

**Universidad de las Ciencias Informáticas**



# **Método para la desambiguación sintáctica de textos de la legislación cubana**

**Tesis para optar por el grado científico de:  
Máster en Informática Avanzada**

**Autor:**

Ing. Yordanis García Leiva

**Tutores:**

Dr. C. Juan Pedro Febles Rodríguez

MsC. Yarina Amoroso Fernández

**La Habana, 2016**

**“Año 58 del Triunfo de la Revolución”**

## DECLARACIÓN DE AUTORÍA

Declaro que soy el único autor de este trabajo y autorizo a la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firma la presente a los \_\_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

\_\_\_\_\_  
Ing. Yordanis García Leiva

\_\_\_\_\_  
Dr. C. Juan Pedro Febles Rodríguez

\_\_\_\_\_  
MsC. Yarina Amoroso Fernández

**AGRADECIMIENTOS**

*A nuestro eterno Comandante en Jefe, por ser faro y guía de la juventud cubana y creador de esta majestuosa universidad.*

*A mi mamá, mi papá, mi hermana y mi abuela, por ser la familia más bella del mundo y darme ánimo en todo momento.*

*A mis tutores, que sin su ayuda no hubiera sido posible este logro, profe Yarina gracias por siempre confiar en mí y a Febles gracias por estar siempre ahí, cuando lo necesitaba.*

*A mis segundos tutores y amigos, Reinier Silverio, Isabel y Danae gracias por la ayuda.*

*A ese hermano de la carrera y para toda la vida, Héctor Luis y su familia.*

*A los verdaderos amigos con los que la UCI me ha premiado Leosnavy, Dayanis, Maikel, Dunia, Orlando, Jorlen, Juan David, Sandy, Cervela, Nailin, Elizabeth, Doris, Yennis, Berenice, Marilé, Robin, Yamila y Mailen Edith.*

*A mis compañeros de apartamento, Arnel, Yulia y Yosvany*

*Al equipo de trabajo del laboratorio 419.*

*A los miembros del consejo de dirección de CEGEL.*

*A los profesores del claustro de la maestría.*

*A mi familia de Cojimar y a Leo, Yanet y esos chiquitines que tienen.*

*Gracias a todos los que de una forma u otra han tenido que ver con este resultado.*

DEDICATORIA

*A la memoria de nuestro eterno Comandante en Jefe, por la idea de crear esta  
majestuosa universidad.*

*A alguien que siempre tendré presente y sé que cada día ilumina mi camino  
donde quiera que esté, mi abuelo Polo.*

*A dos personas que son el faro que me guía todos los días mi camino, mami y  
papi.*

*A esa segunda fuente inspiradora de mis resultados, mi hermanita.*

*A mi abuela por siempre estar al tanto de mis resultados.*

*A esa larga lista de amigos que la vida y la UCI me han permitido crear.*

*A todas las personas que de una forma u otra han contribuido a este resultado.*

**RESUMEN**

La ambigüedad es el término que hace referencia a las estructuras gramaticales que pueden entenderse de varios modos y admitir distintas interpretaciones. Por otra parte, el término desambiguación se refiere a las operaciones necesarias a realizar sobre un texto con el propósito de reducir el nivel de ambigüedad que este pueda presentar. En la actualidad existen técnicas que permiten reducir estos problemas, definidas a través del Procesamiento del Lenguaje Natural, disciplina que combina la lingüística y la informática, con el fin de modelar el lenguaje humano desde el punto de vista computacional. Las soluciones que aplican estas técnicas, tienen como deficiencias: el alto consumo de recursos computacionales y que están desarrolladas para dominios e idiomas específicos, que en pocas ocasiones incluyen el español y el análisis de textos jurídicos, siendo estos últimos el dominio de aplicación de la presente investigación. El método propuesto se desarrolla a partir del uso de técnicas del Procesamiento del Lenguaje Natural y se organiza en tres partes. En el diseño de este se definen reglas heurísticas, gramáticas y un lexicón, con la finalidad de ser aplicados en la identificación de ambigüedades sintácticas presentes en textos jurídicos y brindar información al usuario, que les permita reducir los problemas detectados y elevar la comprensión de estos textos. El resultado fue validado por expertos en informática y Derecho, haciendo uso de métodos cualitativos y cuantitativos, demostrándose en todos los casos buena aceptación y satisfacción de la solución propuesta.

**PALABRAS CLAVES:** ambigüedad, desambiguación, gramáticas, lexicón, método, Procesamiento del Lenguaje Natural, reglas heurísticas

**ÍNDICE DE CONTENIDOS**

<b>INTRODUCCIÓN.....</b>	<b>10</b>
<b>CAPÍTULO 1: Ambigüedad y desambiguación de textos.....</b>	<b>17</b>
1.1 Marco conceptual de la investigación .....	17
1.1.1 Conjunciones .....	20
1.1.2 Preposiciones .....	22
1.1.3 Estructura gramatical de los textos .....	23
1.2 Métodos para la desambiguación de textos .....	25
1.3 Soluciones informáticas para la desambiguación de textos .....	26
1.4 Conclusiones parciales .....	29
<b>CAPÍTULO 2: DIAGNÓSTICO Y PROPUESTA DE SOLUCIÓN.....</b>	<b>31</b>
2.1 Diagnóstico sobre soluciones informáticas para reducir la ambigüedad sintáctica en textos de la legislación cubana .....	31
2.2 Descripción del método propuesto.....	33
2.2.1 Identificación de posibles ambigüedades sintácticas.....	34
2.2.2 Procesamiento de posibles ambigüedades sintácticas.....	41
2.2.3 Brindar información al usuario.....	50
2.3 Conclusiones parciales .....	52
<b>CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA DE SOLUCIÓN .....</b>	<b>53</b>
3.1. Cuasi Experimento .....	53
3.2. Grupo focal.....	57
3.3. Escalamiento de Osgood.....	61
3.4. Triangulación de resultados.....	63
3.5. Conclusiones parciales.....	65
<b>CONCLUSIONES GENERALES .....</b>	<b>67</b>
<b>RECOMENDACIONES.....</b>	<b>68</b>
<b>REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>69</b>
<b>ANEXOS .....</b>	<b>74</b>

**ÍNDICE DE FIGURAS**

Figura 1. Resultados del diagnóstico. Fuente: elaboración propia ..... 32

Figura 2. Flujo para el desarrollo del método de desambiguación sintáctica propuesto.  
Fuente: elaboración propia ..... 33

Figura 3. Actividades de la primera parte del método de desambiguación sintáctica  
propuesto. Fuente: elaboración propia..... 35

Figura 4. GLC definida para el método propuesto. Fuente: elaboración propia..... 36

Figura 5. Lexicón definido para el método propuesto. Fuente: elaboración propia ..... 37

Figura 6. Flujo a seguir para validar la correcta estructuración gramatical de los textos,  
utilizando el lexicón y la GLC. Fuente: elaboración propia ..... 38

Figura 7. Comparación entre la lista de etiquetas extraídas de una oración y la GLC.  
Fuente: elaboración propia ..... 39

Figura 8. Pasos y condiciones para la ejecución de la segunda parte del método  
propuesto. Fuente: elaboración propia..... 42

Figura 9. Flujo a seguir a partir de la implementación de la regla heurística 4. Fuente:  
elaboración propia ..... 43

Figura 10. Aplicación de la regla heurística 4 en un texto. Fuente: elaboración propia .... 44

Figura 11. Flujo a seguir para señalar en un texto, el sentido de las preposiciones  
sintácticamente ambiguas. Fuente: elaboración propia..... 47

Figura 12. Representación a través de una gramática del sentido de las preposiciones  
sintácticamente ambiguas. Fuente: elaboración propia..... 48

Figura 13. Proceso de identificación de un sustantivo que precede una preposición  
sintéticamente ambigua en un texto. Fuente: elaboración propia..... 49

Figura 14. Sentidos que atribuye el sustantivo sala a las preposiciones sintácticamente  
ambiguas. Fuente: elaboración propia ..... 50

Figura 15. Señalización de posibles causas de ambigüedades en un texto. Fuente:  
elaboración propia ..... 51

Figura 16. Resultado de aplicar el tercer paso del método propuesto. Fuente: elaboración  
propia ..... 51

Figura 17. Nivel de incoherencia. Fuente: elaboración propia..... 56

Figura 18. Nivel de legibilidad. Fuente: elaboración propia ..... 56

## ÍNDICE DE FIGURAS

Figura 19. Evaluación de la comprensión del texto. Fuente: elaboración propia .....	57
Figura 20. Niveles de satisfacción de los usuarios en la escala de Osgood. Fuente: elaboración propia .....	63



**ÍNDICE DE TABLAS**

Tabla 1. Conjunciones coordinantes y sus clasificaciones. Fuente: (Torrego, 2007)..... 21

Tabla 2. Preposiciones y sentidos en una oración. Fuente: (Slager, 2004) ..... 22

Tabla 3. Resultados de aplicar el cuasi experimento. Fuente: elaboración propia ..... 55

Tabla 4. Expertos seleccionados para aplicar Grupo focal. Fuente: elaboración propia... 59

Tabla 5. Subgrupos de debate para aplicar Grupo focal. Fuente: elaboración propia ..... 60

Tabla 6. Resumen de las valoraciones de escala realizadas por los usuarios sobre la aplicación del método de desambiguación sintáctica propuesto. Fuente: elaboración propia ..... 62

Tabla 7. Resumen de la triangulación de métodos. Fuente: elaboración propia..... 64

## INTRODUCCIÓN

El enriquecido número de fenómenos lingüísticos y reglas gramaticales que caracterizan los lenguajes naturales<sup>1</sup> complejizan en ocasiones el entendimiento de determinados escritos, originando ambigüedad en la comprensión e interpretación de los mismos. La ambigüedad en un escrito se identifica cuando a través del análisis de un contexto no es posible determinar el significado de una palabra o una oración en general. La existencia de textos ambiguos da lugar a la ocurrencia de incertidumbre, duda e indefinición en la comprensión de un contexto.

La ambigüedad en los textos parte de las características del lenguaje natural, en el cual existen múltiples expresiones y palabras que pueden tener diferentes significados, en dependencia de las circunstancias de uso. Este problema puede tratarse desde distintas perspectivas, desde la ambigüedad debida a palabras polisémicas<sup>2</sup>, hasta la ambigüedad producida por las diferentes interpretaciones que pueda tener una oración (Pérez, 2009).

La ambigüedad puede estar presente en textos de diferentes orígenes. Los textos jurídicos no están exceptos de la existencia de este fenómeno que interfiere en la comprensión de los mismos.

El problema de las lagunas legales o inconsistencias en la legislación es antiguo y continúa existiendo. La solución tradicional al mismo consiste en dejar en manos de los jueces, cotidianos usuarios de la legislación, la interpretación de las normas para subsanar estos problemas. Sin embargo, abundan casos donde, a pesar de la jurisprudencia y otros elementos con los que cuenta el poder judicial para lograr ecuanimidad, existe cierto grado de ambigüedad en las normas (Sznur, 2012).

Los textos jurídicos se caracterizan por el empleo de (Monteagudo, 2013):

---

<sup>1</sup>Lenguaje Natural: Es la lengua o idioma hablado o escrito por humanos para propósitos generales de comunicación (Hernández, 2000).

<sup>2</sup>Palabras que tienen varios significados, ejemplo Sierra es una *Herramienta para cortar madera* y a la vez significa *Cordillera de montañas de picos afilados* (Gil, 1990).

