tienen ambas frases, se procede al tratamiento de la preposición *de* y de la conjunción *y*. En la frase select la preposición *de* aparece uniendo los sustantivos nombre y cliente. Inicialmente se cuenta con los siguientes conjuntos de columnas y tablas asociadas a los sustantivos nombre y cliente (esta información se obtiene consultando el diccionario de dominio):

Sustantivo	nombre	de	cliente
Columnas	Categories.CategoryName Customers.CompanyName Customers.ContactName Employees.LastName Employees.FirstName Shippers.ShipName Products.ProductName Suppliers.CompanyName Suppliers.ContactName		CustomerCustomerDemo.CustomerID CustomerCustomerDemo.CustomerTypeID CustomerDemographics.CustomerTypeID CustomerDemographics.CustomerDesc Customers.CustomerID Orders.CustomerID CustomerCustomerDemo.CustomerID CustomerCustomerDemo.CustomerTypeID
Tablas	∅		Customers

Al aplicar las operaciones sobre los conjuntos, tenemos:

Operación: intersección	nombre de cliente
Intersección columnas	∅
Intersección columnas nombre y tablas cliente	Customers.CompanyName Customers.ContactName
Intersección columnas cliente y tablas nombre	∅
Unión tablas	Customers

El resultado se puede apreciar en la siguiente tabla.

Análisis de la frase select		
frase select	**Columnas referidas**	**Tablas referidas**
nombre de clientes	Customers.CompanyName Customers.ContactName	Customers

La selección de tablas y columnas para la frase where requiere de obtener la columna correspondiente al valor “Zaanse koeken”, la cual se obtiene buscándola en las tablas Customers y Products de la base de datos, ya que son las tablas obtenidas hasta el momento. El resultado se puede ver en esta tabla:

Análisis de la frase where		
frase where	**columnas referidas**	**tablas referidas**
Productos	Product.ProductID	Products
'Zaanse koeken' (valor)	Product.ProductName	Products

Los conjuntos obtenidos son las tablas y las columnas marcadas, las cuales se utilizarán para la construcción del grafo. En primer lugar se marcan las tablas en el grafo, y se etiquetan con sus respectivas columnas marcadas. En caso de tener una condición, se marcan la columna con el valor y operador correspondientes. El grafo de la figura 4.2 es el grafo desconectado inicial.

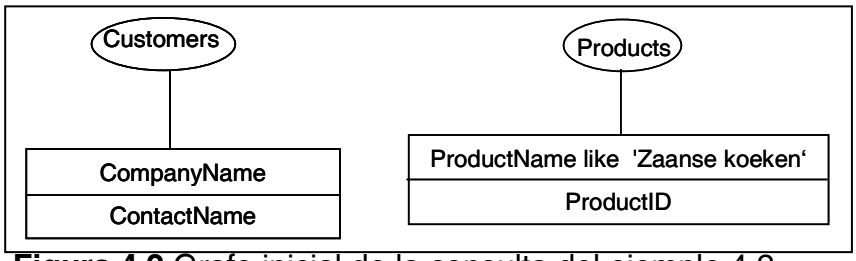


Figura 4.2 Grafo inicial de la consulta del ejemplo 4.2

El siguiente paso es determinar las reuniones implícitas a partir de una búsqueda preferente por amplitud. El grafo final resultante (sin incluir las columnas marcadas) se puede observar en la figura 4.3.

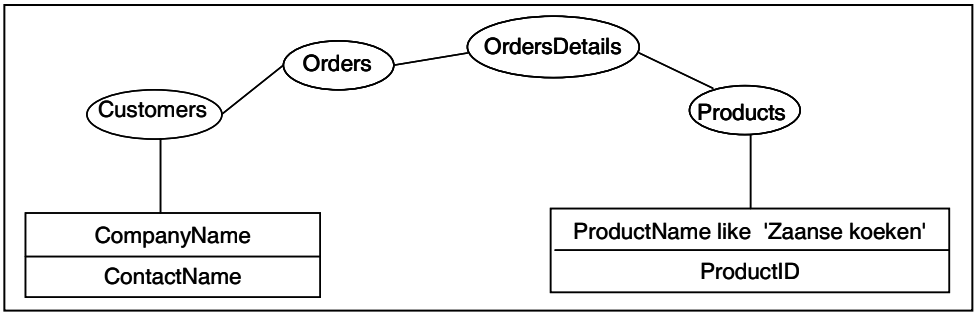


Figura 4.3 Grafo final de la consulta del ejemplo 4.2

Como se puede observar fue necesario agregar al grafo las tablas Orders y OrdersDetails. Para finalizar, se obtiene la consulta en SQL basándose en la información del grafo relacional.

La traducción a SQL:

```
SELECT   Customers.CompanyName,
         Customers.ContactName, Products.ProductID,
         Products.ProductName
FROM     Customers, Products, Orders, OrderDetails
WHERE    ((Products.ProductName                LIKE
```

```
AND     'Zaanse_koeken'))
AND     Customers.CustomerID = Orders.CustomerID
AND     OrderDetails.OrderID = Orders.OrderID
        Products.ProductID = OrderDetails.ProductID
```

4.2 EXPERIMENTACIÓN

Para obtener los casos de prueba se formaron dos grupos de 50 alumnos cada uno de nivel licenciatura del Instituto Tecnológico de Ciudad Madero para que formularán un conjunto de consultas a las bases de datos Northwind y Pubs, las cuales han sido usadas por otras interfaces [51, 52], proporcionándoles únicamente los esquemas de éstas.

Como resultado se obtuvieron dos corpus, uno de 198 consultas distintas para la base de datos Northwind y otro de 70 consultas distintas para la base de datos Pubs (ver Anexo B).

Estas consultas se tipificación con base en la clasificación propuesta en la tabla 3.4 de este documento. Las tablas 4.1 y 4.2 contienen el porcentaje y el número de consultas que pertenecen a cada tipo para los corpus de las bases de datos Northwind y Pubs respectivamente.

Tabla 4.1 Corpus para la base de datos Northwind

1	4	2.02%
2	2	1.01%
3	2	1.01%
4	0	--

5	0	--
6	0	--
7	0	--
8	1	0.50%
9	27	13.64%
10	55	27.80%
11	21	10.61%
12	57	28.80%
13	0	--
14	1	0.50%
15	0	--
16	0	--
17	28	14.14%
Total	198	100%

Tabla 4.2 Corpus para la base de datos Pubs

1	3	4.29%
2	2	2.86%
3	1	1.43%
4	1	1.43%
5	0	--
6	0	--
7	0	--
8	0	--
9	2	2.86%
10	41	58.57%
11	4	5.71%

12	3	4.29%
13	0	--
14	4	5.71%
15	0	--
16	1	1.43%
17	8	11.42%
Total	70	100%

De esta tipificación se puede observar que los tipos 5 al 8 y del 13 al 16 tienen una frecuencia de participación muy baja, con lo que contribuye a un porcentaje mínimo del total de las consultas. Por lo que para propósitos de éste estudio se decidió delimitar los tipos de consulta de la manera siguiente:

Tipo 1: Información completamente explícita con o sin condición.
Tipo 2: Tablas explícitas y columnas implícitas con o sin condición.
Tipo 3: Columnas explícitas y tablas implícitas con o sin condición.
Tipo 4: Columnas y tablas implícitas con o sin condición.
Tipo 5: Con funciones de SQL.
Tipo 6: Sin respuesta en la base de datos.

4.2.1 EXPERIMENTO 1

Objetivo. Obtener la frecuencia de las partes invariables de una oración que se presentan en las consultas que integran los corpus obtenidos en este trabajo.

.

Procedimiento. Se contabilizó la frecuencia con la que aparecen las partes invariables de la oración.

Resultados. Se obtuvieron las 15 partes invariables de la oración que se presentan con mayor frecuencia. En la tabla 4.3 se muestran los resultados mostrando la clasificación de las palabras y la figura 4.4 nos permite visualizar la tendencia.

Tabla 4.3 Frecuencia de las Partes Invariables en el Corpus

Clase	Parte Invariable	Frecuencia
Preposición propia	de	385
Conjunción coordinante copulativa	que	117
Preposición propia	en	71
Conjunción coordinante copulativa	y, e	68
Preposición propia	con	33
Preposición propia	a	30
Adverbio de cantidad	más	13
Conjunción subordinante final, Preposición propia	para	13
Conjunción subordinante espacial, Adverbio de lugar	donde	12
Preposición propia	entre	9
Preposición propia	por	9
Adverbio de negación	no	8
Conjunción coordinante disyuntiva	o	8
Conjunción subordinante final	a que	6

Conjunción subordinante de modo y condicional, Adverbio de modo	como	4
Conjunción subordinante comparativa	más...de	4

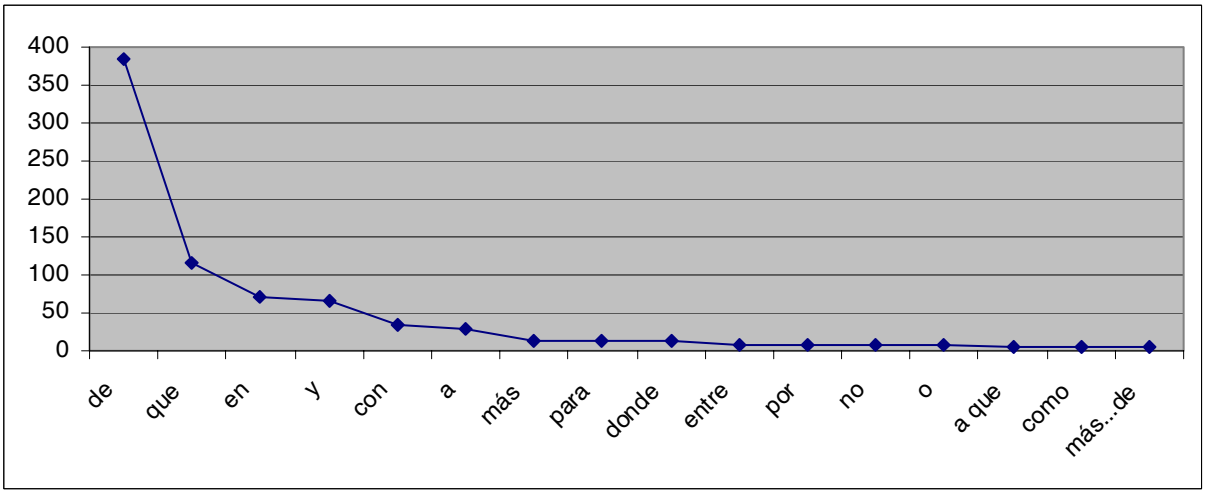


Figura 4.4 Frecuencia de las Partes Invariables en el Corpus

4.2.2 EXPERIMENTO 2

Objetivo. Demostrar que las técnicas usadas aplican a varias bases de datos y son sencillas de utilizar.

Procedimiento. Se ejecuta el módulo que construye el diccionario de dominio para las bases de datos Northwind y Pubs.

Resultados. La configuración de cada base de datos requirió un tiempo menor a 10 minutos. Además este proceso puede ser realizado por cualquier persona, ya que se realiza automáticamente sólo con introducir el nombre de la base de datos.

4.2.3 EXPERIMENTO 3

Objetivo. Obtener el porcentaje de éxito en la traducción de las consultas de los corpus obtenidos para las bases de datos Northwind y Pubs, sin utilizar el tratamiento de la información implícita de las consultas.

Procedimiento. Dado que el tipo de prueba más utilizado es el porcentaje de éxito, se optó por utilizar éste, para lo cual se introdujeron las consultas de ambos corpus en la interfaz implementada, eliminando el módulo para la obtención de las relaciones del grafo semántico.

Resultados. Los resultados obtenidos en este experimento se encuentran concentrados en las tablas 4.4 y 4.5 para las bases de datos Northwind y Pubs, respectivamente.

Tabla 4.4 Resultados del experimento 3 para el corpus de la base de datos Northwind

Consultas a la BD Northwind	tipo 1	tipo 2	tipo 3	tipo 4	tipo 5	tipo 6	total	%	%éxito
Contestadas correctamente	26	33	3	21	0	0	83	42	74
Contestadas con información extra	1	19	20	23	0	0	63	32	
Contestadas incorrectamente	4	5	1	9	23	5	47	23	26
No contestadas	0	0	0	5	0	0	5	3	
Total	31	57	24	58	23	5	198	100	100

Tabla 4.5 Resultados del experimento 3 para el corpus de la base de datos Pubs.

Consultas a la BD Pubs	tipo 1	tipo 2	tipo 3	tipo 4	tipo 5	tipo 6	total	%	%éxito
Contestadas correctamente	7	15	2	0	0	0	24	34	44
Contestadas con información extra	0	7	0	0	0	0	7	10	
Contestadas Incorrectamente	0	7	6	13	10	1	37	53	56
No contestadas	0	0	0	0	1	1	2	3	
Total	7	29	8	13	11	2	70	100	100

4.2.4 EXPERIMENTO 4

Objetivo. Obtener el porcentaje de éxito en la traducción de las consultas de los corpus obtenidos para las bases de datos Northwind y Pubs. Utilizando el tratamiento de la información implícita de las consultas.

Procedimiento. Utilizando las mismas condiciones del experimento 3, se introdujeron las consultas de ambos corpus en la interfaz implementada utilizando el módulo para la obtención de las relaciones del grafo semántico.

Resultados. Los resultados para este experimento están concentrados en tabla 4.6 para la base de datos Northwind y en la tabla 4.7 para Pubs.

Tabla 4.6 Resultados del experimento 4 para el corpus de la base de datos Northwind

Consultas a la BD Northwind	tipo 1	tipo 2	tipo 3	tipo 4	tipo 5	tipo 6	total	%	%éxito
Contestadas correctamente	31	57	19	49	0	0	156	79	84
Contestadas con información extra	0	0	5	5	0	0	10	5	
Contestadas incorrectamente	0	0	0	1	23	5	29	15	16
No contestadas	0	0	0	3	0	0	3	1	
Total	31	57	24	58	23	5	198	100	100

Tabla 4.7 Resultados del experimento 4 para el corpus de la base de datos Pubs.

Consultas a la BD Pubs	tipo 1	tipo 2	tipo 3	tipo 4	tipo 5	tipo 6	total	%	%éxito
Contestadas correctamente	7	29	8	12	0	0	56	80	80
Contestadas con información extra	0	0	0	0	0	0	0	0	
Contestadas incorrectamente	0	0	0	1	10	1	12	17	20
No contestadas	0	0	0	0	1	1	2	3	
Total	7	29	8	13	11	2	70	100	100

4.2.5 EXPERIMENTO 5

Objetivo. Obtener el porcentaje de éxito para consultas presentadas de forma interrogativa e imperativa, después de haber entrenado a un grupo de alumnos.