- tecnicismos, cultismos<sup>3</sup> y términos arcaizantes<sup>4</sup>
- uso de expresiones lexicalizadas<sup>5</sup>
- el futuro de subjuntivo, tanto simple como compuesto (si entre los herederos hubiere; si al *expirar dicho plazo hubiere cumplido*)
- los pronombres átonos pospuestos (*determinase el pago*)
- una sintaxis enrevesada con una clara tendencia a alargar los periodos sintácticos y a usar abundantemente los gerundios.

Cada una de estas características hace que los textos jurídicos se caractericen por ser ambiguos. Así lo han avalado en las últimas décadas, algunos autores reconocidos en el estudio del tema, definiendo que:

Los textos jurídicos y su complejidad es un tema que se ha abordado desde diversas perspectivas, llegando en la mayoría de los casos a resaltar los problemas de comprensión que se generan como resultado de su opacidad<sup>6</sup>. Esta opacidad es producto de una sintaxis denominada por muchos como enmarañada, arcaizante y barroca (Cucatto, 2013).

El lenguaje jurídico plantea diversos problemas de naturaleza semántica y sintáctica. La mayor parte de los textos jurídicos son lingüísticamente ambiguos, porque el lenguaje utilizado es en sí mismo ambiguo (Maña, 2014).

La ambigüedad que caracteriza a los textos jurídicos, ha dado origen al desarrollo en diferentes partes del mundo de soluciones informáticas que contribuyen a la interpretación y comprensión de documentos legislativos. Ejemplo: los trabajos relacionados con el tema de la reducción de la ambigüedad en textos jurídicos, desarrollados principalmente en centros vinculados a la Universidad de Alicante en España (Aarnio, 1987). En el área de América Latina se encuentran, los métodos de interpretación jurídica, desarrollados en la Biblioteca Jurídica Virtual del Instituto de Investigaciones Jurídicas de la Universidad

---

<sup>3</sup>Palabra cuya morfología sigue muy estrechamente su origen etimológico griego o latino, sin obedecer los cambios que la evolución de la lengua castellana siguió a partir de su origen en el latín (Wijnakker, 2011).

<sup>4</sup> Que tiende a lo antiguo (Arte, 1998).

<sup>5</sup>Representan expresiones que no tienen una definición precisa y su sentido exacto puede variar según el autor que las emplee (Calvo, 2009).

<sup>6</sup>Poco claro (RAE, 2016).

Nacional Autónoma de México (UNAM), los cuales facilitan una mejor interpretación de la Ley mexicana (Paredes, 2011); así como el análisis de consistencia realizado mediante métodos formales a la Legislación de Defensa del Consumidor en Argentina, este trabajo fue realizado por un grupo de investigación del Departamento de Computación de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires (UBA) y tiene como propósito la detección de inconsistencia basados en lógica al ámbito legal, que permita reducir la ambigüedad y mejorar la comprensión de esta ley argentina (Sznur, et al, 2015).

Es importante señalar que estos trabajos solo han llegado a un nivel teórico, pues constituyen guías de buenas prácticas e instrucciones metodológicas para la interpretación de una norma. En ninguno de los casos se ofrece una solución práctica que haga posibles a través de un software el tratamiento de la ambigüedad en textos legales.

En Cuba existen pocos avances en el diseño e implementación de soluciones informáticas que posibiliten mejorar la comprensión de textos jurídicos y reducir la ambigüedad existente en los mismos. La Sociedad Cubana de Derecho e Informática, organización académica que contribuye al desarrollo de la informática jurídica en Cuba, ha realizado algunos trabajos en relación al tema, los cuales solo llegan a un nivel teórico, ejemplo de esto lo constituye la investigación sobre similaridad en textos jurídicos, desarrollada por la Dra. Marta Blanquier Ascano (Ascaño, 2015), profesora de la Facultad de Matemática y Computación de la Universidad de La Habana y miembro de la organización antes mencionada.

Como resultado de las características propias de los textos jurídicos, los documentos legislativos en Cuba se caracterizan por la existencia de fenómenos lingüísticos como las palabras polisémicas, provocando que, a la hora de analizar un escrito, el mismo pueda tener más de una interpretación, dando lugar a incertidumbre en su comprensión; es decir, no se tiene una visión clara del significado que puede tener el contexto analizado. Además, provoca inconsistencia en el análisis de los contenidos, pues no existe una comprensión exacta de lo que realmente se está leyendo, a menos que sea un especialista en el tema, por tanto, el resultado del análisis de los textos es bajo, provocando confusiones en su interpretación.

Todo lo expuesto anteriormente sustenta la **contradicción fundamental** de la presente investigación, expresada en la necesidad de mejorar la comprensión de textos jurídicos a

partir de su naturaleza ambigua, teniendo en cuenta la falta de procedimientos, métodos y herramientas que existe para esclarecer los términos ambiguos presentes en los mismos, con el propósito de que las personas no especializadas en temas legales puedan tener una comprensión más clara de lo que están leyendo.

La **situación problemática** antes descrita permite identificar el siguiente **problema científico**:

¿Cómo reducir la ambigüedad sintáctica presente en textos de la legislación cubana contribuyendo a la mejora en la comprensión de los mismos?

A partir del problema de científico antes propuesto se define como **objeto de estudio**: el proceso de desambiguación de textos.

Para ello se identifica como **campo de acción**: la desambiguación sintáctica de textos de la legislación cubana.

Determinándose como **objetivo general**: desarrollar un método basado en Procesamiento del Lenguaje Natural que permita reducir la ambigüedad sintáctica presente en textos de la legislación cubana, contribuyendo a la mejora en la comprensión de los mismos.

Del objetivo general se desglosan los siguientes **objetivos específicos**:

- Elaborar el marco teórico referencial de la investigación en función de los términos relacionados con la ambigüedad y métodos de desambiguación de textos.
- Diagnosticar la situación actual del objeto de estudio en el sistema judicial cubano, buscando el nivel de informatización que existe en la esfera jurídica en relación al tema de investigación.
- Desarrollar un método basado en Procesamiento del Lenguaje Natural, que permita reducir la ambigüedad sintáctica presente en textos de la legislación cubana.
- Validar la solución propuesta a través de métodos científicos.

Teniendo en cuenta lo anteriormente planteado se define la siguiente **hipótesis**:

El desarrollo de un método basado en Procesamiento del Lenguaje Natural permitirá reducir la ambigüedad sintáctica presente en textos de la legislación cubana, contribuyendo a la mejora en la comprensión de los mismos.

A continuación, se explican los **métodos de investigación** utilizados en la tesis:

### **Métodos teóricos:**

**Analítico – sintético:** el análisis fue utilizado para descomponer el problema de investigación en elementos por separado que permiten profundizar en el estudio del proceso de desambiguación de textos legales. La síntesis permite la unión de las partes previamente analizadas, descubrir las relaciones esenciales que existen entre estas y sintetizar todo como una solución integrada.

**Inductivo – deductivo:** la inducción permitió realizar un razonamiento que parte de lo particular a lo general, reflejando en el resultado final lo común que existe en cada bibliografía estudiada. La deducción permite a través de los conocimientos generales adquiridos durante la investigación inferir otros conocimientos lógicos como las conclusiones parciales y generales de la investigación.

**Hipotético – deductivo:** este método fue utilizado en la definición y verificación de la hipótesis de la investigación y en la inferencia de las conclusiones parciales y generales del trabajo. Además, permite inferir nuevas líneas de investigación a partir de los resultados parciales y generales obtenidos.

**Análisis histórico – lógico:** se utilizó para identificar las tendencias actuales en relación al estado de los procesos de desambiguación y mejora en la comprensión de textos legales, permitiendo conocer y estudiar los autores más reconocidos y los trabajos más importantes realizados hasta el momento, vinculados con el tema de investigación.

**Modelación:** se empleó con el propósito de descubrir nuevos conocimientos sobre el proceso de desambiguación de textos, utilizando diagramas y figuras que representan los resultados de la investigación y ayudan a comprender los mismos.

### **Métodos empíricos:**

**Entrevista:** se aplicó a especialistas en derecho (juristas), con el objetivo de obtener información verbal y real de cómo se manifiesta la presencia de términos ambiguos en la legislación cubana y los efectos que estos traen para su comprensión.

**Encuesta:** se instrumentó en las etapas de diagnóstico y validación de los resultados de la investigación. Durante el diagnóstico se aplicó con el propósito de obtener información sobre el estado actual del objeto de estudio en el sistema judicial cubano. En la etapa de validación hizo posible recolectar los datos necesarios para aplicar los métodos utilizados en el capítulo 3.

**Escala de Osgood:** se aplicó con el propósito de evaluar las opiniones de los usuarios en relación a la aplicación del método de desambiguación sintáctica propuesto.

**Experimental:** se aplicó un cuasi experimento de tipo pre prueba-post prueba con grupos intactos con el propósito de constatar la relación causa efecto entre la variable independiente sobre la dependiente.

**Grupo focal:** se aplicó con el propósito de evaluar la utilidad del método de desambiguación sintáctica propuesto y obtener recomendaciones que contribuyan a su perfeccionamiento.

### **Aporte práctico:**

El desarrollo del trabajo tiene como aporte práctico la obtención de un método para la desambiguación sintáctica en textos legales, que, al ser implementado y aplicado a través de una solución informática, brinda una propuesta de desambiguación sintáctica al usuario o detalla las causas que originan los problemas de ambigüedad sintáctica identificados en el texto analizado. El resultado de la tesis constituye un avance dentro del campo de la informática jurídica en Cuba, en específico en la rama de la legimática<sup>7</sup>, que está reconocida como parte instrumental de las ciencias de la legislación.

---

<sup>7</sup> Ciencia que realiza el tratamiento de información legislativa (Soares, 2008).

### **Estructura de la tesis:**

El presente trabajo está estructurado en las siguientes partes:

- **Introducción:** se fundamenta el valor científico del problema y se recoge el diseño teórico y metodológico de la investigación.
- **Capítulo I:** se exponen los fundamentos teóricos referenciales en función de los términos relacionados con la ambigüedad y desambiguación sintáctica de textos. Se realiza un análisis de los métodos y soluciones de desambiguación existentes y se justifican las razones por las cuales se toma como referencia el método seleccionado para el diseño de la solución propuesta.
- **Capítulo II:** se describe el método propuesto para la desambiguación sintáctica de textos de la legislación cubana. Además, se realiza un diagnóstico sobre el estado actual del objeto de estudio de la presente investigación en el sistema judicial cubano.
- **Capítulo III:** se aplican los métodos científico Cuasi experimento, Grupo focal y Escalamiento de Osgood, con el propósito de comprobar la validez y utilidad del método de desambiguación sintáctica propuesto.

Luego se presentan las **conclusiones, recomendaciones, referencias bibliográficas** y los **anexos**.



## **CAPÍTULO 1: Ambigüedad y desambiguación de textos**

En el capítulo se referencian las principales definiciones asociadas al objeto estudio y campo de acción de la presente investigación, expuestas por expertos en el tema. El análisis de estos conceptos se realiza con el propósito de lograr, por parte del autor y personas que consulten el presente documento, una mejor comprensión de los contenidos abordados en el trabajo, sobre la identificación de ambigüedades en textos y la desambiguación de estos. Otro de los temas abordados en el capítulo es el estudio y análisis de los principales trabajos desarrollados hasta el momento vinculados al objetivo de la investigación, con la intención de resaltar la necesidad de realizar la misma.