Procedimiento. A un grupo de 10 alumnos se les da un entrenamiento para realizar consultas. Este entrenamiento consistió en darles las condiciones para la

formulación de consultas expresadas en el apartado 3.8 de este documento y algunos ejemplos. Después de esto se les entregó los corpus obtenidos y se le pidió que elaboraran las consultas en forma interrogativa e imperativa de acuerdo a las condiciones. Se seleccionaron 50 consultas conservando los porcentajes obtenidos en la tipificación, las cuales se introdujeron en la interfaz implementada.

Resultados. Los resultados para este experimento se pueden encontrar en los Anexos C y D, además se presentan los porcentajes obtenidos en la tabla 4.8 para la base de datos Northwind y en la tabla 4.9 para Pubs.

Tabla 4.8 Resultados del experimento 5 para el corpus de la base de datos Northwind

Consultas	Correctas		Incorrectas	
Imperativas	45	90%	5	10%
Interrogativas	43	86%	7	14%

Tabla 4.9 Resultados del experimento 5 para el corpus de la base de datos Pubs

Consultas	Correctas		Incorrectas	
Imperativas	41	82%	9	18%
Interrogativas	40	80%	10	20%

Para ambos corpus se obtuvo un porcentaje de éxito mayor en consultas imperativas.

4.2.6 EXPERIMENTO 6

Objetivo. Obtener el porcentaje de éxito en la traducción de consultas presentadas de forma interrogativa e imperativa, de los corpus obtenidos para las

bases de datos Northwind y Pubs, eliminando el módulo para el tratamiento de la preposición *de* y la conjunción *y.*

Procedimiento. Se introdujeron las 50 consultas obtenidas en el experimento 5 a la interfaz implementada, eliminando el módulo para el tratamiento de la preposición *de* y la conjunción *y*.

Resultados. En este experimento la mayoría de las consultas fue contestada con la información correcta, pero se recuperaron columnas extras. Esto indica que el tratamiento de la preposición *de* y la conjunción *y* ayuda al análisis semántico del traductor para obtener el resultado correcto.

4.2.7 ANÁLISIS DE RESULTADOS

Se puede observar que las técnicas utilizadas en esta interfaz permiten una fácil configuración para diferentes bases de datos. Además los experimentos realizados muestran que al eliminar módulos como el del tratamiento de la información implícita y el del tratamiento de la preposición *de* y la conjunción *y*, los porcentajes de éxito son menores o la información proporcionada por la interfaz contiene columnas extra.

Se puede apreciar también que con un entrenamiento al usuario consistente de reglas simples y ejemplos claros el porcentaje de éxito aumenta.

Los resultados obtenidos en este trabajo dan evidencia de una mejoría en el porcentaje de consultas traducidas en comparación con otras interfaces. Además se muestra que el uso de técnicas que sean independientes de la información contenida en la base de datos, permite la portabilidad del dominio.

5 CONCLUSIONES Y TRABAJOS FUTUROS

En este capítulo se presentan las conclusiones obtenidas durante el desarrollo de esta investigación, la cual aborda el problema de traducción de consultas en lenguaje natural a una base de datos. También enlista las publicaciones y trabajos derivados de esta investigación. Además se proponen trabajos futuros, con el fin de mejorar la precisión en la traducción, y aumentar el tipo de consultas que pueda realizar el usuario.

5.1 CONCLUSIONES

El objetivo de esta investigación fue desarrollar un módulo traductor de lenguaje natural español a SQL para un sistema de consultas a una base de datos, que mejore el desempeño o porcentaje de consultas contestadas correctamente con respecto a las técnicas hasta ahora utilizadas, y que permita al módulo de traducción ser independiente de la base de datos.

En la revisión de trabajos relacionados se observa el uso de diferentes técnicas que abordan la traducción de consultas, sin embargo también se aprecia la

dificultad de configuración, la cual es un elemento importante en la portabilidad del dominio. Este trabajo muestra que es factible incorporar algunos elementos de diseño de la base de datos y de análisis semántico de la consulta para resolver el problema de configuración de una mejor manera.

La siguiente lista enumera en orden de importancia las contribuciones de la presente investigación:

1. El uso de las operaciones de unión e intersección de la teoría de conjuntos para dar tratamiento a la preposición "de" y a la conjunción "y" las cuales conforman las partes invariables de la oración en el idioma español, para mejorar la calidad de la respuesta a la consulta de usuario. Ejemplos de este tratamiento son mostrados en la sección 3.8.2. En la investigación desarrollada no se encontró evidencia del uso de la teoría de conjuntos para resolver problemáticas relacionadas con la semántica de las consultas a bases de datos, por lo que se considera una aportación de este trabajo.
2. Se propuso la generación de un grafo para modelar las relaciones entre las tablas de la base de datos. Este grafo utiliza los metadatos de la base de datos para encontrar las relaciones entre las tablas y poder ser generado automáticamente.
3. A partir de las condiciones de diseño de la base de datos y de formulación de las consultas se proponen una serie de criterios para identificar las frases SELECT y WHERE en la consulta del usuario, las cuales se describen en la sección 3.8.2.
4. Se propuso la utilización de las descripciones de las columnas y tablas de los metadatos para obtener la información semántica necesaria (ver sección 3.1). En los trabajos revisados se observó que ninguna técnica utiliza los metadatos con este fin.
5. Para obtener la consulta etiquetada que requiere el traductor, se implementó un diccionario de sinónimos general, un módulo de software

para obtener los metadatos de la base de datos, un diccionario de dominio, y un preprocesador que utiliza los módulos anteriores, este desarrollo tecnológico servirá de infraestructura para futuros proyectos y es descrito en las secciones 3.1, 3.2 y 3.3.

6. Para generar los casos de prueba se obtuvieron dos corpus realizados en condiciones similares por dos grupos distintos de personas, los cuales se tipificaron y se pudieron apreciar resultados semejantes (ver tablas 4.1 y 4.2). En el análisis a la tipificación de consultas, se observó la problemática que se presenta en las consultas con información implícita.

Estas contribuciones hacen posible que el traductor tenga una mayor portabilidad de dominio, debido a que la técnica de traducción propuesta se mantiene independiente de la información contenida en la base de datos. Además, de los resultados experimentales obtenidos en esta tesis, se puede mencionar que el traductor responde un mayor porcentaje de consultas, por lo que las hipotesis planteadas en esta tesis fueron alcanzadas.
Sin embargo, quiero mencionar que el porcentaje de consultas siempre será cuestionable, debido principalmente a la falta de benchmarks tanto para bases de datos como para tipos de consultas, además, los desarrolladores de interfaces independientes del dominio no muestran lo difícil que es configurar su interfaz para que alcance porcentajes superiores al 80%, debo reconocer la seriedad de los desarrolladores de Precise al abordar esta problemática en su trabajo[62].

5.1.1 PUBLICACIONES RELACIONADAS AL TEMA

De esta investigación doctoral se generaron una serie de artículos que han sido publicados en revistas y congresos nacionales e internacionales, los cuales se enlistan a continuación:

*"Spanish Natural Language Interface for a Relational Database Querying System". **Lecture Notes in Artificial Intelligence** (Text, Speech and Dialogue), Vol. 2448, ISSN 0302-9743, Springer-Verlag, Sep. 2002, pp. 123-130.*

*"A Domain Independent Natural Language Interface to Databases Capable of Processing Complex Queries". **Lecture Notes in Artificial Intelligence** (MICAI 2005: Advances in Artificial Intelligence), ISSN 0302-9743, Springer-Verlag, 2005.*

"Implementación de un analizador sintáctico para una interfaz de lenguaje natural". Rodolfo A. Pazos, Juan Javier González Barbosa. XII Congreso Interuniversitario de Electrónica, Computación, Eléctrica (CIECE 2002). Coahuila, México. 2002

"Implementación de un analizador gramatical para el idioma español". Rodolfo A. Pazos, Juan Javier González Barbosa. XII Congreso Interuniversitario de Electrónica, Computación, Eléctrica (CIECE 2002). Coahuila, México. 2002

"Diseño y construcción de una interfaz de lenguaje natural en español". Rodolfo A. Pazos, Juan Javier González Barbosa. 14º Encuentro Nacional de Investigación Científica y Tecnológica del Golfo de México. Veracruz, México. 2002.

"Técnicas de Traducción utilizadas para consultas a bases de datos en lenguaje natural". Juan Javier González Barbosa, Erika Alarcón Ruiz. XIII Congreso Interuniversitario de Electrónica, Computación, Eléctrica (CIECE 2003). Morelos, México. 2003

"Técnicas de traducción de consultas en Lenguaje natural a Lenguaje formal". Rodolfo A. Pazos, Juan Javier González Barbosa. 10mo Congreso Internacional de Investigación en Ciencias computacionales (CIICC 2003). Morelos, México, 2003.

5.1.2 TRABAJOS DERIVADOS DE ESTA TESIS

Esta tesis doctoral también ha generado trabajos de tesis de licenciatura y maestría como apoyo a la formación de recursos humanos al motivar a estudiantes para seguir el camino de la investigación. A continuación se describen brevemente las tesis que fueron desarrolladas bajo mi dirección y asesoría por alumnos del Instituto Tecnológico de Ciudad Madero.

"Implementación de un Analizador Gramatical del Lenguaje Español", tesis de maestría desarrollada por la alumna Ana Patricia Domínguez Sánchez, agosto de 2002. En este trabajo de investigación se realizó una revisión de los proyectos de interfaces de lenguaje natural para consultas a bases de datos y las técnicas de traducción utilizadas y se implementó un analizador gramatical para ser usado en una ILNBD de dominio específico. Con la información obtenida se pudo determinar la importancia de la independencia del dominio de una ILNBD y la dificultad que ésta implica. Esta tesis aportó el estado del arte para el proyecto doctoral.

"Diseño de la Interfaz de Lenguaje Natural para consultas a bases de datos", tesis de licenciatura realizada por la alumna Erika Alarcón Ruiz, agosto de 2002. En este trabajo se presenta el diseño y la implementación de una interfaz de lenguaje natural para consultas a la base de datos de alumnos del ITCM vía Internet, que realiza la interpretación de una consulta en lenguaje natural utilizando el analizador gramatical desarrollado por Ana Patricia Domínguez. Este proyecto permitió desarrollar una ILNBD para ser usada en Internet, pero no aportó ningún elemento a mi tesis doctoral.

"Desarrollo de un modulo de predicados lógicos para consultas a la base de datos de alumnos del ITCM", tesis de licenciatura desarrollada por el alumno Alejandro Mendoza Mejia, agosto de 2002. En este proyecto se desarrolló el analizador semántico usando predicados lógicos para la ILNBD implementada por Erika

Alarcón Ruíz. Este analizador está diseñado para determinar el significado de las consultas más comunes a la base de datos de alumnos del ITCM.

En las tesis de licenciatura de Erika Alarcón Ruíz y de Alejandro Mendoza Mejía, en conjunto se desarrolló una ILNBD dependiente del dominio. La aportación de ambos proyectos de licenciatura al proyecto doctoral, fue clarificar las desventajas del uso de la técnica de predicados lógicos y mostrar la dificultad para el logro de la independencia del dominio.

"Consultas a una base de datos en lenguaje natural utilizando múltiples tablas", tesis de licenciatura realizada por las alumnas Maria del Rosario Smith Salas y Kaarem Castellanos Jongitud, diciembre de 2003. Este proyecto es una continuación de los trabajos de licenciatura anteriores, en los cuales se desarrolló una ILNBD utilizando la técnica de predicados lógicos para acceder a la información de una sola tabla. La presente investigación desarrolló predicados lógicos que involucran el uso de múltiples tablas.

"Diseño de una técnica basada en grafos semánticos para la traducción de consultas de lenguaje natural a lenguaje formal", tesis de maestría realizada por la alumna Erika Alarcón Ruiz, agosto de 2004. Este trabajo presenta una arquitectura para la traducción de consultas en español a SQL sin la necesidad de tediosas configuraciones. La independencia del dominio es el principal problema a resolver. En este proyecto se diseñó una técnica de traducción para consultas explícitas utilizando un grafo semántico que organiza la información de cada palabra de la consulta según el contexto de la base de datos que se está utilizando. El diseño muestra el tratamiento de la preposición "*de"* y la conjunción "*y"* utilizando la teoría de conjuntos, lo cual ayuda al análisis semántico del traductor. En esta tesis queda plasmado parte del diseño de mi propuesta doctoral.

"Construcción de un preprocesador de consultas en lenguaje natural a una base de datos", tesis de maestría desarrollada por el alumno Alejandro Mendoza Mejia,

noviembre de 2004. En este trabajo se implementó un preprocesador de consultas cuyo objetivo es construir un diccionario de dominio de una manera fácil, para lo cual fue necesario construir un diccionario de sinónimos general para que pudiera ser utilizado por cualquier base de datos y un diccionario de metadatos que es generado de manera automática de acuerdo con la base de datos que se está utilizando. El preprocesador también analiza la consulta hecha por el usuario y la etiqueta con información léxica, sintáctica y semántica. El proyecto doctoral utiliza la arquitectura y el código del prepocesador desarrollado por esta tesis.

"Representación de una consulta de Lenguaje Natural a través de un grafo", tesis de licenciatura realizada por el alumno José Francisco Delgado Orta, marzo de 2005. En esta tesis se realiza la implementación de un grafo semántico ponderado para representar una consulta explicita. Algunas funciones de código para la construcción del grafo, son utilizadas en el proyecto doctoral.

"Construcción de un Diccionario de Valores para una Interfaz de Lenguaje Natural para bases de datos", proyecto de titulación para licenciatura opción IV, realizado por la alumna Yuridia Torres Romero, mayo de 2005. En este proyecto se construyó un diccionario que almacena los valores de tipo cadena que se encuentran en una base de datos. Este diccionario es útil cuando la información almacenada en la base de datos no se actualiza constantemente; sin embargo, este diccionario no es utilizado en mi proyecto doctoral.