### **1.1 Marco conceptual de la investigación**

El desarrollo del marco conceptual de la presente investigación está basado en el análisis de las definiciones expuestas por reconocidos autores en el estudio de la ambigüedad en los textos y el proceso de desambiguación de los mismos, a partir de la bibliografía consultada.

La doctora Sulema Torres Ramos define en su tesis doctoral que la ambigüedad es el término que hace referencia a aquellas estructuras gramaticales que pueden entenderse de varios modos o admitir distintas interpretaciones y dar, por consiguiente, motivo a dudas, incertidumbre o confusión (Torres, 2009).

Por otro lado, el profesor Eneko Agirre, define que la ambigüedad es una característica inherente al lenguaje natural y ocurre cuando una palabra o frase puede tener más de un sentido o interpretación, hecho conocido como polisemia<sup>8</sup> (Agirre & Edmonds, 2006).

En otra de las bibliografías consultadas, el Dr. C. Alexander Gelbukh, plantea que la ambigüedad se presenta en el proceso lingüístico, cuando un término puede admitir distintas interpretaciones a partir de su representación en un texto o cuando existe confusión al tener diversas estructuras y no tener los elementos necesarios para eliminar las incorrectas (Haro & Gelbukh, 2007).

---

<sup>8</sup> Ocurre cuando existen varios significados para una misma palabra (Campos, Simón Cuevas & Olivas, 2014).

## CAPÍTULO 1: Ambigüedad y desambiguación de textos.

Teniendo en cuenta las definiciones anteriores, el autor del presente trabajo considera que la ambigüedad es el resultado de la existencia en un texto de términos con diferentes interpretaciones, dando lugar a incertidumbre, duda o confusión en el análisis de este.

La ambigüedad está dividida en tres tipos fundamentales:

**Ambigüedad léxica:** una misma palabra puede pertenecer a diferentes categorías gramaticales. Ejemplo: La palabra “*corte*”, se puede referir a la forma del verbo cortar o al sustantivo corte de justicia (Pérez, 2009).

**Ambigüedad sintáctica o estructural:** aparece cuando debido a la forma en que se asocian los distintos constituyentes de una oración, esta puede ser interpretada de formas distintas. Siendo a veces casi imposible de solucionar (Pérez, 2009). Ejemplo, en la oración: “Miguel vio a su hermana con unos prismáticos”, se puede entender dos formas diferentes: Miguel uso los prismáticos para ver a su hermana o Miguel vio que su hermana tenía unos prismáticos.

**Ambigüedad semántica:** dentro de este tipo de ambigüedad se diferencian tres clases:

- **Ambigüedad debida a las palabras polisémicas.** En este caso, una misma palabra puede tener distintos significados dependiendo del uso que se le esté dando en cada momento. Ejemplo: “*Entró en el banco. (Se refiere a entidad financiera)*” sin embargo “*Se sentó en el banco del parque. (Se refiere a un asiento)*” (Pérez, 2009).
- **Ambigüedad debida a encontrar una misma estructura sintáctica con diferentes significados.** Ejemplo: “*Todos los estudiantes de secundaria hablan dos lenguas.*” La ambigüedad aquí está en no dejar claro si cada estudiante habla dos lenguas o solo se hablan dos lenguas determinadas (Pérez, 2009).
- **Ambigüedad referencial.** En este caso, es necesario el análisis del texto más allá de los límites de la frase, determinando los antecedentes referenciales de los pronombres. Ejemplo: “*El jamón está en el armario. Sácalo. Ciérralo.*” La ambigüedad aquí está en determinar a quién se hace referencia en las palabras señaladas, si al *jamón* o al *armario* (Pérez, 2009).

## CAPÍTULO 1: Ambigüedad y desambiguación de textos.

En respuesta a la presencia de ambigüedad en un texto, existe el proceso de desambiguación. La desambiguación permite efectuar las operaciones necesarias para que una palabra, frase o texto ambiguo en una oración, disminuya su nivel de ambigüedad (RAE, 2016).

En correspondencia con el proceso de desambiguación de textos se encuentra la técnica de desambiguación del sentido de las palabras (WSD, según las siglas del inglés *Word Sense Disambiguation*). La WSD es el proceso que permite determinar desde el punto de vista computacional el significado de una palabra, para su uso en un contexto dado, a partir de un conjunto de sentidos posibles (Agirre & Edmonds, 2006).

La WSD requiere del uso del Procesamiento de Lenguaje Natural (PLN). Esta disciplina tiene como objetivo, habilitar a las computadoras para que entiendan el texto, procesándolo por su sentido. Para llevar a cabo esta tarea, un sistema de PLN necesita conocer sobre la estructura del lenguaje, la cual se analiza normalmente en cuatro niveles: morfológico, sintáctico, semántico y pragmático. En el nivel morfológico se estudia cómo se construyen las palabras, en el sintáctico cómo combinar las palabras para formar oraciones, en el semántico el significado de las palabras y en el pragmático se estudia cómo el contexto afecta a la interpretación de las oraciones (Torres, 2009).

Al mismo tiempo el profesor Ruslan Mitkov define que el PLN trata todo tipo de fenómenos lingüísticos de forma automática, y se define como una parte esencial de la Inteligencia Artificial que investiga y formula mecanismos que sean computacionalmente efectivos y que faciliten la interacción hombre-máquina (Mitkov, 2012).

Asimismo, la profesora Sonia Vázquez Pérez en su tesis doctoral plantea que, “el Procesamiento del Lenguaje Natural o Lingüística Computacional combina la lingüística y la informática con el fin de modelar el lenguaje humano desde un punto de vista computacional” (Pérez, 2009).

Por otra parte, el profesor Edgar Tello Leal define al PLN como, “un conjunto de técnicas que contextualizan la WSD, tratando de forma automatizada los fenómenos lingüísticos de diversa índole, mediante computadoras” (Leal, 2009).

Una vez analizadas en varias bibliografías las definiciones de PLN, el autor de la presente investigación considera que el PLN constituye la base para aplicar las técnicas de WSD en textos ambiguos.

En el campo del PLN el problema de la ambigüedad puede tratarse desde distintas perspectivas. Desde la ambigüedad debida a palabras polisémicas, hasta la ambigüedad producida por las distintas interpretaciones que pueda tener una oración. Dentro del PLN, se distinguen tres tipos de ambigüedades, la léxica, la sintáctica o estructural y la semántica (Pérez, 2009).

El desarrollo de la presente investigación está basado en el análisis de la ambigüedad sintáctica. La ambigüedad sintáctica, también conocida como estructural, se presenta en oraciones, de tal manera que éstas puedan ser representadas por más de una estructura sintáctica (Torres, 2009). Aparece cuando debido a la forma en que se asocian los distintos constituyentes de una oración, esta puede ser interpretada de diferentes maneras (Pérez, 2009). En resumen, un texto presenta ambigüedad estructural, cuando diferentes interpretaciones para el significado del mismo, dependen de su estructura sintáctica.

Los problemas de ambigüedad estructural en un texto son originados principalmente por determinadas conjunciones y preposiciones presentes en estos o la incorrecta estructura gramatical de los mismos. El análisis de estos elementos constituye la base para el desarrollo del método de desambiguación propuesto en el presente trabajo.

### **1.1.1 Conjunciones**

Las conjunciones son clases de palabras que sintácticamente funcionan como enlaces entre palabras (Mario y María), sintagmas (Mi perro y el tuyo) u oraciones (Luchar para ganar). Morfológicamente son palabras invariables y semánticamente no poseen significado. Las conjunciones se dividen en dos grandes grupos: las coordinantes y las subordinantes (Torrego, 2007).

#### **Conjunciones coordinantes:**

Las conjunciones coordinantes constituyen un puente entre palabras, grupos sintácticos u oraciones sin establecer ninguna relación de dependencia sintáctica. Estas conjunciones generan estructuras compuestas cuando lo que están uniendo son palabras de igual tipo o función. Cuando las conjunciones coordinantes unen proposiciones (oraciones), estas pasan a ser proposiciones coordinadas u oraciones compuestas por coordinación (Torrego, 2007). La tabla 1 muestra las conjunciones coordinantes y sus clasificaciones.

**Tabla 1.** Conjunciones coordinantes y sus clasificaciones. Fuente: (Torrego, 2007)

<b>Conjunciones coordinantes</b>		
<b>Clasificación</b>	<b>Significado</b>	<b>Conjunción</b>
Copulativas	Suma o adición. Se añade un significado.	y, e, ni
Disyuntivas	Alternativa, permiten la elección entre dos o más realidades distintas o entre dos variantes de una misma realidad	o, u
Distributivas	Es una variedad del significado de alternancia.	bien, sea, ya
Adversativas	Contraposición de los dos segmentos unidos.	mas, pero, sin embargo, no obstante, a pesar de, aunque, sino (que)
Explicativas	Unen proposiciones (oraciones) que expresan lo mismo, pero de distinta forma, a fin de que se expliquen mutuamente.	o sea, es decir, esto es

Es necesario aclarar que no todas las conjunciones coordinantes generan ambigüedad sintáctica en un texto. El autor (Zapata, et al, 2008) define como conjunciones coordinantes sintácticamente ambiguas las copulativas (y, e) y las disyuntivas (o, u), pues cuando se emplean de forma excesiva en una oración originan ambigüedad en su comprensión. El siguiente ejemplo muestra como la conjunción [y] genera dificultades en la comprensión de una oración.

*En el proceso jurídico participaron acusado, fiscal y juez y abogado.*

En la oración anterior el uso excesivo de la conjunción coordinante copulativa y, hace que esta sea interpretada de dos formas distintas: en el proceso jurídico intervinieron tres personas, porque fiscal y juez son la misma persona o en el proceso intervinieron cuatro personas (acusado, fiscal, juez y abogado). De igual forma ocurre con el uso excesivo en oraciones del resto de las conjunciones coordinantes que pueden originar ambigüedad sintáctica en una oración (e, o, u).

### 1.1.2 Preposiciones

Las preposiciones son una categoría de palabras invariables con función gramatical dentro del sintagma nominal, que establecen una relación entre dos sintagmas y aportan un matiz significativo. Por sí solas no tienen significado propio, sino que prestan un significado al sintagma que introducen. Se emplean normalmente antepuestas al vocablo que acompañan, y en este sentido se les ha relacionado en ocasiones con los prefijos. Así existen preposiciones como contra que es también un prefijo, contraponer (poner en contra) (Llorach, 1994).

En la tabla 2 se listan las preposiciones definidas en la lengua española y el sentido que estas pueden indicar en una oración.

**Tabla 2.** Preposiciones y sentidos en una oración. Fuente: (Slager, 2004)

Preposición	Sentidos	Preposición	Sentidos
A	Dirección, Tiempo de realización de algo, Complemento directo	Hacia	Dirección
Ante	Situación	Hasta	Límite
Bajo	Lugar, Modo	Incluso	Función de adverbio
Cabe	Considerada ya arcaica	Mediante	Medio
Con	Instrumento, Compañía, Circunstancia, Contenido	Para	Dirección, Objetivo, Tiempo, Finalidad
Contra	Tendencias opuestas entre elementos	Por	Lugar, Causa, Motivo, Tiempo indeterminado o aproximado, Precio, Medio, Modo, Distribución,

## CAPÍTULO 1: Ambigüedad y desambiguación de textos.

			Velocidad, Periodicidad
De	Origen, Pertenencia, Materia, Tema, Contenido	Según	Modo
Desde	Tiempo, Lugar	Salvo	Excepción
En	Lugar, Tiempo, Precio, Medio	Sin	Privación
Entre	Hora, Lugar	Sobre	Lugar, Hora aproximada, Tema
Excepto	Excepción	Tras	Tiempo

Teniendo en cuenta los diferentes sentidos que pueden tener la mayoría de las preposiciones en una oración, existen cuatro de ellas identificadas por el autor Carlos Zapata, como preposiciones sintácticamente ambiguas (a, con, de, en), al generar varias interpretaciones en una oración, a partir de los distintos sentidos que estas pueden representar (Zapata, et al, 2008).

El siguiente ejemplo muestra como la preposición *[en]*, genera más de un significado en una oración.

*El recluso fue capturado en el 1900.*

Al interpretarse la oración no queda claro si el recluso fue hecho prisionero en el año 1900 o en algún lugar nombrado 1900, debido a que entre los sentidos que tiene la preposición *[en]*, se encuentran el de lugar y el de tiempo. De igual forma ocurre con el uso de las otras preposiciones definidas como sintácticamente ambiguas.

### 1.1.3 Estructura gramatical de los textos

La estructura gramatical de los textos, constituye el orden de las palabras, partes de las oraciones y frases que lo componen, expresando así las ideas en forma lógica y adecuada, es decir, sin ambigüedades. La estructura sintáctica de los textos se divide en tres niveles: entre palabras, entre las partes de la oración y en el párrafo (Torrego, 2007).

## CAPÍTULO 1: Ambigüedad y desambiguación de textos.

En el primer nivel las palabras cumplen diversas funciones en las oraciones. Hay dos tipos de palabras que forman los núcleos sintácticos de la oración, el sustantivo, núcleo del sujeto de la oración, hace referencia a la persona, animal, cosa, idea o lugar del que se habla, y el verbo, núcleo del predicado, expresa la acción o estado del sujeto (Torrego, 2007).

Alrededor de estos núcleos se colocan otras palabras que sirven para determinar, precisar, describir, indicar las circunstancias como el tiempo, lugar o estado de las palabras que modifican. Estas palabras se clasifican, según su función en la oración, y tienen un orden lógico en la misma (Torrego, 2007). A continuación, se definen cada una de estas:

**Artículos:** son un tipo de adjetivos, que sirven para señalar o especificar el sustantivo (determinantes). Hay artículos definidos (el, la, los, las) y artículos indefinidos (un, una, unos, unas). Siempre se escriben antes del sustantivo o de la palabra que modifican (Torrego, 2007).

**Adjetivos:** son palabras que describen, limitan o especifican al sujeto u otro adjetivo. Hay varios tipos de adjetivos que se clasifican según su función: posesivos (indican propiedad), demostrativos (indican la distancia entre quien habla y el objeto), calificativos (los que describen alguna característica del objeto), cuantificadores cardinales (los que dicen el número de objetos o sujetos), cuantificadores ordinales (los que marcan el orden en una sucesión), cuantificadores partitivos (los que hablan de una parte o fracción del objeto), cuantificadores indefinidos (los que modifican al sujeto, pero sin ser específicos sobre su cantidad exacta). Los adjetivos siempre se escriben pegados a la palabra que modifican (Torrego, 2007).

Los adjetivos posesivos, demostrativos y cuantificadores, junto con los artículos, son el tipo de palabras llamados determinantes, y generalmente se escriben antes de la palabra que modifican (Torrego, 2007).

Los adjetivos calificativos, generalmente se escriben después del sustantivo que modifican. Cuando se usan varios adjetivos calificativos para un sustantivo, se separan por comas, y se mencionan en forma de lista (Torrego, 2007).



Adverbios: estos modifican a los verbos, a los adjetivos y a otros adverbios, expresando las circunstancias en que suceden. Los principales tipos de adverbio son los de lugar, tiempo, modo y cantidad. Los adverbios siempre se escriben pegados a la palabra que modifican (Torrego, 2007).

A nivel de oración, la estructura gramatical se refiere al orden adecuado entre las partes del sujeto y las partes del predicado. Entre estas dos partes, generalmente se escribe primero el sujeto y luego el predicado.

A nivel de párrafos, la estructura gramatical consiste en ordenar las ideas en forma lógica y jerarquizada. Para hacer una estructura adecuada en un párrafo, primero se debe tener claro el tema y ordenar las ideas principales, de las de mayor importancia a las de menor, y después las ideas complementarias.

### 1.2 Métodos para la desambiguación de textos

En el proceso de resolución de los diferentes tipos de ambigüedades sintáctica presentes en un texto es necesario aplicar y combinar métodos del PLN en función de lograr un proceso de WSD. Estos métodos se clasifican en dos tipos principales: basados en conocimiento (diccionarios y reglas lingüísticas) y los basados en corpus<sup>9</sup> (Leal, 2009).

Los métodos de desambiguación basados en conocimiento utilizan conocimiento lingüístico previamente adquirido. La idea básica consiste en utilizar recursos externos para desambiguar las palabras, tales como reglas heurísticas<sup>10</sup>, diccionarios, ontologías<sup>11</sup>, tesauros<sup>12</sup>, textos sin ningún tipo de etiquetado e incluso recursos de la web (Agirre, 2000).

---

<sup>9</sup> Constituye una colección de textos de un idioma, que se seleccionan de acuerdo a criterios lingüísticos explícitos, con el fin de ser utilizado como muestra representativa de una lengua (Hernández C. P., 2002).

<sup>10</sup> Reglas empíricas definidas a partir del arte y la ciencia del descubrimiento, la invención o la resolución de problemas mediante la creatividad y el pensamiento lateral o pensamiento divergente (Pólya, 1945).

<sup>11</sup> Las ontologías constituyen base de datos que describen los conceptos generales o sobre un dominio en específico, algunas de sus propiedades y cómo los conceptos se relacionan unos con otros (Cruzata, 2011).

<sup>12</sup> Vocabularios controlados que representan las relaciones semánticas con otras palabras y sus significados (Gavilán, 2009).

Los métodos basados en corpus hacen uso de técnicas estadísticas y de aprendizaje automático para inducir modelos del lenguaje a partir de conjuntos de ejemplos textuales. El propósito de un corpus es servir de fuente de datos, proporcionando ejemplos de oraciones y ejemplos de uso de varias palabras, para ser utilizados en algoritmos de aprendizaje automático (Pedersen, 2011).

Los recursos y técnicas utilizadas por ambos métodos han sido aplicados en el desarrollo de soluciones informáticas dirigidas a procesar las ambigüedades en textos. Estas soluciones son descritas en el próximo epígrafe.

### **1.3 Soluciones informáticas para la desambiguación de textos**

El idioma español es complejo de analizar a través de la informática. La mayoría de los trabajos desarrollados hasta el momento sobre el análisis de problemas de ambigüedad presentes en textos, están orientados al idioma inglés. Estos han sido implementados haciendo uso de técnicas del PLN. A continuación, se describen algunos de ellos:

Método supervisado de WSD basado en los modelos de Markov (MM): este método de desambiguación basado en corpus, utiliza modelos de Markov (MM) especializados, haciendo uso de matrices de probabilidades, donde cada estado corresponde a una categoría morfosintáctica<sup>13</sup> y el número de estados corresponde al número de categorías asociadas a una palabra. El método tiene dos pasos fundamentales: la selección de las características relevantes para la tarea de desambiguación, mediante la definición del alfabeto de símbolos utilizado en un MM y la especialización o redefinición de los estados del modelo a partir de la información disponible en los datos de entrenamiento. Este método requiere recursos computacionales escasos para el idioma español, como es el caso de los corpus anotados semánticamente (Marco, 2004).

Método probabilístico basado en una gramática lexicalizada: este método constituye una gramática que proporciona mayor información sintáctica por cada categoría gramatical. El método combina probabilidades sintácticas, las cuales permiten seleccionar una categoría sintáctica de un conjunto de categorías asociadas a una palabra, y probabilidades semánticas, que posibilitan la selección de una regla sintáctica entre un conjunto de reglas asociadas a una oración (Miyao, 2011).

---

<sup>13</sup> Representa los niveles básicos en que se agrupan los fenómenos lingüísticos: fónico, morfosintáctico y semántico (Bosque, et al, 2009).

Método de aprendizaje supervisado a partir de un corpus de textos anotados semánticamente para la resolución de la ambigüedad semántica de las palabras: se necesita una fase previa de aprendizaje antes de poder construir y almacenar un clasificador para cada palabra; en esta fase se recogen los ejemplos del corpus y se incorporan al modelo de probabilidad para hacer la estimación de la función de clasificación. El método utiliza los Modelos de Máxima entropía (MME)<sup>14</sup> para realizar la asignación de sentidos a cada palabra y un algoritmo de aprendizaje que permite estudiar los ejemplos y asignar pesos a las palabras que hacen parte de los mismos (Cueto, 2004).

Método de desambiguación léxica: consiste en asignar automáticamente el sentido de las palabras que aparecen dentro del contexto de una oración, recurriendo a *WordNet Domains*<sup>15</sup>, el cual se usa para recopilar ejemplos de los diferentes dominios asociados con los significados semánticos de las palabras. El valor agregado de esta propuesta es que etiqueta cada palabra, asignándole los dominios a los cuales puede pertenecer la misma. Los dominios se encuentran ordenados de mayor a menor, de acuerdo con la importancia que tenga la palabra en el dominio (Vázquez, et al, 2003).

Método de resolución de ambigüedad léxica basado en el Modelo de Espacio Vectorial (MEV): cada sentido de una palabra es representado con un vector, así como el contexto de la palabra a desambiguar. Las entradas del algoritmo están representadas por los vectores, que son procesados mediante el algoritmo LVQ (Learning Vector Quantization). Mediante una función de similitud se comparan los vectores que representan el contexto de cada palabra a desambiguar con cada uno de los vectores de sus sentidos. El sentido representado por el vector de mayor similitud será el designado como sentido desambiguado (Valdivia, et al, 2011).

Método para la desambiguación sintáctica de tipo coordinativo y preposicional: este método está basado en la definición de un conjunto de reglas heurísticas que permiten desambiguar textos que presentan ambigüedad sintáctica causada por conjunciones o preposiciones (Zapata, et al, 2008).

---

<sup>14</sup> Constituyen modelos matemáticos que pertenecen a la familia de los clasificadores exponenciales o logaritmos lineales, utilizados como marcos probabilísticos para el aprendizaje automático (Phillips, et al, 2006).

<sup>15</sup> Base de datos léxica para el idioma inglés.

Natural Language Toolkit (NLTK): es un paquete de herramientas y recursos libres para un espectro amplio de tareas dentro de PLN. NLTK está destinada a apoyar la investigación y la enseñanza en PLN o áreas relacionadas, que incluyen la lingüística empírica, las ciencias cognitivas, la inteligencia artificial, la recuperación de información y el aprendizaje de la máquina. El paquete integra diferentes tipos de corpus: texto simple, texto etiquetado con su categoría, etiquetado con sintaxis superficial, etiquetado con sintaxis profunda e incluso simples listas de palabras o léxicos. Cada corpus ofrece una serie de métodos para leer sus datos palabra por palabra, oración por oración, párrafo por párrafo o por unidades de etiquetado (Manterola, et al, 2010).