"Diseño de una técnica para la traducción de consultas con información implícita a una base de datos", tesis de maestría realizada por la alumna Myriam Janeth Rodríguez Martínez, octubre de 2005. Esta tesis es resultado de la tipificación de consultas obtenida del corpus. Debo mencionar que la resolución de consultas implícitas no era parte de los objetivos de mi tesis doctoral; sin embargo, dada la importancia y el avance en el diseño de la técnica para resolver este tipo de consultas, se procedió a agregarla al proyecto doctoral. El diseño desarrollado en este trabajo modela el significado de una consulta en lenguaje natural mediante un

grafo semántico de la base de datos. En esta tesis se utiliza el diccionario de valores desarrollado por Yuridia Torres Romero, debido a que los valores tipo cadena aportan información importante en el proceso de traducción.

5.2 LIMITACIONES

Es importante señalar que este trabajo no revisa la información semántica que aportan los verbos, por lo cual no puede responder consultas como "*Dame los nombres de los empleados que nacieron el 15/09/1975*".

Otro tipo de consultas que esta interfaz no puede responder es cuando la consulta está incompleta o no tiene información suficiente como la siguiente: "*Muéstrame el nombre de la compañía*", en este caso el usuario no indica de cuál compañía requiere información (clientes, proveedores o fletadores).

5.3 TRABAJOS FUTUROS

Con el fin de mejorar la respuesta de la consulta, se propone el siguiente trabajo que, debido a la dificultad que implica, podría considerarse como tesis doctoral:

- Diseñar e implementar un módulo que permita interactuar con el usuario con el objeto de mejorar la interpretación de la consulta, utilizando algún tipo de diálogo inteligente con el usuario. Aunque existen algunas interfaces que abordan esta problemática no se da un diálogo como tal, y sólo muestran posibles respuestas para que el usuario elija cuál desea. Esto si bien es practico, no siempre muestra la respuesta correcta.

Para aumentar los tipos de consulta que el usuario pueda introducir a la interfaz se proponen los siguientes trabajos:

- Crear una técnica que relacione elementos de la base de datos con verbos, pero con el fin de mantener la facilidad en la configuración esta técnica tendría que ser automática o semiautomática, lo cual le daría el grado de complejidad suficiente para ser tratada en una tesis de doctorado.

- Diseñar un módulo para el tratamiento de consultas que requieren de funciones de SQL y otro modulo de funciones especiales que permita dar respuesta a algunas de las consultas que la alta gerencia requiera.

- Diseñar un módulo para el manejo de errores, el cual permitiría informar al usuario cuál es la causa del error cuando no se obtiene la traducción de una consulta.

Estos últimos trabajos aunque importantes, requieren sobre todo de elementos de programación, por lo cual considero deberían ser tratados en una tesis de licenciatura o maestría.

10. REFERENCIAS

[1] TA Associates News. http://www.ta.com/news/07-17-00.html

[2] NLUC: Natural Language Understanding Consortium. http://www.nluc.com/ver_sp/Estrategia /Estrategia.htm

[3] InQuira Press Release 07-17-02. http://inquira.com/releases/pr_020717.html

[4] Flores Vázquez Juana María, Ing. Matadamas Hernández José Manuel; Sistema de Interpretación de Texto; Instituto Tecnológico de Celaya, Universidad de Guanajuato, Facultad de Ingeniería Mecánica, Eléctrica y Electrónica, México.

[5] Rodríguez S., Carretero J.; Corrector ortográfico de libre distribución basado en reglas de derivación; Primer encuentro del grupo de usuarios de TeX hispanohablantes. EGUTH'99 pp: 44-52. Septiembre 1999 http://www.datsi.fi.upm.es/~coes/publications.html

[6] González José C., Goñi José M. y Nieto Amalio F.; ARIES: a ready for use platform for engineering Spanish-processing tools; En Digest of the Second Language Engineering Convention, paginas 219-226, Londres, Octubre 1995. http://wotan.mat.upm.es/~aries/papers.html

[7] Monedero Juan, González José C., Goñi José M., Iglesias Carlos A., y Nieto Amalio F.; Obtención automática de marcos de subcategorización verbal a partir de texto etiquetado: el sistema SOAMAS; En Actas del XI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN'95), paginas 241-254, Bilbao, Septiembre 1995. http://wotan.mat.upm.es/~aries/papers.html

[8] Zarate Antonio, Pazos Rodolfo, Gelbukh y Padron Isabel; A Portable Natural Interface for Diverse Databases Using Ontologies; Computational Linguistics and Intelligent Text Processing, paginas 494-505, México, Febrero 2003.

[9] México – Wikipedia. http://es.wikipedia.org/wiki/M%C3%A9xico

[10] Euro-Ecole. http://grandin.ecsd.net/pais.htm

[11] Ana M. Popescu , Oren Etzioni, Henry Kautz,"Towards a theory of natural language interfaces to databases," University of Washington, D.C.

http://www.cs.washington.edu/research/projects/webware1/www/precise/precise.html

[12] Boris Katz, Sue Felshin, Deniz Yuret. “Omnibase: Uniform Acces to Heterogeneous Data for Question Answering”, Artificial Intelligence Laboratory, Cambridge, MA. 2002.

[13] Oren Etzioni, Alexander Yates, Daniel Weld. A Reliable Natural Language Interface to Household Appliances.
University of Washington.

[14] Paulo Reis, Joao Matias, Nuno Mamede. “A Natural Language Interface to Databases. A new dimension for an approach”. 1997.
http://citeseer_nj.nec.com/reis97edite.html

[15] J.L. Binot, L. Debille, D. Sedlock and B. Vandecapelle. “Natural Language Interfaces: A New Philosophy”. SunExpert Magazine, 1991.

[16] Sergio Chapa, CINVESTAV, México.
http://www.cs.cinvestav.mx/BDChapa/poctzin/intro.htm

[17] X.Meng, S.Wang, and K.F. Wong. “Overview of a Chinese Natural Language Interface to Databases: NChiql”. 2001.

[18] Wesley W. Chu, Frank Meng, Gladis Kong. “Query Formulation from High-level Concepts for Relational Databases ”. University of California.
http://citeseer.nj.nec.com/zhang99query.html

[19] E.F. Codd. “A Relational Model for Large Shared Data Banks”.
Communications of the ACM, 1970.

[20] I. Androutsopoulos, G. Ritchie, and P. Thanisch, “Natural Language Interfaces to Databases An Introduction,” Natural Language Engineering, Vol. 1, No. 1,1995.

[21] W. A. Woods, R.M. Kaplan, y B.N. Webber. “The Lunar Sciences Natural Language Information System” Cambridge, Massachusetts, 1972.

[22] E.F. Codd. “Seven Steps to RENDEZVOUZ with the casual user”. North-Holland Publishers, 1974.

[23] G. Hendrix, E. Sacerdoti, D. Sagalowicz, y J. Slocum. “Developing a Natural Language Interface to Complex Data.” ACM Transactions on Database

System, 1978.

[24] D.L. Waltz. “An English Language Question Answering System for a Large Relational Database”. Communications of the ACM, 1978.

[25] R.J.H. Scha. Philips. “Question Answering System PHILIQA1”. ACM, 1977.

[26] D. Warren y F. Pereira. “An Efficient Easily Adaptable System for Interpreting Natural Language Queries”. Computacional Linguistics, 1982.

[27] Dávila Pérez Rogelio. “Una interface en Español para Bases de Datos Expresadas en Lógica”, Tesis Profesional de Maestría. Universidad de Essex, Inglaterra 1986.

[28] W. Ballard, E. Stumberger. “Semantic Acquisition in TELI: A Transportable, User-Customized Natural Language Processor” AT&T Bell Laboratories, 1987.

[29] P.Resnik. “Access to Multiple Underlyng Systems in Janus” Bolt Beranek and Newman Inc. Cambridge, Massachusetts, 1989. C.D. Hafner. “Interaction of Knowledge Sources in a Portable Natural Language Interface”. In roceedings of the 22nd Annual Meeting of ACL, Stanford, California, 1984.

[30] C.D. Hafner. Semantic of Temporal Data Queries and Temporal Data. In Proceedings of the 23rd Annual Meeting of ACL, Chicago, Illinois, 1985.

[31] C.D. Hafner. Semantic of Temporal Data Queries and Temporal Data. In Proceedings of the 23rd Annual Meeting of ACL, Chicago, Illinois, 1985.

[32] B.W. Ballard. “The Syntax and Semantics of User-Defined Modifiers in a Transportable Natural Language Processor”. In Proceedings of the 22nd Annual Meeting of ACL, Stanford, California, 1984.

[33] B.W. Ballard, J.C. Lusth, and N.L. Tinkham. “A Transportable, Knowledge based Natural Language Processor”. ACM Transactions on Office Information Systems, 1984.

[34] F. Damerau. “Operating statistics for the transformational question answering system”. American Journal of Computational Linguistic, 1981.

[35] F. Damerau. “Operating statistics for the transformational question answering system”. ACM Transactions on Office Information System, 1985.

[36] Templeton Marjorie, Burger John.”PROBLEMS IN NATURAL-LANGUAGE

INTERFACE TO DSMS WITH EXAMPLES FROM EUFID". System Development Corporation.Santa Mortice, California, 1985.

[37] B.J. Grosz. "TEAM: A Transportable Natural-Language Interface System". In Proceedings of the 1st. Conference on Applied Natural Language Processing, Santa Monica, California, 1983.

[38] B.J. Grosz, D.E. Appelt, P.A. Martin, and F.C.N. Pereira. "TEAM: An Experiment in the Design of Transportable Natural-Language Interfaces". Artificial Intelligence, 1987.

[39] H. Thompson, B. Thompson, "INTRODUCING ASK, A SIMPLE KNOWLEDGEABLE SYSTEM". California Institute of Technology,1986.

[40] BBN Systems and Technologies. BBN Parlance Interface Software – System Overview, 1989.

[41] Evaristo Daniel Chay Coyoc, ITESM Campus Morelos.Mayo 1990. http://w3.mor.itesm.mx/~jtorres/Datos.html

[42] R&D Activities; PASO PC315 PROJECT. http://www.vai.dia.fi.upm.es/ing/projects/paso.htm

[43] Gladys Rocher Silva. "Traducción de Queries en Prolog a SQL" Tesis de Licenciatura. Universidad de las Américas Puebla. http://mailweb.udlap.mx/~tesis/lis/rocher_s_gr/indice.html

[44] Procesamiento de lenguaje natural. 1998. http://gplsi.dlsi.ua.es/gplsi/areas.htm

[45] Nick Cercone, Paul McFetridge, Fred Popowish, Dan Fass, Chris Groeneboer, Gary Hall. "The SystemX, Natural Language Interface: Design, Implementation and Evaluation", Centre for Systerns Science, Simon Fraser University, British Columbia, 1993.

[47] L.R. Harris. "Experience with INTELLECT: Artificial Intelligence Technology Transfer". The AI Magazine, 1984.

[48] J.L. Manferdelli. Natural Languages. Sun Technology, 1989.

[49] Q&A. http://www.quickanswer.com/query2.htm

[50] Supun Ruwanpura; SQ-HAL, 2000. http://www.csse.monash.edu.au/hons/projects/2000/Supun.Ruwanpura

[51] English Language FrontEnd Software, 1999. http://www.elf-software.com/links.htm

[52] Adam Blum. Microsoft Corporation, 1998. http://msdn.microsoft.com/library/en-us/dnenq/html/msdm_eqwebdev.asp

[53] I. Androutsopoulos, G. Ritchie and P. Thanish. “MASQUE/SQL, An Efficient and Portable Language Query Interface for Relational Databases”. Department of Artificial Intelligence, University of Edinburgh,1993.

[54] J. Chae and S. Lee. “Frame-based Descomposition Method for Korean Language Query Processing. Computer Processing of Oriental Languages. 1998.

[55] A. Klein, J. Matiasek, y H. Trost. “The treatment of noun phrase queries in a natural language database access system”. In COLING ACL’98. 1998.

[56] Russian Research Institute of Artificial Intelligence; InBase http://www.inbase.artint.ru/nl/kadry-eng.asp

[57] VILIB Virtual Library; Tematics for libraries. 1999. http://www.wi-im.uni-koeln.de/vilib/

[58] AVENTINUS; Advanced Information System for Multinational Drug Enforcement. http://www.dcs.shef.ac.uk/nlp/funded/aventinus.html.

[59] Hanmin Jung, Gary Geunbae Lee, “Multilingual Question Answering with High Portability on Relational Databases,” Department of Computer Science and Engineering Pohang University of Science and Technology Hyoja-dong, Nam-gu, Pohang,Kyungbuk, Korea.

[60] W. C. Ogden and P. Bernickl, “Using Natural Language Interfaces”, Computer Research Laboratory New Mexico State University, Las Cruces, New Mexico 88003.

[61] S. Whittaker y P. Stenton, “User Studies and the Design of Natural Language Systems”, In Proceedings of the 4th Conference of the European Chapter of ACL, Manchester, England. Hewlett-Packard Laboratories, April 1989.

[62] Oren Etzioni, Alex Armanasu and Ana Maria Popescu. “Modern Natural Language Interfaces to Databases: Composing Statistical Parsing with

Semantic Tractability". University of Washington. 2002.

[63] G. Guida y G. Mauri, "A Formal Bases for Performance Evaluation of Natural Language Understanding Systems", Istituto di Matematica, Informatica e Sistemistica, Universita di Udine, Udine, Italy.

[64] G. G. Chowdhury, "Natural Language Processing", Dept. of Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XH, UK

[65] Stratica, N.,Kosseim, L.,Desai, B.: NLIDB Templates for Semantics Parsing. In: Proceedings of Applications of Natural Language to Data Bases (NLDB2003). pp 235-241,. http//www.cs.concordia.ca/~kosseim/research.html.

[66] Quiroz, F.:Un modelo de metadatos para información estadística.. Departamento de Sistemas Informaticos y Bases de Datos de INEGI. Boletín de politica informatica. No. 1, 2003.

[67] H. Uszkoreit, "DiET: Diagnostic and Evaluation Tools for Natural Language Applications." [Consulta: 19 de marzo de 2005], http://www.dfki.de/lt/projects/diet-e.html

[68] Montero, J.M.: Sistemas de Conversión Texto Voz. B.S. thesis. Universidad Politécnica de Madrid. http://lorien.die.upm.es/ juancho.

[69] Procesamiento del Lenguaje Natural. Escuela Superior de Ingeniería. Ingeniería. Ingeniería técnica en Informática de Gestión. Introducción a la Inteligencia Artificial. Depto. De Lenguajes y Sistemas Informáticos. Universidad de Cádiz.
http://webs.ono.com/usr008/igpeblan/files/tema5a.pdf

[70] S. Whittaker y P. Stenton, "User Studies and the Design of Natural Language Systems", In Proceedings of the 4th Conference of the European Chapter of ACL, Manchester, England. Hewlett-Packard Laboratories, April 1989.