3LB-SAT (3LB-Herramienta de Anotación Semántica): es una herramienta para el etiquetado semántico de corpus multilingüe. Tiene como principales características: que está orientado a la palabra, permite introducir el corpus en diferentes formatos y usa EuroWordnet<sup>16</sup> para consultar el sentido de las palabras en cuatro idiomas (español, catalán, euskara e inglés). Anota las palabras monosémicas<sup>17</sup> automáticamente; y muestra todas las apariciones de un lema en el texto, siendo posible asociar más de un synset<sup>18</sup> con la aparición de un lema (Bisbal, et al, 2003).

La mayoría de las soluciones antes descritas están basadas en métodos estadísticos y técnicas de inteligencia artificial. Las basadas en métodos estadísticos, a pesar de resolver problemas de ambigüedad, consumen muchos recursos léxicos y computacionales, esto las hace poco convenientes para su uso en el análisis del idioma español (Coutiño, 2006). Además, estos trabajos son muy especializados, pues recurren a fórmulas y estudios complejos que hacen necesaria la presencia de un experto, si se desea mejorar el resultado (Zapata, et al, 2008).

Las soluciones basadas en técnicas de inteligencia artificial requieren de repositorios de información muy extensos (redes semánticas, ontologías específicas de un dominio particular, lexicones<sup>19</sup> o corpus) para llegar a una solución coherente. Estos repositorios

---

<sup>16</sup> Diccionario electrónico semántico que tiene como fin la construcción de base de datos léxico-semánticas y contiene relaciones semánticas entre palabras de diez idiomas de la Comunidad Europea (Verdejo, et al, 2012).

<sup>17</sup> Palabras que cuentan con un significado único (RAE, 2016).

<sup>18</sup> Conjunto de sinónimos estrictos para una palabra (López, 2002).

<sup>19</sup> Plural de lexicón, que significa tipo de diccionario que contiene una amplia gama de información sobre palabras específicas de un lenguaje (López, 2002).

son escasos para el idioma español o pueden pertenecer a dominios muy restringidos, lo cual limita su aplicación en dominios específicos (Coutiño, 2006).

Teniendo en cuenta que en Cuba existen pocas experiencias en el desarrollo de soluciones analíticas o algorítmicas para definir un corpus, las características de las soluciones antes descritas y la necesidad de utilizar conocimientos previos al dominio de la investigación, aplicando reglas heurísticas y diccionarios de términos, se considera oportuno utilizar recursos de los métodos basados en conocimiento del PLN para llegar a una solución en la desambiguación de textos jurídicos.

Para el desarrollo de la solución se toma como referencia el “Método para la desambiguación sintáctica de tipo coordinativo y preposicional”, del autor Carlos Zapata (Zapata, et al, 2008). La decisión es tomada considerando que, de las soluciones estudiadas, esta es la única desarrollada para el idioma español que utiliza recursos de los métodos basados en conocimiento. Los cuales pueden ser adaptadas con facilidad al dominio de la presente investigación.

El método propuesto constituye un acercamiento al trabajo de Zapata, al basar su estructura en la definición de reglas heurísticas. Pero el alcance es mayor, teniendo en cuenta que se analizan otras causas que originan ambigüedad sintáctica en los textos analizados. Además, se incluyen características de otras de las soluciones estudiadas como la definición de gramáticas y un lexicón, con el propósito de brindar una información detallada al usuario sobre las causas que originan ambigüedad sintáctica en un texto y cómo proceder para reducirla, elementos no considerados en el método referenciado.

### **1.4 Conclusiones parciales**

El análisis de las conjunciones y preposiciones sintácticamente ambiguas, así como los elementos gramaticales a tener en cuenta para la correcta estructuración sintáctica de textos en el idioma español, permitieron identificar las causas que pueden originar problemas de ambigüedad sintáctica en un texto.

El análisis de los métodos del PLN y su aplicación en soluciones informáticas dirigidas a la desambiguación de textos, asistieron la decisión de utilizar recursos de los métodos basados en conocimiento para la definición de la solución propuesta.

## **CAPÍTULO 1: Ambigüedad y desambiguación de textos.**

La definición y aplicación de reglas heurísticas, gramáticas y un lexicón, como aplicación de técnicas del PLN, constituyen la base para el diseño e implementación del método de desambiguación sintáctica propuesto.

## CAPÍTULO 2: DIAGNÓSTICO Y PROPUESTA DE SOLUCIÓN

En el capítulo se realiza un diagnóstico sobre el estado actual del objeto de estudio de la presente investigación en el sistema judicial cubano. En el mismo se describe el nivel de informatización de la esfera jurídica cubana en relación al tema de la desambiguación de textos. En el capítulo se describe también el método para la desambiguación sintáctica de textos de la legislación cubana propuesto en el presente trabajo. La descripción se realiza paso a paso, a partir del uso de técnicas del PLN, con el propósito de destacar el nexo que existe entre sus partes y la flexibilidad que tiene para ser aplicado a otros campos de la ciencia.

### **2.1 Diagnóstico sobre soluciones informáticas para reducir la ambigüedad sintáctica en textos de la legislación cubana**

Un diagnóstico es el proceso mediante el cual se lleva a cabo un análisis para recopilar información que ayude a determinar la situación actual de la organización y detectar sus áreas de mejoramiento. Mediante un diagnóstico se trata de focalizar y evaluar un conjunto de variables que juegan un importante papel en la comprensión, predicción y control del comportamiento de un fenómeno determinado (Ortega, 2016).

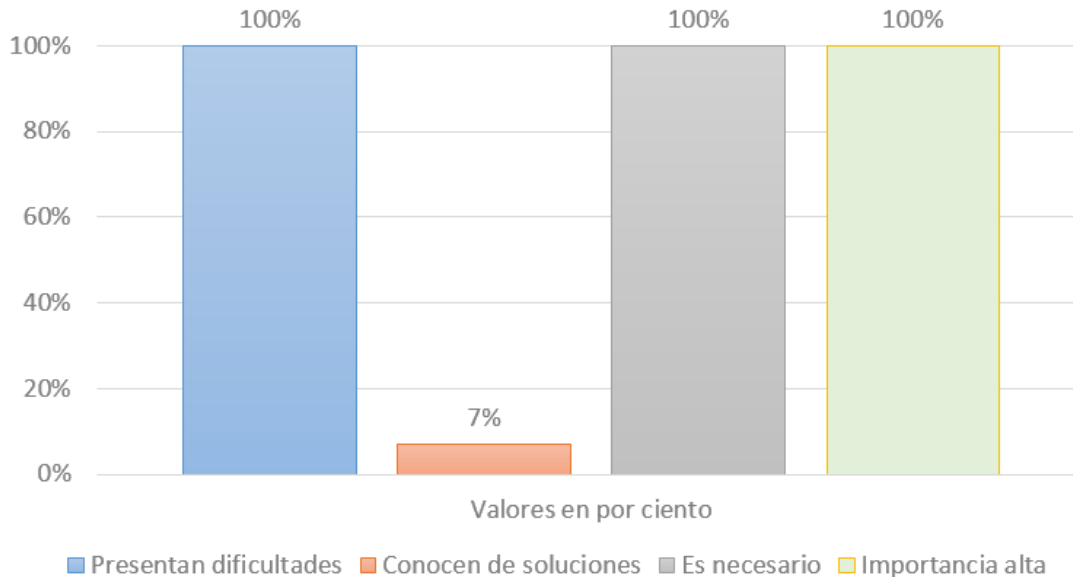
El diagnóstico de la presente investigación se desarrolló a través de una encuesta aplicada a especialistas en temas jurídicos, ver Anexo 1. Los encuestados fueron seleccionados teniendo en cuenta su vínculo con la Sociedad Cubana de Derecho e Informática, con el propósito de obtener información de personas relacionadas con el área de la ciencia a la que responde el método propuesto, la informática jurídica. Las personas seleccionadas fueron informadas sobre el objetivo de la investigación. También se les dio una breve explicación sobre el proceso de desambiguación de textos y su importancia para elevar la comprensión de los escritos. La encuesta se aplicó con el propósito de comprobar el comportamiento de los siguientes elementos tenidos en cuenta en la definición de la propuesta de solución:

- Dificultades presentadas por los encuestados en la comprensión de un determinado texto jurídico a causa de posibles ambigüedades existentes en este.
- Conocimientos sobre el desarrollo en Cuba de trabajos similares al tema de investigación.

## CAPÍTULO 2: DIAGNÓSTICO Y PROPUESTA DE SOLUCIÓN

- Necesidad de una solución informática que auxilie a los usuarios en la detección y reducción de ambigüedades en textos jurídicos.
- La importancia que atribuyen los encuestados al método propuesto.

La encuesta fue aplicada a una muestra de 29 personas. A continuación, se muestran los principales resultados obtenidos con la aplicación de este instrumento, ver figura 1.



**Figura 1.** Resultados del diagnóstico. Fuente: elaboración propia

La gráfica anterior muestra los siguientes datos:

- El 100 % de las personas encuestadas ha presentado dificultades en la comprensión de un escrito jurídico, como resultado de posibles ambigüedades presentes en este.
- Solo el 7 % de los encuestados conocen de la existencia de trabajos dirigidos a la desambiguación de textos jurídicos en Cuba, pero en todos los casos solo se ha llegado a nivel teórico en el desarrollo de estos.
- El 100 % de los profesionales encuestados, consideran viable y necesaria la existencia de una solución informática que los auxilie en la comprensión de textos jurídicos con problemas de ambigüedad.

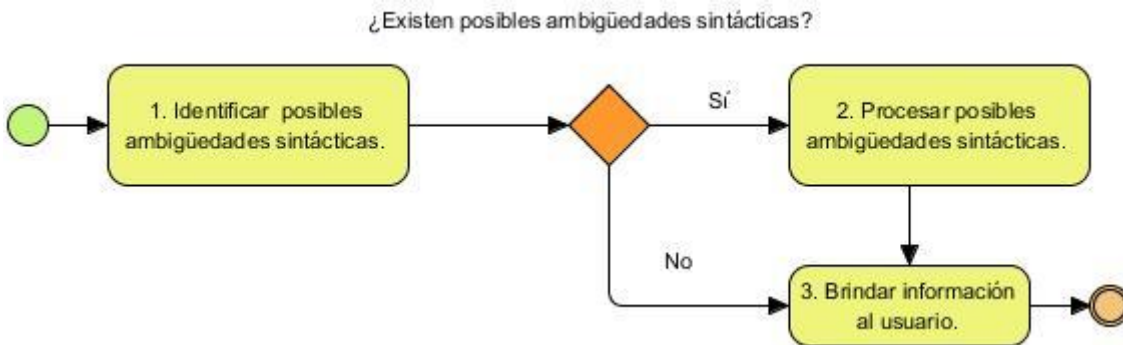


- El 100 % de la muestra de personas encuestadas le atribuye una importancia alta al desarrollo del presente trabajo.

Los resultados obtenidos con el desarrollo del diagnóstico de la presente investigación, confirman los elementos analizados en el capítulo 1, sobre el reducido número de soluciones de este tipo existentes para el idioma español. De esta forma se evidencia la importancia del desarrollo del método propuesto, como herramienta informática para el auxilio de los usuarios en la comprensión de textos jurídicos con problemas de ambigüedad sintáctica. Además, se demuestra la novedad e influencia del trabajo en la evolución de la informática jurídica en Cuba, a partir del nivel de importancia que los usuarios encuestados atribuyen a su desarrollo.