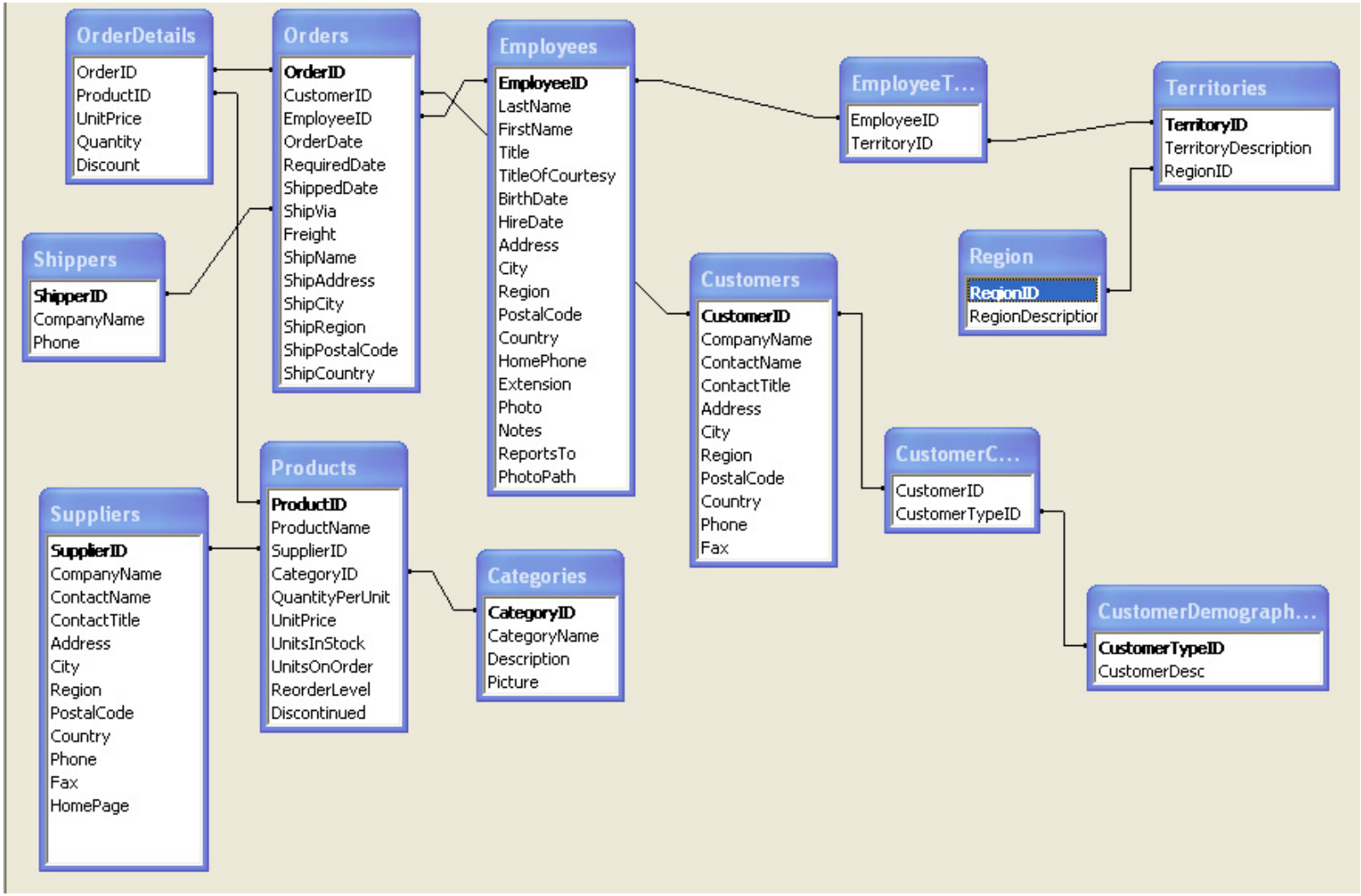
[71] C. J. Date. Introducción a los Sistemas de Bases de Datos. 7ª. Edición. Editorial Prentice may. 2001

[72] Ronald Fagin. "Acyclic Database Schemes (of Various Degrees): A Painless Introduction". IBM Research. 1983

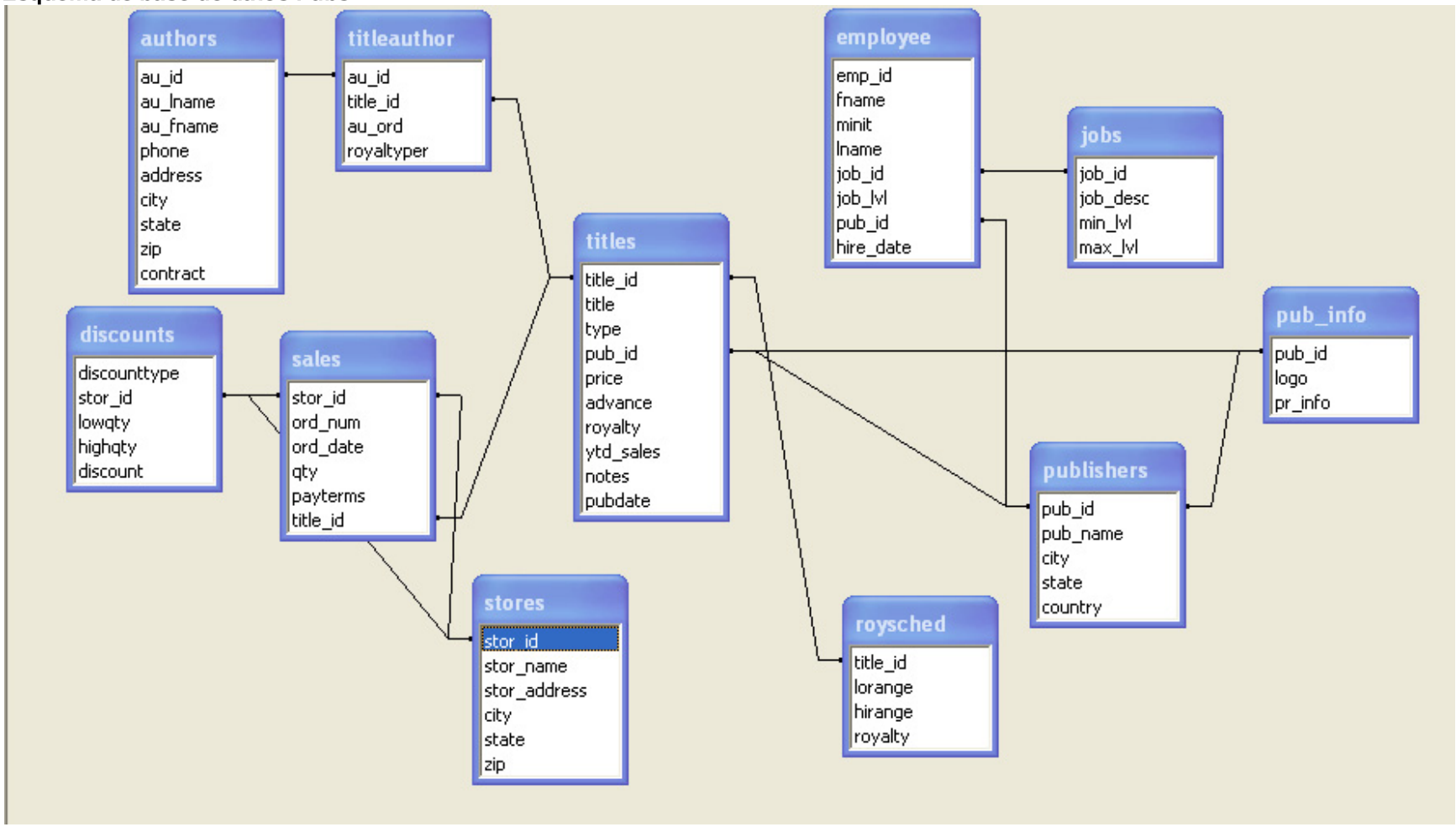
[73] Real Academia Española: Gramática Descriptiva de la Lengua Española. Espasa
Calpe, 1999.

Anexo A Esquemas

Esquema de la base de datos Northwind



Esquema de base de datos Pubs



Anexo B Corpus obtenidos

Corpus de la base de datos Northwind

Tipos 1, 5, 9, 13 Columnas y tablas explícitas

Tipo 1 Columnas y tablas explícitas sin funciones SQL y sin condición

1	Dame el nombre de todas las categorías
2	Haz un listado de todos los proveedores que contenga nombre de la compañía, ciudad y país
3	Dame la descripción de las categorías de productos
4	Dame los nombres de contactos de los clientes

Tipo 5 Columnas y tablas explícitas con funciones SQL y sin condición

Tipo 9 Columnas y tablas explícitas sin funciones SQL y con condición

5	Dame el puesto del empleado con apellido Fuller
6	Muéstrame los nombres de los empleados que tengan el puesto sales representative
7	Obtén el apellido del empleado con el puesto de vicepresident
8	Dame la fecha de la orden con identificador 10249
9	Dame el nombre del producto y la cantidad ordenada para los productos con identificador igual a 4
10	Dame el nombre de las compañías de clientes de la ciudad de México
11	Dame el teléfono del proveedor con identificador 3
12	Dame la descripción de la categoría con identificador 5
13	¿Cuál es el nombre de la categoría que tiene el identificador 6?
14	¿Qué descripción tiene el nombre de la categoría Dairy Products?
15	¿Qué numero de identificador de categoría contiene la descripción cheeses?
16	Nombre del navío de las ordenes que van al país de Francia y México
17	Dame la descripción de la categoría que tiene el identificador de la categoría numero 4
18	Quiero saber el puesto del empleado con fecha de nacimiento 30/08/1963
19	Cual es el identificador de categoría de nombre Seafood
20	Dame el nombre y el territorio del empleado con identificador 3
21	Muestra las imágenes de todos las categorías
22	Cual es el flete de la orden con fecha de envio 09/07/1996
23	Dame la descripción del territorio y la descripción de la región del identificador de región 3

24	Dame el apellido del empleado con fecha de contratación 23/07/1996
25	Cual es el teléfono particular del empleado con identificador de territorio 30346
26	Cual es la fecha de contratación del empleado con identificador de territorio 03049
27	Dame las unidades de reserva y descripción del producto con identificador igual a 2
28	Cual es la identificador del empleado con descripción de territorio Filadelfia
29	Dame el nombre del producto cuya cantidad ordenada sea mayor de 10
30	Dame el nombre del producto cuya cantidad ordenada sea mayor que 100
31	Dame el identificador de la orden cuyo monto total sea menor de 20

Tipo 13 Columnas y tablas explícitas con funciones SQL y con condición

Tipos 2, 6, 10, 14 Columnas implícitas y tablas explícitas

Tipo 2 Columnas implícitas y tablas explícitas sin funciones SQL y sin condición

32	Quiero todos los productos
33	Dame las categorías de los productos

Tipo 6 Columnas implícitas y tablas explícitas con funciones SQL y sin condición

Tipo 10 Columnas implícitas y tablas explícitas sin funciones SQL y con condición

34	Muestra los nombres de los clientes de CANADA
35	Cual es el nombre de los proveedores de alemania
36	Muestra el nombre de los proveedores de Australia
37	Muéstrame todos los empleados de la ciudad Seattle
38	Dame los proveedores de la ciudad de Tokio
39	Dame la categoría del identificador 1
40	Muéstrame las ordenes con fecha de envio 16/07/1996
41	¿Cuál es el precio del producto chang?
42	Dame los precios por unidad de los productos 3 y 6
43	Dame los códigos postales de los proveedores 1, 3 y 4
44	Un listado de todos los clientes que sean representantes de ventas
45	Muéstrame el costo total de las ordenes que van a brasil
46	¿Que identificador de categoría tiene dairy products?
47	Detalles de la Orden 10250
48	Visualiza los proveedores en USA
49	Dame los nombres de los productos de la categoría 1

50	Dame una lista de los clientes y proveedores que son de Londres
51	Quiero el tipo de comida dentro de la categoría Granos/Cereales
52	Dame la descripción de la categoría condiments
53	Dame el nombre del producto 5
54	Dame la descripción de la categoría confections
55	Dame el nombre de los clientes que viven en Londres
56	Dame todas las ordenes con fecha de envío 21/08/1996 a la ciudad de Madrid
57	Dame el precio y la cantidad del producto 42
58	Dame la descripción de la categoría 6
59	Muéstrame todas las ordenes de mas de $1000
60	Detalles de la orden 10248
61	Dame el precio unitario de la orden 10249
62	Dame a que categoría pertenece la descripción cheeses
63	Listar el nombre del producto que cueste 10
64	Dame la descripción de la categoría de lácteos
65	Dame la categoría de dulces
66	Muéstrame el precio del producto 6
67	Dame el precio de la orden 10248
68	Muestra los proveedores de Nueva Orleans
69	Muestra los proveedores de Nueva Inglaterra
70	Muestra los proveedores de Japón
71	Muéstrame todos los nombres del producto cuya categoría sea beverages
72	Dame la fecha de envío de la orden 10277
73	Muestra el identificador del empleado Michael Suyama
74	Cual es el identificador de producto de Geitost
75	Dame el precio unitario y descripción del producto con identificador de producto igual a 4
76	Cual es la descripción del cliente con identificador ANTON
77	Muestra la fotografía del empleado con identificador de territorio 30346
78	¿Que nombres de productos contiene la categoría 2?
79	Muéstrame todos los nombres del producto cuya categoría es beverages
80	Detalles de la orden del cliente VINET
81	Dame la categoría de la pasta
82	Dame la categoría del producto chai y su precio
83	Muestra la descripción del producto Aniseed Syrup
84	Muéstrame el identificador de la orden para los productos de la categoría condiments
85	Dame todas las fechas de orden que ha hecho el empleado Dodsworth

86	Dame los productos con precio mayor a 10
87	Lista los productos y cantidad que su precio por unidad sea menor de 20
112	Muéstrame todos los productos con la categoría procedure e identificador de producto 39

Tipo 14 Columnas implícitas y tablas explícitas con funciones SQL y con condición

88	Proporcióname el numero de productos que se hicieron en la orden 10248

Tipos 3, 7, 11, 15 Columnas explícitas y tablas implícitas

Tipo 3 Columnas explícitas y tablas implícitas sin funciones SQL y sin condición

89	Dame todos los nombres de las compañías
90	Dame los nombres de la compañía

Tipo 7 Columnas explícitas y tablas implícitas con funciones SQL y sin condición

Tipo 11 Columnas explícitas y tablas implícitas sin funciones SQL y con condición

91	Dame el nombre de la compañía con el identificador 1
92	Dame el domicilio de la compañía de nombre Around the horn
93	Dame el nombre del contacto de la compañía de nombre Alfredo Futterkiste
94	Dame el teléfono de la compañía con identificador 1
95	Nombre de la compañía con identificador 1
96	Dame el nombre del contacto que lleva el nombre de la compañía Toms Spezialitaten
97	Muéstrame el teléfono de la compañía United Package
98	Muéstrame el nombre del navío donde la ciudad de envío sea México
99	Cual es el nombre del navío que tiene como país de envío Brasil
100	Dame el teléfono de la compañía speedy express
101	¿Cuál es el apellido de la persona que tiene el nombre Nancy?
102	Dame los nombres de las compañías que se encuentran en la ciudad de México
103	¿Cuál es el titulo del contacto que tiene la dirección Obere Str.57?
104	Dame el nombre de la ciudad de la compañía Exotic Liquids
105	Dame el nombre del país de la compañía Tokio Traders
106	Dame el teléfono de la compañía federal shipping
107	En que ciudad se encuentra la compañía exotic liquids

108	Dame los teléfonos de las compañías united package y speedy express
109	Dame la ciudad del apellido peacock
110	Muéstrame el teléfono de la compañía fletera United Package
111	En que ciudad se encuentra la compañía Exotic Liquids

Tipo 15 Columnas explícitas y tablas implícitas con funciones SQL y con condición
Tipos 4, 8, 12, 16 Columnas y tablas implícitas

Tipo 4 Columnas y tablas implícitas sin funciones SQL y sin condición

Tipo 8 Columnas y tablas implícitas con funciones SQL y sin condición

113	A que envío pertenece el flete de mayor costo

Tipo 12 Columnas y tablas implícitas sin funciones SQL y con condición

114	¿Cuáles son los nombres de las compañías que estén situadas en México df?
115	Que clientes me compraron fletes mayores a 50
116	Datos de la compañía Tokio Traders
117	Dame la dirección de la compañía donde trabaja Ana Trujillo
118	Dame la fecha de contratación de Margaret Peacock
119	Dame el nombre de la compañía y el contacto que se encuentran en Tokio
120	Dame el identificador de United Package
121	Dame la dirección de la taqueria Antonio moreno
122	Dame el titulo, fecha de nacimiento y ciudad de Margaret
123	Dame el teléfono de speedy express
124	Dame las direcciones y los nombres de las compañías de los representantes de ventas
125	Dame el código postal y la ciudad de Exotic Liquids
126	Dame el nombre, apellido y fecha de nacimiento del vicepresidente de ventas
127	Dame la dirección de la taqueria Antonio moreno
128	Dame la fecha de cumpleaños de andrew fuller
129	Dame el numero de teléfono de federal shipping
130	Dame el apellido de Nancy
131	Dame la dirección de Alfredo futterkiste
132	Dame la dirección de margaret
133	Dame el teléfono de la compañía speedy express
134	Muéstrame la descripción de los beverages
135	Quiero la dirección de la taqueria Antonio moreno
136	Dame el nombre de las personas que vivan en México df

137	Dame el registro de la compañía de Londres
138	Donde trabaja ana trujillo
139	Quien vive en la calle mataderos
140	Quienes viven en México df
141	Dame el identificador de la carne
142	¿En que ciudad vive margaret peacock?
143	Dame los nombres de contactos de los representantes de ventas
144	Dame las unidades de reserva de chang
145	Muestra la fecha de cumpleaños de andrew fuller
146	Cual es la fecha de contratación de Robert King
147	Muestra la dirección de Laura Callahan
148	Cual es el titulo de cortesía de Nancy Davolio
149	Cual es la dirección de Leverling
150	Dame el teléfono particular de Steven Buchanan
151	Cual es el precio unitario de ikura
152	Dame las unidades de reserva de tofu
153	Muestra el descuento de Pavlova
154	Cual es el nombre del contacto de Grandma Kelly's Homestead
155	Cual es el titulo de contacto de PB knackebrod
156	Cual es el nombre de la compañía 16
157	Muestra la dirección de Karkki Oy
158	Dame el nombre de contacto de 18
159	Muestra la región de Ma Maison
160	Muestra el nombre de la compañía de 29
161	Dame el nombre de contacto del Centro Comercial Moctezuma
162	Dame el titulo del contacto de Comercio Mineiro
163	Muestra la dirección de la familia Arquibaldo
164	Muestra el código postal de GREAL
165	Cual es el fax de North/South
166	Dame la descripción del producto con identificador de producto 19
167	Cual es la descripción del producto Pavlova
168	Dame el nombre de la compañía que tiene el producto Mishi Kobe Nike
169	¿Cuantas unidades de reserva tiene la categoría seafood?
170	Cual es la extensión del identificador de territorio 90405

Tipo 16 Columnas y tablas implícitas sin funciones SQL y con condición

Tipo 17 (con funciones especiales)

171	Dame el apellido de los empleados con fecha de contratación mayor a 01/01/1992
172	Dame una lista de las ordenes realizadas entre 04/07/1996 y 11/07/1996 y que sea mayor a $51.30
173	Dame el nombre de los empleados que nacieron entre 1950 y 1960
174	Dame los nombres de los empleados contratados entre marzo del 92 y marzo del 93
175	Dame el nombre de todos los empleados que nacieron en el mes de febrero
176	Dar el monto total de las ordenes realizadas el 08/06/1996
177	Dame el identificador de todas las ordenes hechas el día de ayer
178	Dame las primeras 5 categorías
179	Dame los productos que tienen una reserva menor que el nivel de pedido
180	Dame todas las ordenes de los primeros 15 días de julio de 1997
181	Muéstrame los empleados que cumplen en el mes de enero
182	¿Qué ordenes fueron atendidas los primeros 15 días del mes de julio?
183	Cual es el nombre de las compañías que no son de estados unidos
184	En que fecha se realizan mas envíos
185	En que fecha hubo menos envíos
186	En que fecha hubo mas contratados
187	Cual es el nombre del empleado que haya nacido después del empleado que nació en 1948
188	Dame las unidades en reserva del identificador de la orden 10269 (9)
189	Muestra el nombre del contacto que tiene el producto chai
190	Dame el identificador del producto que se envío el 16/07/1996
191	Dame el nombre de los clientes que han ordenado productos el mes pasado
192	¿Cuántas unidades se vendieron en la orden del 4 de junio de 1996, enviado de Francia
193	Cual es el nombre de la compañía que tiene fecha de envío 12/07/1996
194	Cual es el nombre del contacto que tiene la fecha de orden 30/07/1996
195	Cual es el identificador de producto que tiene identificador de cliente GODOS
196	Cual es el descuento del navío que delicia
197	Dame el nombre de las compañías que empiezan con a
198	Dame los nombres de empleados que empiecen con N

Corpus de la base de datos Pubs

Tipos 1, 5, 9, 13 Columnas y tablas explícitas

Tipo 1 Columnas y tablas explícitas sin funciones SQL y sin condición

1	Dame los titulos de los libros
2	Visualiza los tipos de descuentos
3	Cual es la dirección de la editorial y su ciudad

Tipo 5 Columnas y tablas explícitas con funciones SQL y sin condición(0)

Tipo 9 Columnas y tablas explícitas sin funciones SQL y con condición

4	Cual es el titulo del libro con identificador TC4203
5	Selecciona el titulo donde el precio sea igual a $19.99 y el tipo sea bussines

Tipo 13 Columnas y tablas explícitas con funciones SQL y con condición

Tipos 2, 6, 10, 14 Columnas implícitas y tablas explícitas

Tipo 2 Columnas implícitas y tablas explícitas sin funciones SQL y sin condición

6	Lista los empleados con su respectivo cargo
7	Que puesto ocupa cada empleado

Tipo 6 Columnas implícitas y tablas explícitas con funciones SQL y sin condición

Tipo 10 Columnas implícitas y tablas explícitas sin funciones SQL y con condición

8	A que almacen pertenece la siguiente dirección 679 carson st.
9	Que empleado tiene como identificador H-B39728F
10	A que almacen corresponde el identificador 7131

11	Que autores viven en la ciudad de Oakland
12	Que trabajador tiene como nivel de trabajo 227
13	Que trabajador tiene su fecha de contratación como 13/02/1991
14	Que libros son del tipo bussines
15	Quien es el autor del titulo the busy
16	Selecciona todos los libros en donde su adelanto sea mayor a %5000
17	Mostrar los libros cuyo precio es mayor a $19.99 y son de tipo bussines
18	Que editorial se encuentra en Alemania
19	Quien es el autor del libro the Gourmet
20	Quien es el autor del libro the busy
21	Selecciona todos los libros del autor Smith
22	Que libros son de la editorial Algodata Infosystems
23	Dame los titulos del autor green
24	Que puesto tiene el empleado francisco
25	Que puesto tiene el empleado Francisco Chang
26	Selecciona el descuento para el almacén 8042
27	Que puesto tiene el empleado Paolo y su fecha de contratación
28	Dime los libros que fueron vendidos en la fecha 13/09/2004 y que son diferentes del tipo business
29	Cual es la dirección del almacen Barnum's
30	Cual es el numero de teléfono del autor cheryl
31	Cual es la ciudad de la editorial New Moon Books
32	Cual es el identificador del empleado paolo accort
33	Cual es el nivel de trabajo de philip cramer
34	Cual es el precio del identificador de editor 1389
35	Cuales son los titulos de la editorial GGG&G
36	Cual es la clave y el precio del libro you can
37	Cual es la dirección y el teléfono del autor del libro you can
38	En que ciudad se encuentra el autor Jonson White
39	Que apellido tiene el empleado Pedro
40	En que ciudad se encuentra el almacen Bookbeat
41	En que ciudad se encuentra la editorial lucerne publishing
42	En que estado se encuentra la editorial Ramona publishers
43	En que ciudad se ubica la tienda Erick the read books
44	Titulo de los libros cuyos editores se encuentran en Texas
45	Que nombre y dirección tiene el empleado que trabaja para la editorial GGG&G
46	Que descripción tiene el puesto del empleado VPA30890F

47	Obtener el nombre del almacen donde se encuentra el libro cooking with
48	En que fecha se realizo el contrato del empleado PTC11962M

Tipo 14 Columnas implícitas y tablas explícitas con funciones SQL y con condición

49	Cuantos autores son de la ciudad de Berkeley
50	Cual es el numero de empleados de la editorial Scotney book
51	Cual es el numero de ventas realizadas el 14/09/1994
52	Dame el numero de libros vendidos el 13/19/1994

Tipos 3, 7, 11, 15 Columnas explícitas y tablas implícitas

Tipo 3 Columnas explícitas y tablas implícitas sin funciones SQL y sin condición

53	Que identificador tienen los títulos

Tipo 7 Columnas explícitas y tablas implícitas con funciones SQL y sin condición

Tipo 11 Columnas explícitas y tablas implícitas sin funciones SQL y con condición

54	A que ciudad pertenece el código postal 89076
55	A que ciudad corresponde la dirección 567 Pasadena Ave
56	Cual es el adelanto del numero de orden 6871
57	Dame la fecha de contratación de pedro

Tipo 15 Columnas explícitas y tablas implícitas con funciones SQL y con condición

Tipos 4, 8, 12, 16 Columnas y tablas implícitas

Tipo 4 Columnas y tablas implícitas sin funciones SQL y sin condición

58	Que titulos contiene cada editorial

Tipo 8 Columnas y tablas implícitas con funciones SQL y sin condición

Tipo 12 Columnas y tablas implícitas sin funciones SQL y con condición

59	Que cantidad de silicon valley es vendida

60	En que editorial trabaja victoria ashworth
61	Nombre del almacen donde se encuentra the busy

Tipo 16 Columnas y tablas implícitas sin funciones SQL y con condición

62	Cuantos números de ejemplares tiene el libro the busy

Tipo 17 (difíciles y con funciones especiales)

63	Cuantos ejemplares del libro the busy se vendieron el 14 de septiembre
64	Todos los empleados que tengan un puesto
65	Cual es el nombre de la editorial que no tenga estado y a su vez no este en USA
66	Cual es el titulo del libro que no tiene precio
67	Cuántas ventas se realizaron en el año de 1992
68	Selecciona todas las editoriales del mismo país
69	Cual es el libro mas barato de tipo business
70	Quien es el empleado que tiene mas tiempo trabajando

Anexo C Traducción de consultas limpias del corpus de la base de datos Northwind

Tipo 1 Columnas y tablas explícitas sin funciones SQL y sin condición (1)

	Consulta Imperativa	Traducción	Consulta Interrogativa	Traducción
1	Dame el nombre de todas las categorías	SELECT Categories.CategoryName FROM Categories	¿Cuál es el nombre de todas las categorías?	SELECT Categories.CategoryName FROM Categories

Tipo 9 Columnas y tablas explícitas sin funciones SQL y con condición (8)

6	Dame el apellido de los empleados con fecha de contratación mayor a 01/01/1992	SELECT Employees.FirstName, Employees.LastName FROM Employees WHERE ((Employees.HireDate > #1992/01/01#))	¿Cuál es el apellido de los empleados con fecha de contratación mayor a 01/01/1992?	SELECT Employees.FirstName, Employees.LastName FROM Employees WHERE ((Employees.HireDate > #1992/01/01#))
12	Dame el nombre de las compañías de clientes de la **ciudad de México**	SELECT Customers.CompanyName FROM Customers WHERE ((Customers.City LIKE '%México%'))	¿Cuál es el nombre de las compañías de clientes de la **ciudad de México**?	SELECT Customers.CompanyName FROM Customers WHERE ((Customers.City LIKE '%México%'))
19	Dame el flete de la orden con fecha de envio 09/07/1996	SELECT Orders.Freight FROM Orders WHERE ((Orders.ShippedDate = #1996/07/09#))	¿Cual es el flete de la orden con fecha de envio 09/07/1996?	SELECT Orders.Freight FROM Orders WHERE ((Orders.ShippedDate = #1996/07/09#))
20	Dame el apellido del empleado con fecha de contratación 23/07/1996	SELECT Employees.FirstName, Employees.LastName FROM Employees WHERE ((Employees.HireDate = #1996/07/23#))	¿Cuál es el apellido del empleado con fecha de contratación 23/07/1996?	SELECT Employees.FirstName, Employees.LastName FROM Employees WHERE ((Employees.HireDate = #1996/07/23#))
25	Dame el identificador de la orden cuyo monto total sea menor de 20	SELECT Orders.OrderID FROM Orders WHERE ((Orders.Freight < 20))	¿Cuál es el identificador de la orden cuyo monto total es menor de 20?	SELECT Orders.OrderID FROM Orders WHERE ((Orders.Freight < 20))
28	Dame el identificador del empleado con descripción de territorio Philadelphia	SELECT Employees.EmployeeID, EmployeeTerritories.EmployeeID, Territories.TerritoryID, Territories.TerritoryDescription FROM Employees, EmployeeTerritories, Territories WHERE ((Territories.TerritoryDescription LIKE '%Philadelphia%')) AND Territories.TerritoryID =	Cual es el identificador del empleado con descripción de territorio Philadelphia	SELECT Employees.EmployeeID, EmployeeTerritories.EmployeeID, Territories.TerritoryID, Territories.TerritoryDescription FROM Employees, EmployeeTerritories, Territories WHERE ((Territories.TerritoryDescription LIKE '%Philadelphia%')) AND Territories.TerritoryID =