## 2.2 Descripción del método propuesto

El desarrollo de este trabajo constituye una continuidad del método para la desambiguación sintáctica de tipo coordinativo y preposicional, propuesto en el año 2008 por el autor Carlos Zapata (Zapata, et al, 2008). Sin embargo, como se explica en el capítulo 1 de la presente investigación, el método que se propone en esta tesis de maestría, tiene mayor alcance y a pesar de estar dirigido al análisis de textos jurídicos, puede ser fácilmente adaptado a otros dominios. La solución propuesta hace uso de los métodos basados en conocimiento (reglas heurísticas y diccionarios), definidos por el PLN y está organizada en tres partes. En la figura 2 se muestra el flujo a seguir para el desarrollo del método propuesto. Para la representación de cada uno de los flujos que describen las partes del método propuesto, se emplearon diagramas de procesos, con el propósito de realizar una mejor descripción de la solución.



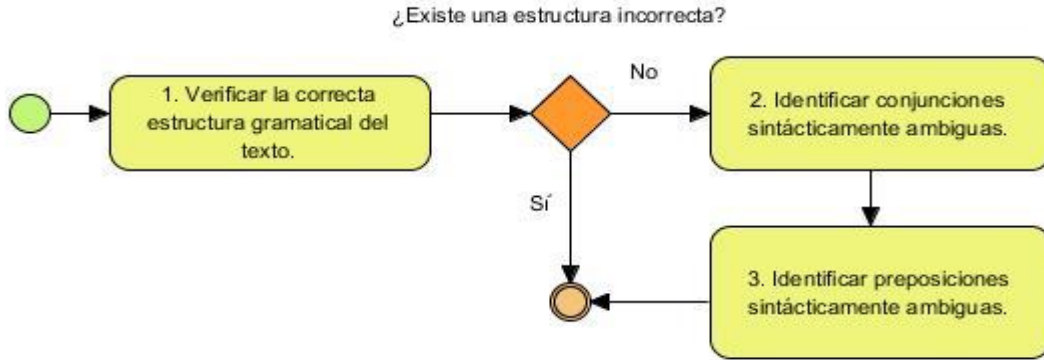
**Figura 2.** Flujo para el desarrollo del método de desambiguación sintáctica propuesto. Fuente: elaboración propia

A continuación, se describen brevemente las partes del método propuesto, definidas en el flujo anterior, teniendo en cuenta que en próximos epígrafes se realiza una descripción más ampliada de las mismas:

1. Identificar posibles ambigüedades sintácticas: se identifican los posibles tipos de ambigüedades sintácticas que pueden estar presentes en los textos analizados.
2. Procesar posibles ambigüedades sintácticas: se procesan las posibles ambigüedades sintácticas detectadas, con el fin de determinar su origen y en determinados casos ofrecer una propuesta de desambiguación al usuario.
3. Brindar información al usuario: se muestran los resultados al usuario, brindando una información detallada sobre las causas que originan los problemas de ambigüedad sintáctica identificados en los textos analizados, con el propósito que estos sepan cómo proceder a su reducción.

### **2.2.1 Identificación de posibles ambigüedades sintácticas**

En la primera parte del método de desambiguación sintáctica que responde al objetivo general de la presente investigación, se identifican las posibles ambigüedades sintácticas existentes en un texto. Es importante aclarar que en toda la descripción del método propuesto se utiliza el término posible ambigüedad, con el propósito de especificar que no en todos los casos el uso de conjunciones y preposiciones clasificadas como sintácticamente ambiguas, originan ambigüedad sintáctica al ser identificadas en un texto. El método propuesto constituye una herramienta de auxilio para los usuarios que lo aplican teniendo en cuenta que los resultados proporcionados por este, una vez ejecutadas sus tres partes, brindan la posibilidad a los usuarios de decidir si apoyarse o no en el método para lograr una mejor comprensión del texto en análisis. El flujo de la figura 3 muestra los pasos a seguir para la ejecución de la primera parte del método propuesto.



**Figura 3.** Actividades de la primera parte del método de desambiguación sintáctica propuesto.

Fuente: elaboración propia

A continuación, se describen de forma detallada los pasos definidos en el flujo de la figura 3:

**Primer paso:** Verificar la correcta estructura gramatical del texto.

En el desarrollo de este paso se utiliza la técnica de diccionario, como aplicación de los métodos basados en conocimiento, del PLN. La verificación de la correcta estructura gramatical del texto se realiza por medio de la definición de una gramática libre de contexto y un lexicón que actúa como diccionario.

Las gramáticas libres de contexto (GLC) constituyen una gramática formal<sup>20</sup> en la que cada regla de producción es de la forma  $A \rightarrow B$  (Alvarez, 2013). Estas son gramáticas de tipo 2 o gramáticas independientes del contexto que generan lenguajes libres o independientes del contexto. Estos lenguajes pueden ser reconocidos por autómatas de pila<sup>21</sup> determinísticos o no determinísticos (Martínez, 2014). En el caso del método propuesto, se debe diseñar una GLC donde cada regla de producción represente la forma correcta en que debe estar estructurada gramaticalmente una oración. En la figura 4 se muestra un ejemplo de esta.

<sup>20</sup> Una gramática formal es una estructura matemática con un conjunto de reglas de formación que definen las cadenas de caracteres admisibles en un determinado lenguaje formal o lengua natural (Marin, et al, 2013).

<sup>21</sup> Un autómata de pila es un modelo matemático de un sistema que recibe una cadena constituida por símbolos de un alfabeto y determina si esa cadena pertenece al lenguaje que el autómata reconoce (Hopcroft, et al, 2001).

## CAPÍTULO 2: DIAGNÓSTICO Y PROPUESTA DE SOLUCIÓN

O->Asm Ssm V Asm Ssm  
O->Asm Ssm V Asf Ssf  
O->Apm Spm Vsp Apm Spm  
O->Asm Ssm V Asf Ssf  
O->Asf Ssf V Asm Ssm  
O->Asf Ssf V Asf Ssf  
O->Apm Spm V Asm Ssm  
O->Apm Spm V Asf Ssf  
O->Apf Spf V Asm Ssm  
O->Apf Spf V Asf Ssf  
O->Apm Spm V Apm Spm  
O->Apf Spf V Apf Spf

**Figura 4.** GLC definida para el método propuesto. Fuente: elaboración propia

La GLC representada en la figura 4 define algunas reglas de producción que validan la correcta estructuración de una oración. La primera de estas reglas permite evaluar una oración con la siguiente estructura: (Asm: Artículo en número singular género masculino) + (Ssm: Sustantivo en número singular género masculino) + (V: Forma verbal en singular) + (Asm: Artículo en número singular género masculino) + (Ssm: Sustantivo en número singular género masculino), por tanto, la oración: *El juez terminó el juicio*, sería válida al ser evaluada en esta primera regla de producción de la GLC antes definida. De igual forma se deben continuar definiendo reglas de producción hasta completar la gramática y cubrir todas las posibles estructuras de las oraciones que pueden formar parte de un texto jurídico.

El lexicón que se propone en esta primera parte del método, constituye una lista de palabras definidas en un fichero en forma de diccionario, etiquetadas en función de la categoría gramatical a la cual pertenecen. Este diseño se realiza a partir los elementos definidos por la profesora Manuela Sassi en (Sassi, et al, Saba, 1981), donde escribe sobre el desarrollo de analizadores gramaticales para el idioma español, uno de los pocos referentes que existen sobre la aplicación de la lingüística computacional a la lengua española.

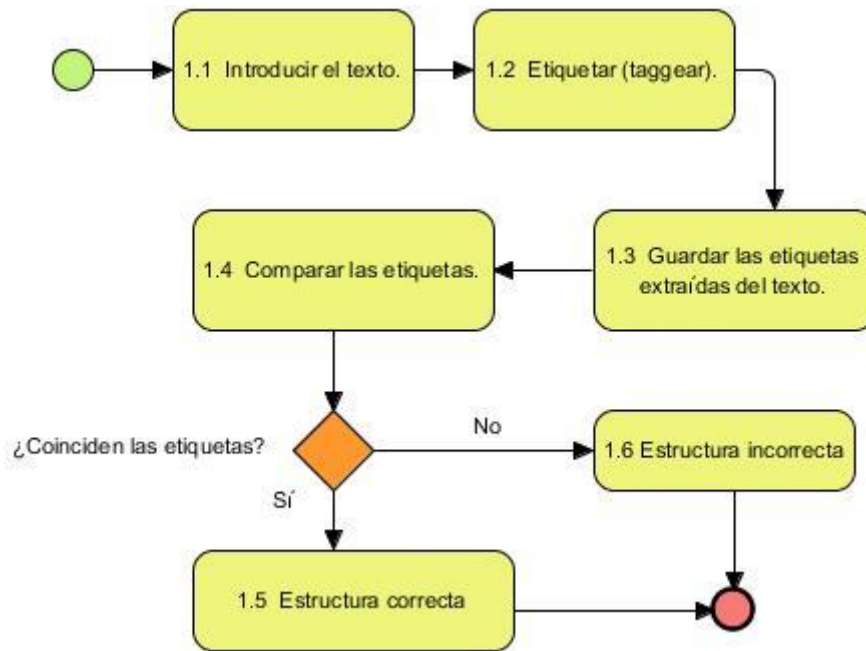
El lexicón es diseñado con el fin de señalar en los textos jurídicos, la categoría gramatical a la cual pertenece cada palabra y luego controlar la correcta estructura del texto por medio de las reglas de producción definidas en la GLC. En la figura 5 se muestra una parte del lexicón a definir para la aplicación del método propuesto.

```
Asm->el  
Asm->al  
Asf->la  
Apm->los  
Apf->las  
Ssm->juez  
Spm->jueces  
Ssf->jueza  
Spf->juezas  
Ssm->tribunal  
Spm->tribunales
```

**Figura 5.** Lexicón definido para el método propuesto. Fuente: elaboración propia

El lexicón representado en la figura 5 muestra las etiquetas definidas para algunos artículos y sustantivos del idioma español, teniendo en cuenta sus categorías sintácticas en cuanto a género y número. Ejemplo el artículo *el*, se etiqueta como *Asm* (Artículo en número singular género masculino) y el sustantivo *jueza* es etiquetado como *Ssf* (Sustantivo en número singular género femenino). De igual forma se deben etiquetar el resto de los términos del idioma español utilizados en textos jurídicos, para poder ser identificados en los textos analizados y evaluados en las reglas de producción de la GLC.

A continuación, en la figura 6 se desglosa, a través de un diagrama de procesos, el flujo a seguir para validar la correcta estructuración gramatical de los textos, con la ejecución del paso número uno de la primera parte del método propuesto, utilizando el lexicón y la GLC.