		EmployeeTerritories.TerritoryID AND EmployeeTerritories.EmployeeID = Employees.EmployeeID		EmployeeTerritories.TerritoryID AND EmployeeTerritories.EmployeeID = Employees.EmployeeID
30	Dame la descripción que tiene el nombre de la categoría **"Dairy Products"**	SELECT Categories.Description FROM Categories WHERE ((Categories.CategoryName LIKE '%Dairy_Products%'))	¿Que descripción tiene el nombre de la categoría "Dairy Products"?	SELECT Categories.Description FROM Categories WHERE ((Categories.CategoryName LIKE '%Dairy_Products%'))
33	Dame la descripción de la categoría que tiene el identificador de la categoría numero 4	SELECT Categories.Description FROM Categories WHERE ((Categories.CategoryID = 4))	¿Cuál es la descripción de la categoría que tiene identificador de la categoría número 4?	SELECT Categories.Description FROM Categories WHERE ((Categories.CategoryID = 4))

Tipo 2 Columnas implícitas y tablas explícitas sin funciones SQL y sin condición (1)

37	Dame las categorías de los productos	SELECT Products.CategoryID, Categories.CategoryID, Categories.CategoryName FROM Categories, Products WHERE Products.CategoryID = Categories.CategoryID	¿Cuáles son las categorías de los productos?	SELECT Products.CategoryID, Categories.CategoryID, Categories.CategoryName FROM Categories, Products WHERE Products.CategoryID = Categories.CategoryID

Tipo 10 Columnas implícitas y tablas explícitas sin funciones SQL y con condición (15)

38	Dame el nombre de la compañía de los clientes con identificador ANTON	SELECT Customers.CompanyName FROM Customers WHERE ((Customers.CustomerID LIKE '%ANTON%'))	¿Cual es la nombre de la compañía de los clientes con identificador ANTON?	SELECT Customers.CompanyName FROM Customers WHERE ((Customers.CustomerID LIKE '%ANTON%'))
39	Muéstrame todos los nombres del producto cuya categoría es beverages NOTA: El resultado de la consulta es correcto, pero la consulta SQL no lo es, por que iguala 'beverages' a CategoryID, pero debido a que también lo iguala con CategoryName es por eso que el resultado es correcto. Funciona correctamente si la consulta LN fuera:	SELECT Products.ProductName, Categories.CategoryID, Categories.CategoryName FROM Categories, Products WHERE ((Categories.CategoryID LIKE '%beverages%' OR Products.CategoryID LIKE '%beverages%' OR Categories.CategoryName LIKE '%beverages%')) AND Products.CategoryID = Categories.CategoryID	¿Cuáles son los nombres de los productos cuya categoría es beverages?	SELECT Products.ProductName, Categories.CategoryID, Categories.CategoryName FROM Categories, Products WHERE ((Categories.CategoryID LIKE '%beverages%' OR Products.CategoryID LIKE '%beverages%' OR Categories.CategoryName LIKE '%beverages%')) AND Products.CategoryID = Categories.CategoryID

	Muéstrame todos los nombre del producto cuyo nombre de categoría es beverages			
41	Dame los nombres de los productos que contiene la categoría 2	SELECT Products.ProductName, Categories.CategoryID, Categories.CategoryName FROM Categories, Products WHERE ((Categories.CategoryID = 2 OR Products.CategoryID = 2 OR Categories.CategoryName = '2')) AND Products.CategoryID = Categories.CategoryID	¿Que nombres de productos contiene la categoría 2?	SELECT Products.ProductName, Categories.CategoryID, Categories.CategoryName FROM Categories, Products WHERE ((Categories.CategoryID = 2 OR Products.CategoryID = 2 OR Categories.CategoryName = '2')) AND Products.CategoryID = Categories.CategoryID
46	Lista los nombres de los proveedores de Germany	SELECT Suppliers.CompanyName, Suppliers.ContactName FROM Suppliers WHERE ((Suppliers.Country LIKE 'Germany'))	Cual es el nombre de los proveedores de Germany	SELECT Suppliers.CompanyName, Suppliers.ContactName FROM Suppliers WHERE ((Suppliers.Country LIKE 'Germany'))
49	Dame los precios por unidad de los productos con identificador de producto 3 y 6	SELECT Products.UnitPrice FROM Products WHERE ((Products.ProductID = 3) OR (Products.ProductID = 6))	¿Qué precios por unidad tienen los productos con identificador de producto 3 y 6?	SELECT OrderDetails.UnitPrice, Products.UnitPrice FROM OrderDetails, Products WHERE ((OrderDetails.ProductID = 3 OR Products.CategoryID = 3 OR Products.ProductID = 3 OR Products.ProductName = '3' OR Products.SupplierID = 3) OR (OrderDetails.ProductID = 6 OR Products.CategoryID = 6 OR Products.ProductID = 6 OR Products.ProductName = '6' OR Products.SupplierID = 6)) AND Products.ProductID = OrderDetails.ProductID
54	Dame la descripción de la categoría condiments	SELECT Categories.Description FROM Categories WHERE ((Categories.CategoryName LIKE 'condiments'))	¿Cuál es la descripción de la categoría condiments?	SELECT Categories.Description FROM Categories WHERE ((Categories.CategoryName LIKE 'condiments'))
58	Dame el precio y la cantidad del producto con identificador de producto 42	SELECT Products.QuantityPerUnit, Products.UnitPrice FROM Products WHERE ((Products.ProductID = 42))	¿Cuál es el precio y la cantidad del producto con identificador de producto 42?	SELECT Products.QuantityPerUnit, Products.UnitPrice FROM Products WHERE ((Products.ProductID = 42))
60	Dame el precio unitario de la orden con identificador 10249	SELECT OrderDetails.UnitPrice FROM OrderDetails WHERE ((OrderDetails.OrderID = 10249 OR OrderDetails.ProductID = 10249))	¿Cuál es el precio unitario de la orden con identificador 10249?	SELECT OrderDetails.UnitPrice FROM OrderDetails WHERE ((OrderDetails.OrderID = 10249 OR OrderDetails.ProductID = 10249))

62	Dame la descripción de la categoría de Dairy	SELECT Categories.Description FROM Categories WHERE ((Categories.CategoryName LIKE '%Dairy%'))	¿Cuál es la descripción de la categoría de Dairy?	SELECT Categories.Description FROM Categories WHERE ((Categories.CategoryName LIKE '%Dairy%'))
64	Dame el precio de la orden con identificador 10248	SELECT Orders.Freight, OrderDetails.UnitPrice FROM OrderDetails, Orders WHERE ((Orders.CustomerID = '10248' OR Orders.EmployeeID = 10248 OR OrderDetails.OrderID = 10248 OR Orders.OrderID = 10248 OR OrderDetails.ProductID = 10248)) AND OrderDetails.OrderID = Orders.OrderID	¿Cuál es el precio de la orden con identificador 10248?	SELECT Orders.Freight, OrderDetails.UnitPrice FROM OrderDetails, Orders WHERE ((Orders.CustomerID = '10248' OR Orders.EmployeeID = 10248 OR OrderDetails.OrderID = 10248 OR Orders.OrderID = 10248 OR OrderDetails.ProductID = 10248)) AND OrderDetails.OrderID = Orders.OrderID
73	Dame los proveedores de la ciudad de Tokio	SELECT Suppliers.SupplierID FROM Suppliers WHERE ((Suppliers.City LIKE '%Tokio%'))	¿Quiénes son los proveedores de la ciudad de Tokio?	SELECT Suppliers.SupplierID FROM Suppliers WHERE ((Suppliers.City LIKE '%Tokio%'))
76	Un listado de todos los clientes que sean "Sales Representative"	SELECT Customers.CustomerID FROM Customers WHERE ((Customers.ContactTitle LIKE 'Sales_Representative'))	¿Qué clientes son "Sales Representative"?	SELECT Customers.CustomerID FROM Customers WHERE ((Customers.ContactTitle LIKE 'Sales_Representative'))
82	Dame una lista de los clientes que son de London	SELECT Customers.CustomerID FROM Customers WHERE ((Customers.City LIKE 'London'))	¿Qué clientes son de London?	SELECT Customers.CustomerID FROM Customers WHERE ((Customers.City LIKE 'London'))
85	Dame la categoría de sweet	SELECT Categories.CategoryID, Categories.CategoryName FROM Categories WHERE ((Categories.Description LIKE '%sweet%'))	¿Cuál es la categoría sweet?	SELECT Categories.CategoryID, Categories.CategoryName FROM Categories WHERE ((Categories.Description LIKE '%sweet%'))
90	Dame la categoría de la pasta	SELECT Categories.CategoryID, Categories.CategoryName FROM Categories WHERE ((Categories.Description LIKE '%pasta%'))	¿Cuál es la categoría de la pasta?	SELECT Categories.CategoryID, Categories.CategoryName FROM Categories WHERE ((Categories.Description LIKE '%pasta%'))
93	Muéstrame todos los productos con la categoría "produce" e identificador de producto 39 NOTA: Se reescribe Muéstrame todos los productos con el nombre de categoría "produce" e identificador de producto 39 En categoría de debe especificar columna por que hay ambigüedad por	SELECT Products.ProductID, Products.ProductName, Categories.CategoryID, Categories.CategoryName FROM Categories, Products WHERE ((Categories.CategoryName LIKE '%produce%') AND (Products.ProductID = 39)) AND Products.CategoryID = Categories.CategoryID	¿Qué productos tienen nombre de categoría "produce" e identificador de producto 39?	SELECT Products.ProductID, Products.ProductName, Categories.CategoryID, Categories.CategoryName FROM Categories, Products WHERE ((Categories.CategoryName LIKE '%produce%') AND (Products.ProductID = 39)) AND Products.CategoryID = Categories.CategoryID

	las columnas que tienes el nombre de la tabla			

Tipo 14 Columnas implícitas y tablas explícitas con funciones SQL y con condición

97	Proporcióname el numero de productos que se hicieron en la orden 10248	NO	¿Cuántos productos que se hicieron en la orden 10248?	NO

Tipo 3 Columnas explícitas y tablas implícitas sin funciones SQL y sin condición (1)

99	Dame los nombres de la compañía de clientes	SELECT Customers.CompanyName FROM Customers	¿Cuáles son los nombres de las compañías de clientes?	SELECT Customers.CompanyName FROM Customers

Tipo 11 Columnas explícitas y tablas implícitas sin funciones SQL y con condición (6)

106	Muéstrame el teléfono del cliente cuya compañía es "United Package"	SELECT Customers.Phone FROM Customers WHERE ((Customers.CompanyName LIKE '%United_Package%'))	¿Cuál es el teléfono del cliente cuya compañía es "United Package"?	SELECT Customers.Phone FROM Customers WHERE ((Customers.CompanyName LIKE '%United_Package%'))
108	Dame el nombre del navío que tiene como país de envío Brazil	SELECT Orders.ShipName FROM Orders WHERE ((Orders.ShipCountry LIKE 'Brazil'))	Cual es el nombre del navío que tiene como país de envío Brazil	SELECT Orders.ShipName FROM Orders WHERE ((Orders.ShipCountry LIKE 'Brazil'))
110	Dame el apellido del **empleado** que tiene el nombre Nancy	SELECT Employees.FirstName, Employees.LastName FROM Employees WHERE ((Employees.FirstName LIKE '%Nancy%' OR Employees.LastName LIKE '%Nancy%'))	¿Cuál es el apellido del empleado que tiene el nombre Nancy?	SELECT Employees.FirstName, Employees.LastName FROM Employees WHERE ((Employees.FirstName LIKE '%Nancy%' OR Employees.LastName LIKE '%Nancy%'))
112	Dame el puesto del contacto **del cliente** que tiene la dirección **"Obere Str.57"**	SELECT Customers.ContactTitle FROM Customers WHERE ((Customers.Address LIKE '%Obere_Str.57%'))	¿Cuál es el puesto del contacto **del cliente** que tiene la dirección "Obere Str.57"?	SELECT Customers.ContactTitle FROM Customers WHERE ((Customers.Address LIKE '%Obere_Str.57%'))
116	En que ciudad se encuentra la compañía de proveedores **"exotic liquids"**	SELECT Suppliers.City FROM Suppliers WHERE ((Suppliers.CompanyName LIKE '%exotic_liquids%'))	¿En que ciudad se encuentra la compañía de proveedores "**exotic liquids**"?	SELECT Suppliers.City FROM Suppliers WHERE ((Suppliers.CompanyName LIKE '%exotic_liquids%'))
121	Dame el nombre del contacto	SELECT Customers.ContactName,	¿Cual es el nombre del	SELECT Customers.ContactName,

	del cliente que tiene la fecha de orden 30/07/1996	Orders.OrderID, Orders.OrderDate FROM Customers, Orders WHERE ((Orders.OrderDate = #1996/07/30#)) AND Customers.CustomerID = Orders.CustomerID	contacto **del cliente** que tiene la fecha de orden 30/07/1996?	Orders.OrderID, Orders.OrderDate FROM Customers, Orders WHERE ((Orders.OrderDate = #1996/07/30#)) AND Customers.CustomerID = Orders.CustomerID

Tipo 12 Columnas y tablas implícitas sin funciones SQL y con condición (15)

124	Dame el teléfono del fletador "speedy express"	SELECT Shippers.Phone FROM Shippers WHERE ((Shippers.CompanyName LIKE 'speedy_express'))	¿Cuál es el teléfono del fletador "speedy express"?	SELECT Shippers.Phone FROM Shippers WHERE ((Shippers.CompanyName LIKE 'speedy_express'))
126	Dame los nombres de las compañías de clientes que están situadas en "México D.F."	SELECT Customers.CompanyName FROM Customers WHERE ((Customers.City LIKE 'México_D.F.'))	¿Cuáles son los nombres de las compañías de clientes que están situadas en "México D.F."?	SELECT Customers.CompanyName FROM Customers WHERE ((Customers.City LIKE 'México_D.F.'))
129	Dame la fecha de contratación **del empleado** Margaret Peacock	SELECT Employees.HireDate FROM Employees WHERE ((Employees.FirstName LIKE 'Margaret') OR (Employees.LastName LIKE 'Peacock'))	¿Cuál es la fecha de contratación **del empleado** Margaret Peacock?	SELECT Employees.HireDate FROM Employees WHERE ((Employees.FirstName LIKE 'Margaret') OR (Employees.LastName LIKE 'Peacock'))
131	Dame el identificador del fletador "United Package"	SELECT Shippers.ShipperID FROM Shippers WHERE ((Shippers.CompanyName LIKE 'United_Package'))	¿Cuál es el identificador del fletador "United Package"?	SELECT Shippers.ShipperID FROM Shippers WHERE ((Shippers.CompanyName LIKE 'United_Package'))
136	Dame el código postal y la ciudad del **proveedor "Exotic Liquids"**	SELECT Suppliers.City, Suppliers.PostalCode FROM Suppliers WHERE ((Suppliers.CompanyName LIKE 'Exotic_Liquids'))	¿Cuál es el código postal y la ciudad del **proveedor "Exotic Liquids"?**	SELECT Suppliers.City, Suppliers.PostalCode FROM Suppliers WHERE ((Suppliers.CompanyName LIKE 'Exotic_Liquids'))
138	Dame la dirección **del cliente "Antonio moreno taquería"**	SELECT Customers.Address FROM Customers WHERE ((Customers.CompanyName LIKE 'Antonio_moreno_taquería'))	¿Cuál es la dirección **del cliente "Antonio moreno taquería"?**	SELECT Customers.Address FROM Customers WHERE ((Customers.CompanyName LIKE 'Antonio_moreno_taquería'))
139	Dame la fecha de nacimiento **del empleado** andrew fuller	SELECT Employees.BirthDate FROM Employees WHERE ((Employees.FirstName LIKE 'andrew') OR (Employees.LastName LIKE 'fuller'))	¿Cuál es la fecha de nacimiento **del empleado** andrew fuller?	SELECT Employees.BirthDate FROM Employees WHERE ((Employees.FirstName LIKE 'andrew') OR (Employees.LastName LIKE 'fuller'))
140	Dame el numero de	SELECT Shippers.Phone	¿Cuál es el numero de	SELECT Shippers.Phone

	teléfono del **fletador "federal shipping"**	FROM Shippers WHERE ((Shippers.CompanyName LIKE 'federal_shipping'))	teléfono del **fletador "federal shipping"?**	FROM Shippers WHERE ((Shippers.CompanyName LIKE 'federal_shipping'))
154	Dame el titulo de cortesía **de la empleada Nancy Davolio**	SELECT Employees.TitleOfCourtesy FROM Employees WHERE ((Employees.FirstName LIKE 'Nancy') OR (Employees.LastName LIKE 'Davolio'))	Cual es el titulo de cortesía de la empleada Nancy Davolio	SELECT Employees.TitleOfCourtesy FROM Employees WHERE ((Employees.FirstName LIKE 'Nancy') OR (Employees.LastName LIKE 'Davolio'))
162	Dame el nombre de la compañía **de los proveedores con identificador** 16	SELECT Suppliers.CompanyName FROM Suppliers WHERE ((Suppliers.SupplierID = 16))	Cual es el nombre de la compañía de los proveedores con identificador 16	SELECT Suppliers.CompanyName FROM Suppliers WHERE ((Suppliers.SupplierID = 16))
168	Dame el puesto del contacto del cliente "Comércio Mineiro"	SELECT Customers.ContactTitle FROM Customers WHERE ((Customers.CompanyName LIKE 'Comércio_Mineiro'))	¿Cuál es el puesto del contacto del cliente "Comércio Mineiro"?	SELECT Customers.ContactTitle FROM Customers WHERE ((Customers.CompanyName LIKE 'Comércio_Mineiro'))
172	Dame la descripción de la categoría de producto Pavlova	SELECT Categories.Description, Products.ProductID, Products.ProductName FROM Categories, Products WHERE ((Products.ProductName LIKE 'Pavlova')) AND Products.CategoryID = Categories.CategoryID	¿Cual es la descripción de la categoría de producto Pavlova?	SELECT Categories.Description, Products.ProductID, Products.ProductName FROM Categories, Products WHERE ((Products.ProductName LIKE 'Pavlova')) AND Products.CategoryID = Categories.CategoryID
174	Dame los teléfonos de las fletadoras "united package" y "speedy express"	SELECT Shippers.Phone FROM Shippers WHERE ((Shippers.CompanyName LIKE 'united_package') OR (Shippers.CompanyName LIKE 'speedy_express'))	¿Cuáles son los teléfonos de las fletadoras "united package" y "speedy express"?	SELECT Shippers.Phone FROM Shippers WHERE ((Shippers.CompanyName LIKE 'united_package') OR (Shippers.CompanyName LIKE 'speedy_express'))
178	Dame las unidades de reserva de los productos de la categoría seafood	SELECT Products.UnitsInStock, Categories.CategoryID, Categories.CategoryName FROM Categories, Products WHERE ((Categories.CategoryName LIKE 'seafood')) AND Products.CategoryID = Categories.CategoryID	¿Cuantas unidades de reserva de productos tiene la categoría seafood?	SELECT Products.UnitsInStock FROM Products WHERE ((Products.CategoryID LIKE '%seafood%'))
180	Dame la extensión del empleado con identificador de territorio 90405	SELECT Employees.Extension, Territories.TerritoryID, Territories.TerritoryDescription FROM EmployeeTerritories, Territories WHERE ((EmployeeTerritories.TerritoryID = '90405' OR Territories.TerritoryID = '90405')) AND Territories.TerritoryID = EmployeeTerritories.TerritoryID	Cual es la extensión del empleado con identificador de territorio 90405	SELECT Employees.Extension, Territories.TerritoryID, Territories.TerritoryDescription FROM EmployeeTerritories, Territories WHERE ((EmployeeTerritories.TerritoryID = '90405' OR Territories.TerritoryID = '90405')) AND Territories.TerritoryID = EmployeeTerritories.TerritoryID
181	Lista los clientes que compraron fletes mayores a 50	SELECT Customers.CustomerID FROM Customers WHERE ((Customers.Fax > '50'))	Que clientes me compraron fletes mayores a 50	SELECT Customers.CustomerID FROM Customers WHERE ((Customers.Fax > '50'))

Anexo D Traducción de consultas limpias del corpus de la base de datos Pubs

Tipo 1 Columnas y tablas explícitas sin funciones SQL y sin condición

	Consulta Imperativa	Traducción	Consulta Interrogativa	Traducción
1	Dame los titulos de los libros	SELECT titles.title FROM titles	¿Cuáles son los titulos de los libros?	SELECT titles.title FROM titles
2	Visualiza los tipos de descuentos	SELECT discounts.discounttype FROM discounts	¿Cuáles son los tipos de descuentos?	SELECT discounts.discounttype FROM discounts
3	Dame la direccion **y la ciudad** de las editoriales	SELECT publishers.city FROM publishers	¿Cual es la dirección y la ciudad de las editoriales?	SELECT publishers.city FROM publishers

Tipo 9 Columnas y tablas explícitas sin funciones SQL y con condición

4	Dame el titulo del libro con identificador TC4203	SELECT titles.title FROM titles WHERE ((titles.pub_id LIKE '%TC4203%' OR titles.title_id LIKE '%TC4203%'))	¿Cual es el titulo del libro con identificador TC4203?	SELECT titles.title FROM titles WHERE ((titles.pub_id LIKE '%TC4203%' OR titles.title_id LIKE '%TC4203%'))
5	Selecciona el titulo donde el precio sea igual a $19.99 y el tipo sea business	SELECT titles.title FROM titles WHERE ((titles.price = 19.99) AND (titles.type LIKE '%business%'))	¿Qué titulo tiene el precio igual a $19.99 y el tipo business?	SELECT titles.title FROM titles WHERE ((titles.price = 19.99) AND (titles.type LIKE '%business%'))

Tipo 2 Columnas implícitas y tablas explícitas sin funciones SQL y sin condición

6	Lista los empleados con su respectivo cargo	SELECT employee.emp_id, employee.fname, jobs.job_desc, employee.job_id, jobs.job_id, employee.lname FROM employee, jobs WHERE employee.job_id = jobs.job_id	¿Qué cargo tienen los empleados?	SELECT employee.emp_id, employee.fname, jobs.job_desc, employee.job_id, jobs.job_id, employee.lname FROM employee, jobs WHERE employee.job_id = jobs.job_id
7	Dame el puesto que ocupa cada empleado	SELECT employee.emp_id, employee.fname, jobs.job_desc, employee.job_id, jobs.job_id, employee.lname FROM employee, jobs WHERE employee.job_id = jobs.job_id	¿Que puesto ocupa cada empleado?	SELECT employee.emp_id, employee.fname, jobs.job_desc, employee.job_id, jobs.job_id, employee.lname FROM employee, jobs WHERE employee.job_id = jobs.job_id

Tipo 10 Columnas implícitas y tablas explícitas sin funciones SQL y con condición

8	Dame el almacen al pertenece la siguiente dirección "679 carson st."	SELECT stores.stor_id, stores.stor_name FROM stores WHERE ((stores.stor_address LIKE '%679_carson_st.%'))	¿Qué almacen tiene la dirección "679 carson st."?	SELECT stores.stor_id, stores.stor_name FROM stores WHERE ((stores.stor_address LIKE '%679_carson_st.%'))
9	Muéstrame el empleado con identificador H-B39728F	SELECT employee.emp_id, employee.fname, employee.lname FROM employee WHERE ((employee.emp_id LIKE '%H-B39728F%' OR employee.job_id LIKE '%H-B39728F%' OR employee.pub_id LIKE '%H-B39728F%'))	¿Que empleado tiene como identificador H-B39728F?	SELECT employee.emp_id, employee.fname, employee.lname FROM employee WHERE ((employee.emp_id LIKE '%H-B39728F%' OR employee.job_id LIKE '%H-B39728F%' OR employee.pub_id LIKE '%H-B39728F%'))
11	Lista los autores que viven en la ciudad de Oakland	SELECT authors.au_fname, authors.au_id, authors.au_lname FROM authors WHERE ((authors.city LIKE '%Oakland%'))	Que autores viven en la ciudad de Oakland	SELECT authors.au_fname, authors.au_id, authors.au_lname FROM authors WHERE ((authors.city LIKE '%Oakland%'))
12	Dame los trabajadores con nivel 227	SELECT employee.emp_id, employee.fname, employee.lname FROM employee WHERE ((employee.job_lvl = 227))	Que trabajadores tienen nivel 227	SELECT employee.emp_id, employee.fname, employee.lname FROM employee WHERE ((employee.job_lvl = 227))
13	Dame el trabajador que tiene su fecha de contratación como 13/02/1991	SELECT employee.emp_id, employee.fname, employee.lname FROM employee WHERE ((employee.hire_date = #1991/02/13#))	Que trabajador tiene su fecha de contratación como 13/02/1991	SELECT employee.emp_id, employee.fname, employee.lname FROM employee WHERE ((employee.hire_date = #1991/02/13#))
15	Muestra el autor del titulo "the busy"	SELECT authors.au_fname, authors.au_id, titleauthor.au_id, authors.au_lname FROM authors, titleauthor, titles WHERE ((titles.title LIKE '%the_busy%')) AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id	Quien es el autor del titulo "the busy"	SELECT authors.au_fname, authors.au_id, titleauthor.au_id, authors.au_lname FROM authors, titleauthor, titles WHERE ((titles.title LIKE '%the_busy%')) AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id
16	Selecciona todos los libros en donde su adelanto sea mayor a 5000	SELECT titles.title, titles.title_id FROM titles WHERE ((titles.advance > 5000))	¿Qué libros tienen su adelanto mayor a 5000?	SELECT titles.title, titles.title_id FROM titles WHERE ((titles.advance > 5000))
17	Mostrar los libros cuyo precio es mayor a $19.99 y son de tipo business	SELECT titles.title, titles.title_id FROM titles WHERE ((titles.price > 19.99) AND (titles.type LIKE '%business%'))	¿Qué libros tienen precio mayor a $19.99 y son de tipo business?	SELECT titles.title, titles.title_id FROM titles WHERE ((titles.price > 19.99) AND (titles.type LIKE '%business%'))
18	Muestra la editorial que se encuentra en el	SELECT publishers.pub_id, publishers.pub_name FROM publishers	Que editorial se encuentra en el estado	SELECT publishers.pub_id, publishers.pub_name FROM publishers