**Figura 6.** Flujo a seguir para validar la correcta estructuración gramatical de los textos, utilizando el lexicon y la GLC. Fuente: elaboración propia

A continuación, se describen las actividades definidas en el flujo anterior:

- 1.1 Introducir el texto: se introduce el texto a analizar.
- 1.2 Etiquetar (*taggear*): se etiquetan cada uno de los términos que forman las oraciones del texto en análisis, a partir de las etiquetas definidas para cada palabra en el lexicon. Proceso definido en programación como *taggear*.
- 1.3 Guardar las etiquetas extraídas del texto: se guardan en una lista las etiquetas extraídas del texto, organizadas por oraciones, en el mismo orden en que fueron apareciendo.
- 1.4 Comparar las etiquetas: se comparan las listas de etiquetas extraídas de las oraciones del texto, con las reglas de producción definidas en la GLC, con el propósito de validar que la estructura de cada oración corresponde a una de estas reglas.
- 1.5 Estructura correcta: si al comparar la lista de etiquetas generada de cada oración del texto con las reglas de producción de la GLC, se identifica que alguna regla coincide con la lista, entonces la estructura del texto es correcta.

1.6 Estructura incorrecta: si al comparar la lista de etiquetas generada de cada oración del texto con las reglas de producción de la GLC, se identifica que ninguna regla coincide con la lista, entonces la estructura del texto es incorrecta.

El siguiente ejemplo explica de forma práctica cómo un texto puede ser validado aplicando los pasos descritos en el flujo anterior.

1.1 Se introduce el texto: *El juez terminó el juicio.*

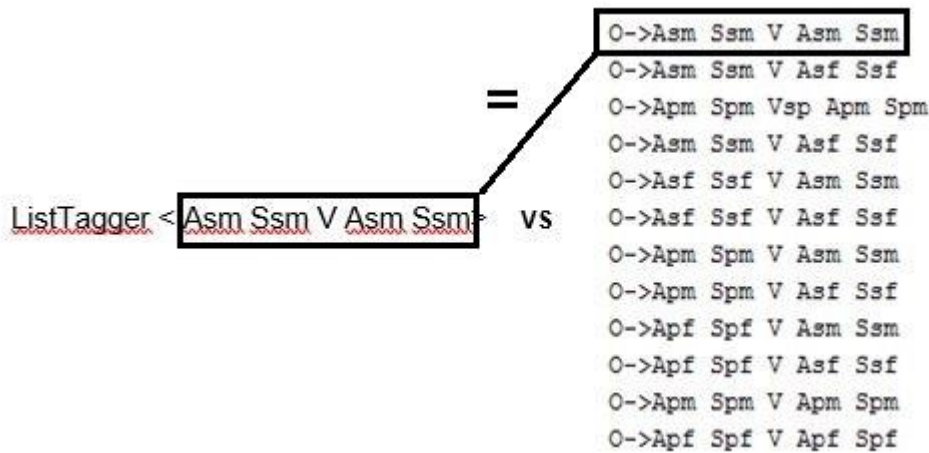
1.2 El texto queda etiquetado, luego de cada término ser evaluado en las etiquetas definidas en el lexicón:

El	juez	terminó	el	juicio
↓	↓	↓	↓	↓
Asm	Ssm	V	Asm	Ssm

1.3 Las etiquetas definidas para cada término se guardan en una lista:

ListTagger < Asm Ssm V Asm Ssm >

1.4 Se comparan las listas de etiquetas definidas por cada oración del texto, con cada regla de producción definida en la GLC. La figura 7 muestra cómo la lista de etiquetas coincide con una de las reglas de producción. Por tanto, el texto tiene una correcta estructura gramatical.



**Figura 7.** Comparación entre la lista de etiquetas extraídas de una oración y la GLC. Fuente: elaboración propia

1.5 Se obtiene un resultado positivo, indicando que la estructura del texto en análisis es correcta.

**Segundo paso:** Identificar conjunciones sintácticamente ambiguas.

Este paso se ejecuta si en el paso anterior no se identifican estructuras incorrectas en el texto en análisis, de existir incoherencias en alguna parte del texto, el usuario luego de ser informado sobre estos problemas, debe proceder a resolverlos para poder continuar con el análisis del texto, aplicando el resto de los pasos del método, como se describe en próximos epígrafes.

En este paso se identifica la existencia de conjunciones sintácticamente ambiguas (e, y, o, u) en el texto. Esta identificación se realiza a través de la implementación de un par de reglas heurísticas, definidas por el autor del presente trabajo, teniendo en cuenta los siguientes elementos:

- El resultado de investigaciones previas vinculadas al tema.
- El análisis de textos donde se demuestra la ocurrencia de ambigüedades sintácticas por el uso excesivo de estas categorías gramaticales.
- Los elementos teóricos analizados en el epígrafe 1.2.1.
- La aplicación de las técnicas del PLN.

A continuación, se listan las reglas heurísticas formuladas:

**Regla heurística 1:** si una oración contiene más de una conjunción sintácticamente ambigua, alternada de forma consecutiva con otros elementos gramaticales y estas pertenecen al grupo de conjunciones coordinantes copulativas (y, e), entonces la oración presenta posible ambigüedad coordinante copulativa.

**Regla heurística 2:** si una oración contiene más de una conjunción sintácticamente ambigua, alternada de forma consecutiva con otros elementos gramaticales y estas pertenecen al grupo de conjunciones coordinantes disyuntivas (o, u), entonces la oración presenta posible ambigüedad coordinante disyuntiva.

**Tercer paso:** Identificar preposiciones sintácticamente ambiguas.

Una vez identificadas las posibles ambigüedades presentes en un texto originadas por el uso excesivo de conjunciones sintácticamente ambiguas, se procede a la ejecución del

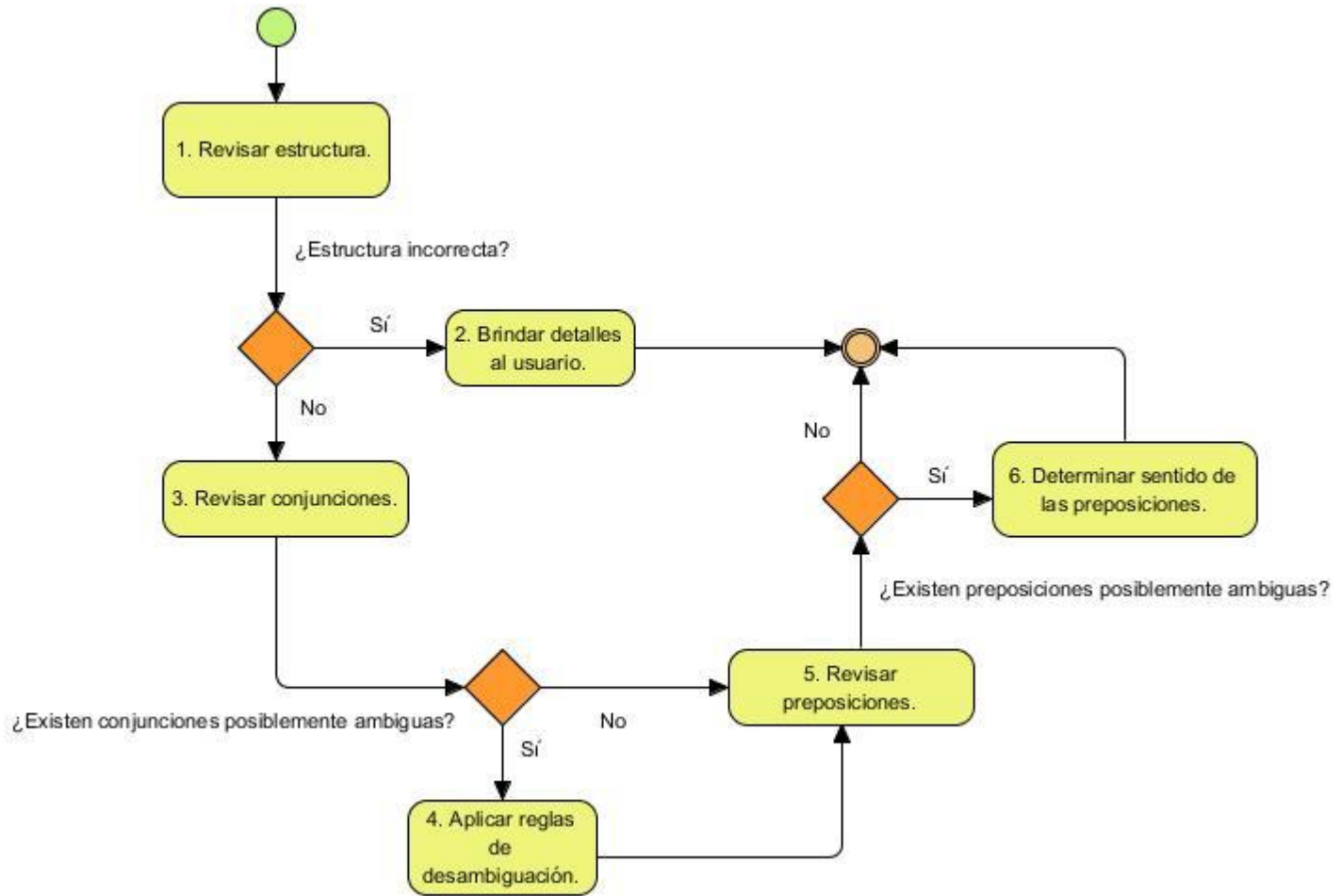


tercer y último paso de la primera parte del método propuesto, con el fin de ubicar en un texto la existencia de preposiciones sintácticamente ambiguas. Para el desarrollo de este paso se debe implementar una nueva regla heurística, formulada a partir de los mismos elementos tenidos en cuenta por el autor de la presente investigación en la definición de las reglas 1, 2 y 3. Esta regla permite identificar si en los textos a analizar existen posibles ambigüedades originadas por preposiciones sintácticamente ambiguas (a, con, de, en). A continuación, se lista la formulación de esta regla:

**Regla heurística 3:** si una oración contiene al menos una preposición sintácticamente ambigua (a, con, de, en) entonces la oración presenta posible ambigüedad preposicional.

### 2.2.2 Procesamiento de posibles ambigüedades sintácticas

Una vez descrita la primera parte del método propuesto, se procede a describir la segunda parte de este. El propósito de este segundo segmento del método es procesar las ambigüedades sintácticas identificadas a través de los pasos descritos en el epígrafe anterior, con el fin de brindar información detallada al usuario sobre las causas que originan problemas de ambigüedad en un texto en análisis. Esta información permite que el usuario conozca cómo proceder a la reducción de las ambigüedades identificadas en el texto y hacer este más legible. En la figura 8 se muestra a través de un diagrama de procesos, los pasos a seguir para la ejecución de esta segunda parte del método propuesto. Es importante aclarar que cada paso de esta segunda parte, se ejecuta en dependencia del resultado obtenido con la implementación de los pasos descritos en la figura 3 del epígrafe anterior.



**Figura 8.** Pasos y condiciones para la ejecución de la segunda parte del método propuesto.  
Fuente: elaboración propia

A continuación, se describen las actividades definidas en el flujo anterior:

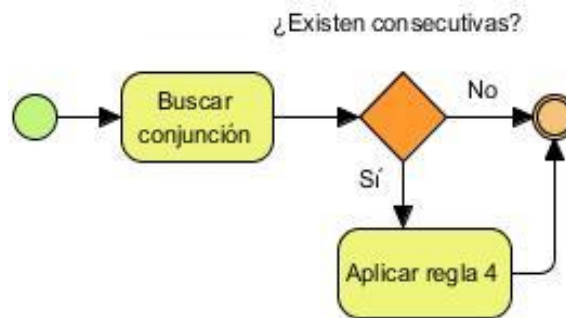
1. Revisar estructura: se analiza el resultado del paso número uno de la primera parte del método, de existir estructuras gramaticales incorrectas en el texto analizado, se brindan detalles al usuario sobre estas incongruencias. Si el texto está correctamente estructurado, se continúa procesando.
2. Brindar detalles al usuario: este paso se ejecuta si se identifican estructuras gramaticales incorrectas en el texto y tiene como propósito mostrar información detallada sobre la ubicación del problema estructural. Una vez representada la parte del texto con problemas, el usuario debe volver a iniciar el proceso. En el próximo epígrafe se explica de forma detallada cómo mostrar la información al usuario.