	estado "CA"	WHERE ((publishers.country LIKE '%CA%' OR publishers.state LIKE '%CA%'))	CA	WHERE ((publishers.country LIKE '%CA%' OR publishers.state LIKE '%CA%'))
20	Dame el autor de "the busy"	SELECT authors.au_fname, authors.au_id, titleauthor.au_id, authors.au_lname FROM authors, titleauthor, titles WHERE ((titles.title LIKE '%the_busy%')) AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id	¿Cual es el autor "the busy"?	SELECT authors.au_fname, authors.au_id, titleauthor.au_id, authors.au_lname FROM authors, titleauthor, titles WHERE ((titles.title LIKE '%the_busy%')) AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id
21	Selecciona todos los libros de Dull	SELECT titles.title, titleauthor.title_id, titles.title_id FROM authors, titleauthor, titles WHERE ((authors.au_lname LIKE 'Dull')) AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id	¿Cuáles son los libros de Dull?	SELECT titles.title, titleauthor.title_id, titles.title_id FROM authors, titleauthor, titles WHERE ((authors.au_lname LIKE 'Dull')) AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id
23	Dame los libros del autor green	SELECT titleauthor.au_id, titleauthor.title_id FROM authors, titleauthor WHERE ((authors.au_lname LIKE 'green')) AND titleauthor.au_id = authors.au_id	¿Cuáles son los libros del autor green?	SELECT titleauthor.au_id, titleauthor.title_id FROM authors, titleauthor WHERE ((authors.au_lname LIKE 'green')) AND titleauthor.au_id = authors.au_id
25	Dame el puesto que tiene el empleado Francisco Chang	SELECT jobs.job_desc, employee.job_id, jobs.job_id FROM employee, jobs WHERE ((employee.fname LIKE 'Francisco') OR (employee.lname LIKE 'Chang')) AND employee.job_id = jobs.job_id	¿Que puesto tiene el empleado Francisco Chang?	SELECT jobs.job_desc, employee.job_id, jobs.job_id FROM employee, jobs WHERE ((employee.fname LIKE 'Francisco') OR (employee.lname LIKE 'Chang')) AND employee.job_id = jobs.job_id
27	Dame el puesto y fecha de contratación que tiene el empleado Paolo?	SELECT employee.hire_date, jobs.job_desc, employee.job_id, jobs.job_id FROM employee, jobs WHERE ((employee.fname LIKE 'Paolo')) AND employee.job_id = jobs.job_id	¿Que puesto y fecha de contratación tiene el empleado Paolo?	SELECT employee.hire_date, jobs.job_desc, employee.job_id, jobs.job_id FROM employee, jobs WHERE ((employee.fname LIKE 'Paolo')) AND employee.job_id = jobs.job_id
29	Dame la dirección del almacen Barnum's	SELECT stores.stor_address FROM stores WHERE ((stores.stor_name LIKE 'Barnum''s'))	Cual es la dirección del almacen Barnum's	SELECT stores.stor_address FROM stores WHERE ((stores.stor_name LIKE 'Barnum''s'))
30	Dame el numero de teléfono del autor cheryl	SELECT authors.phone FROM authors WHERE ((authors.au_fname LIKE 'cheryl'))	¿Cual es el numero de teléfono del autor cheryl?	SELECT authors.phone FROM authors WHERE ((authors.au_fname LIKE 'cheryl'))
31	Dame la ciudad de la editorial "New Moon Books"	SELECT publishers.city FROM publishers WHERE ((publishers.pub_name LIKE 'New_Moon_Books'))	¿Cual es la ciudad de la editorial "New Moon Books"?	SELECT publishers.city FROM publishers WHERE ((publishers.pub_name LIKE 'New_Moon_Books'))
33	Dame el nivel de philip cramer	SELECT employee.job_lvl FROM employee WHERE ((employee.fname LIKE 'philip') OR (employee.lname LIKE 'cramer'))	¿Cual es el nivel de philip cramer?	SELECT employee.job_lvl FROM employee WHERE ((employee.fname LIKE 'philip') OR (employee.lname LIKE 'cramer'))
34	Dame el precio del libro con identificador de	SELECT titles.price FROM pub_info, publishers, titles WHERE ((pub_info.pub_id = '1389' OR	¿Cual es el precio del libro con identificador	SELECT titles.price FROM pub_info, publishers, titles WHERE ((pub_info.pub_id = '1389' OR

	editorial 1389	publishers.pub_id = '1389' OR titles.pub_id = '1389')) AND titles.pub_id = publishers.pub_id AND publishers.pub_id = pub_info.pub_id	de editorial 1389?	publishers.pub_id = '1389' OR titles.pub_id = '1389')) AND titles.pub_id = publishers.pub_id AND publishers.pub_id = pub_info.pub_id
35	Lista los libros de la editorial GGG&G	SELECT titles.pub_id FROM pub_info, publishers, titleauthor, titles WHERE ((pub_info.pr_info LIKE '%GGG&G%' OR publishers.pub_name LIKE 'GGG&G')) AND titleauthor.title_id = titles.title_id AND titles.pub_id = publishers.pub_id AND publishers.pub_id = pub_info.pub_id	¿Cuales son los libros de la editorial GGG&G?	SELECT titles.pub_id FROM pub_info, publishers, titleauthor, titles WHERE ((pub_info.pr_info LIKE '%GGG&G%' OR publishers.pub_name LIKE 'GGG&G')) AND titleauthor.title_id = titles.title_id AND titles.pub_id = publishers.pub_id AND publishers.pub_id = pub_info.pub_id
36	Dame el identificador y el precio del libro "you can"	SELECT titles.price, titles.pub_id, titles.title_id FROM titles WHERE ((titles.title LIKE '%you_can%'))	¿Cual es el identificador y el precio del libro "you can"?	SELECT titles.price, titles.pub_id, titles.title_id FROM titles WHERE ((titles.title LIKE '%you_can%'))
37	Dame la dirección y el teléfono del autor del libro "you can"	SELECT authors.address, authors.phone FROM authors, titleauthor, titles WHERE ((titles.title LIKE '%you_can%')) AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id	¿Cual es la dirección y el teléfono del autor del libro "you can"?	SELECT authors.address, authors.phone FROM authors, titleauthor, titles WHERE ((titles.title LIKE '%you_can%')) AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id
38	Dame la ciudad en donde se encuentra el autor Johnson White	SELECT authors.city FROM authors WHERE ((authors.au_fname LIKE 'Johnson') OR (authors.au_lname LIKE 'White'))	¿En que ciudad se encuentra el autor Johnson White?	SELECT authors.city FROM authors WHERE ((authors.au_fname LIKE 'Johnson') OR (authors.au_lname LIKE 'White'))
39	Dame el apellido que tiene el empleado Pedro	SELECT employee.fname, employee.lname FROM employee WHERE ((employee.fname LIKE 'Pedro'))	¿Que apellido tiene el empleado Pedro?	SELECT employee.fname, employee.lname FROM employee WHERE ((employee.fname LIKE 'Pedro'))
40	Dame la ciudad en donde se encuentra el almacen Bookbeat	SELECT stores.city FROM stores WHERE ((stores.stor_name LIKE 'Bookbeat'))	¿En que ciudad se encuentra el almacen Bookbeat?	SELECT stores.city FROM stores WHERE ((stores.stor_name LIKE 'Bookbeat'))
42	Dame el estado en donde se encuentra la editorial "Ramona publishers"	SELECT publishers.country, publishers.state FROM pub_info, publishers WHERE ((pub_info.pr_info LIKE '%Ramona_publishers%' OR publishers.pub_name LIKE 'Ramona_publishers')) AND publishers.pub_id = pub_info.pub_id	¿En que estado se encuentra la editorial "Ramona publishers"?	SELECT publishers.country, publishers.state FROM pub_info, publishers WHERE ((pub_info.pr_info LIKE '%Ramona_publishers%' OR publishers.pub_name LIKE 'Ramona_publishers')) AND publishers.pub_id = pub_info.pub_id
44	Titulo de los libros cuyos editores se encuentran en TX	SELECT titles.title FROM publishers, titles WHERE ((publishers.state LIKE 'TX')) AND titles.pub_id = publishers.pub_id	¿Cuál es el Titulo de los libros cuyos editores se encuentran en TX?	SELECT titles.title FROM publishers, titles WHERE ((publishers.state LIKE 'TX')) AND titles.pub_id = publishers.pub_id
45	Dame el nombre y dirección del empleado que trabaja para la	SELECT employee.fname, employee.lname FROM employee, pub_info, publishers WHERE ((pub_info.pr_info LIKE '%GGG&G%' OR publishers.pub_name LIKE 'GGG&G')) AND	¿Que nombre y dirección tiene el empleado que trabaja	SELECT publishers.pub_name FROM pub_info, publishers WHERE ((pub_info.pr_info LIKE '%GGG&G%' OR publishers.pub_name LIKE 'GGG&G')) AND

	editorial GGG&G	employee.pub_id = publishers.pub_id AND publishers.pub_id = pub_info.pub_id	para la editorial GGG&G?	publishers.pub_id = pub_info.pub_id
46	Dame la descripción del puesto del empleado VPA30890F	SELECT jobs.job_desc FROM employee, jobs WHERE ((employee.emp_id LIKE 'VPA30890F')) AND employee.job_id = jobs.job_id	¿Que descripción del puesto del empleado VPA30890F?	SELECT jobs.job_desc FROM employee, jobs WHERE ((employee.emp_id LIKE 'VPA30890F')) AND employee.job_id = jobs.job_id
47	Obtener el nombre del almacén donde se encuentra el libro "cooking with"	SELECT stores.stor_name FROM stores, titles, sales WHERE ((titles.title LIKE '%cooking_with%')) AND sales.title_id = titles.title_id AND sales.stor_id = stores.stor_id	¿Cuál es el nombre del almacén donde se encuentra el libro "cooking with"?	SELECT stores.stor_name FROM stores, titles, sales WHERE ((titles.title LIKE '%cooking_with%')) AND sales.title_id = titles.title_id AND sales.stor_id = stores.stor_id
48	Dame la fecha en que se realizo el contrato del empleado PTC11962M	SELECT employee.hire_date FROM employee WHERE ((employee.emp_id LIKE 'PTC11962M'))	¿En que fecha se realizo el contrato del empleado PTC11962M?	SELECT employee.hire_date FROM employee WHERE ((employee.emp_id LIKE 'PTC11962M'))

Tipo 14 Columnas implícitas y tablas explícitas con funciones SQL y con condición

49	Dame la cantidad de autores de la ciudad de Berkeley	NO	¿Cuantos autores son de la ciudad de Berkeley?	NO
50	Dame el numero de empleados de la editorial "Scotney book"	NO	¿Cual es el numero de empleados de la editorial "Scotney book"?	NO
51	Dame el numero de ventas realizadas el 14/09/1994	NO	¿Cual es el numero de ventas realizadas el 14/09/1994?	NO
52	Dame el numero de libros vendidos el 13/19/1994	NO	¿ Cuál es el numero de libros vendidos el 13/19/1994?	NO

Tipo 3 Columnas explícitas y tablas implícitas sin funciones SQL y sin condición

53	Lista los identificadores que tienen los libros	SELECT titleauthor.au_id, titles.pub_id, titles.title, titleauthor.title_id, titles.title_id FROM titleauthor, titles WHERE titleauthor.title_id = titles.title_id	¿Cuáles son los identificadores de los libros?	SELECT titleauthor.title_id, titles.title_id FROM titleauthor, titles WHERE titleauthor.title_id = titles.title_id

Tipo 11 Columnas explícitas y tablas implícitas sin funciones SQL y con condición

54	Dame la ciudad a la que pertenece el código postal 89076	SELECT authors.city, stores.city FROM authors, stores, sales, titles, titleauthor WHERE ((authors.zip = '89076' OR stores.zip = '89076')) AND sales.stor_id = stores.stor_id AND sales.title_id = titles.title_id AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id	¿A que ciudad pertenece el código postal 89076?	SELECT authors.city, stores.city FROM authors, stores, sales, titles, titleauthor WHERE ((authors.zip = '89076' OR stores.zip = '89076')) AND sales.stor_id = stores.stor_id AND sales.title_id = titles.title_id AND titleauthor.title_id = titles.title_id AND titleauthor.au_id = authors.au_id
56	Muestra el adelanto del numero de orden 6871	SELECT titles.advance, sales.payterms FROM sales, titles WHERE ((titles.advance = 6871 OR sales.payterms = '6871')) AND sales.title_id = titles.title_id	¿Cual es el adelanto del numero de orden 6871?	SELECT titles.advance, sales.payterms FROM sales, titles WHERE ((titles.advance = 6871 OR sales.payterms = '6871')) AND sales.title_id = titles.title_id
57	Dame la fecha de contratación de pedro	SELECT employee.hire_date FROM employee WHERE ((employee.fname LIKE 'Pedro'))	¿Cuál es la fecha de contratación de pedro?	SELECT employee.hire_date FROM employee WHERE ((employee.fname LIKE 'Pedro'))

Tipo 4 Columnas y tablas implícitas sin funciones SQL y sin condición

58	Dame los libros que contiene cada editorial	SELECT pub_info.pub_id, publishers.pub_id, titles.pub_id, publishers.pub_name, titles.title, titleauthor.title_id, titles.title_id FROM pub_info, publishers, titleauthor, titles WHERE titleauthor.title_id = titles.title_id AND titles.pub_id = publishers.pub_id AND publishers.pub_id = pub_info.pub_id	¿Cuáles son los libros y su editorial?	SELECT pub_info.pub_id, publishers.pub_id, titles.pub_id, publishers.pub_name, titles.title, titleauthor.title_id, titles.title_id FROM pub_info, publishers, titleauthor, titles WHERE titleauthor.title_id = titles.title_id AND titles.pub_id = publishers.pub_id AND publishers.pub_id = pub_info.pub_id

Tipo 12 Columnas y tablas implícitas sin funciones SQL y con condición

59	Muestra la cantidad de ventas de "silicon valley"	SELECT sales.qty FROM sales, titles WHERE ((titles.title LIKE '%silicon_valley%')) AND sales.title_id = titles.title_id	¿Que cantidad de ventas de "silicon valley"?	SELECT sales.qty FROM sales, titles WHERE ((titles.title LIKE '%silicon_valley%')) AND sales.title_id = titles.title_id
60	Dame la editorial en donde trabaja victoria ashworth	SELECT pub_info.pub_id, publishers.pub_id, publishers.pub_name FROM employee, pub_info, publishers WHERE ((employee.fname LIKE 'victoria') OR (employee.lname LIKE 'ashworth')) AND employee.pub_id = publishers.pub_id AND publishers.pub_id = pub_info.pub_id	¿En que editorial trabaja victoria ashworth?	SELECT pub_info.pub_id, publishers.pub_id, publishers.pub_name FROM employee, pub_info, publishers WHERE ((employee.fname LIKE 'victoria') OR (employee.lname LIKE 'ashworth')) AND employee.pub_id = publishers.pub_id AND publishers.pub_id = pub_info.pub_id
61	Nombre del almacen	SELECT stores.stor_name FROM stores, titles, sales	¿Cuál es el nombre del	SELECT stores.stor_name FROM stores, titles, sales

	donde se encuentra “the busy”	WHERE ((titles.title LIKE '%the_busy%')) AND sales.title_id = titles.title_id AND sales.stor_id = stores.stor_id	almacen donde se encuentra “the busy”?	WHERE ((titles.title LIKE '%the_busy%')) AND sales.title_id = titles.title_id AND sales.stor_id = stores.stor_id

Tipo 16 Columnas y tablas implícitas sin funciones SQL y con condición

62	Obtén los números de ejemplares que tiene el libro “the busy”	SELECT titles.title, titles.title_id, titles.type FROM titles WHERE ((titles.title LIKE '%the_busy%'))	¿Cuántos números de ejemplares tiene el libro “the busy”?	SELECT titles.title, titles.title_id, titles.type FROM titles WHERE ((titles.title LIKE '%the_busy%'))