3. Revisar conjunciones: se analiza el resultado del paso dos de la primera parte del método, de existir conjunciones sintácticamente ambiguas en el texto en análisis, se procede a reducir las ambigüedades que estas provocan.
4. Aplicar reglas de desambiguación: este paso se ejecuta, si se identifican conjunciones sintácticamente ambiguas en el texto. A continuación, se procede a la descripción detallada del mismo.

El desarrollo de este paso parte de la formulación de una nueva regla heurística, a partir del cumplimiento de las reglas 1 y 2 definidas en el epígrafe anterior. El propósito de esta cuarta regla es modificar las oraciones que presentan ambigüedad coordinante copulativa o disyuntiva, con el fin de hacerlas más comprensibles para el usuario.

**Regla 4:** se modifica la oración posiblemente ambigua, sustituyendo por comas todas las conjunciones consecutivas excepto la última.

A continuación, en la figura 9 se desglosa, a través de un diagrama de procesos, el flujo a seguir al aplicar la regla heurística 4:



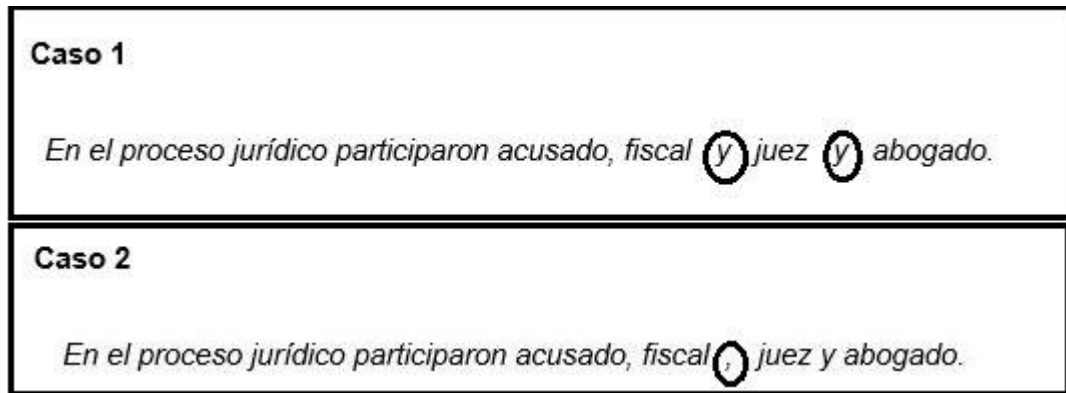
**Figura 9.** Flujo a seguir a partir de la implementación de la regla heurística 4. Fuente: elaboración propia

En el siguiente ejemplo se expone de forma práctica, cómo a través de la implementación de la regla heurística 4, se aplican los pasos del flujo descrito en la figura 9, con el propósito de brindar una propuesta de desambiguación al usuario. El ejemplo parte del análisis de la siguiente oración:

*En el proceso jurídico participaron acusado, fiscal y juez y abogado.*

## CAPÍTULO 2: DIAGNÓSTICO Y PROPUESTA DE SOLUCIÓN

Una vez identificada la existencia de conjunciones sintácticamente ambiguas en el texto en análisis, por medio de la implementación del segundo paso del flujo expuesto en la figura 3, se procede a ubicar las mismas. Para el caso del ejemplo en curso, se tiene la existencia en dos ocasiones de la conjunción [y], alternada de forma consecutiva con dos sustantivos, ver caso 1 de la figura 10. Por tanto, al aplicar la regla heurística 4, la oración se transforma, eliminando la primera [y], por una coma, ver caso 2 de la figura 10, de esta forma se obtiene una propuesta de desambiguación de la oración en análisis.



**Figura 10.** Aplicación de la regla heurística 4 en un texto. Fuente: elaboración propia

5. Revisar preposiciones: se analiza el resultado del paso tres de la primera parte del método, de existir preposiciones sintácticamente ambiguas en el texto en análisis, se procede a reducir las ambigüedades causadas por las mismas a través de la determinación de su sentido.
6. Determinar sentido de las preposiciones: este paso se ejecuta si se identifican preposiciones sintácticamente ambiguas en el texto. A continuación, se procede a la descripción detallada del mismo.

Para la realización el paso se formula un nuevo grupo de reglas heurísticas, a partir de los criterios tenidos en cuenta por el autor en la definición de las reglas 1, 2, 3 y 4. Estas reglas permiten extraer información semántica de las oraciones de los textos analizados. A continuación, se listan las reglas definidas:

**Regla 5:** se identifica la preposición que ha generado la ambigüedad sintáctica.

**Regla 5.1:** si la preposición es “a”:

**Regla 5.1.1:** si la preposición se encuentra sucedida por un sustantivo que representa dirección, entonces el sentido de la preposición es de dirección.

**Regla 5.1.2:** si la preposición se encuentra sucedida por un sustantivo que actúe o represente un punto en el tiempo, entonces el sentido de la preposición es de tiempo.

**Regla 5.1.3:** si la preposición se encuentra sucedida por un sustantivo que actúe como complemento directo, entonces el sentido de la preposición es de complemento directo.

**Regla 5.2:** si la preposición es “con”:

**Regla 5.2.1:** si la preposición se encuentra sucedida por un sustantivo que represente una persona, animal u objeto, entonces el sentido de la preposición es de compañía.

**Regla 5.2.2:** si la preposición se encuentra sucedida por un sustantivo que actúe o represente una circunstancia, entonces el sentido de la preposición es de circunstancia.

**Regla 5.2.3:** si la preposición se encuentra sucedida por un sustantivo que actúe o represente un instrumento, entonces el sentido de la preposición es de instrumento.

**Regla 5.2.4:** si la preposición se encuentra sucedida por un sustantivo relacionado con sustancia o material, entonces el sentido de la preposición es de contenido.

**Regla 5.3:** si la preposición es “de”:

**Regla 5.3.1:** si la preposición se encuentra sucedida por un sustantivo relacionado con un tipo de material, entonces el sentido de la preposición es de materia.

## CAPÍTULO 2: DIAGNÓSTICO Y PROPUESTA DE SOLUCIÓN

**Regla 5.3.2:** si la preposición se encuentra sucedida por un sustantivo relacionado con una persona, animal u objeto, entonces el sentido de la preposición es de pertenencia.

**Regla 5.3.3:** si la preposición se encuentra sucedida por un sustantivo relacionado con un lugar, entonces el sentido de la preposición es de origen.

**Regla 5.3.4:** si la preposición se encuentra sucedida por un sustantivo relacionado con una sustancia o material entonces el sentido de la preposición es de contenido.

**Regla 5.3.5:** si la preposición se encuentra sucedida por un sustantivo relacionado con un tema, entonces el sentido de la preposición es de tema.

**Regla 5.4:** si la preposición es “en”:

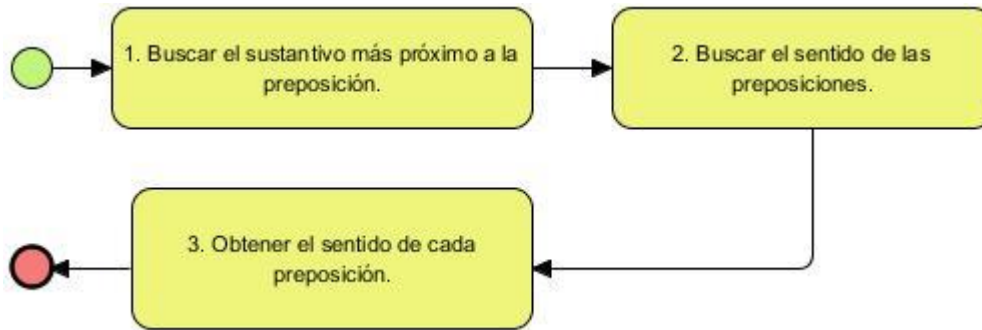
**Regla 5.4.1:** si la preposición se encuentra sucedida por un sustantivo que represente un lugar, entonces el sentido de la preposición es de lugar.

**Regla 5.4.2:** si la preposición se encuentra sucedida por un sustantivo que represente un punto en el tiempo, entonces el sentido de la preposición es de tiempo.

**Regla 5.4.3:** si la preposición se encuentra sucedida por un sustantivo que represente un precio, entonces el sentido de la preposición es de precio.

**Regla 5.4.4:** si la preposición se encuentra sucedida por un sustantivo que represente un medio, entonces el sentido de la preposición es de medio.

En el diagrama de procesos de la figura 11 se describe el flujo a seguir para señalar en un texto el sentido de las preposiciones sintácticamente ambiguas (a, con, de, en), a partir de la implementación de las reglas antes definidas.



**Figura 11.** Flujo a seguir para señalar en un texto, el sentido de las preposiciones sintácticamente ambiguas. Fuente: elaboración propia

A continuación, se describen las actividades del flujo anterior:

1. Buscar el sustantivo más próximo a la preposición: a partir de las preposiciones sintéticamente ambiguas identificadas con la aplicación del tercer paso del flujo descrito en la figura 3, se busca el sustantivo que precede a cada una de estas, haciendo uso de una segunda gramática definida para la segunda parte del método, que recoge una lista de sustantivos y los sentidos que estos brindan a las preposiciones sintéticamente ambiguas (a, con, de, en).
2. Buscar el sentido de las preposiciones: una vez identificado el sustantivo que precede a cada preposición sintéticamente ambigua, se busca el sentido que este brinda a esa preposición. Esta búsqueda se realiza haciendo uso de la gramática definida en la segunda parte del método propuesto.
3. Obtener el sentido de cada preposición sintéticamente ambigua: se obtiene el sentido de cada preposición sintéticamente ambigua, brindado por el sustantivo más próximo escrito después de cada una de estas.

En la figura 12 se muestra un ejemplo de la gramática citada en la descripción de los pasos 1 y 2 del flujo de la figura 11. En esta se asocian los sustantivos, con el sentido que proporcionan a las preposiciones sintácticamente ambigua (a, con, de, en).

```
S->parque#dirección#compañía#dirección#lugar  
S->jueza#complemento#compañía#pertenencia#lugar  
S->juezas#complemento#compañía#pertenencia#lugar  
S->juez#complemento#compañía#pertenencia#lugar  
S->jueces#complemento#compañía#pertenencia#lugar  
S->tribunal#dirección#compañía#pertenencia#lugar  
S->tribunales#dirección#compañía#pertenencia#lugar  
S->ley#complemento#instrumento#pertenencia#lugar  
S->leyes#complemento#instrumento#pertenencia#lugar  
S->sala#dirección#circunstancia#origen#lugar
```

**Figura 12.** Representación a través de una gramática del sentido de las preposiciones sintácticamente ambiguas. Fuente: elaboración propia

A continuación, se explica a través de un ejemplo práctico, cómo obtener desde un texto, el sentido de las preposiciones sintácticamente ambiguas presentes en este, aplicando los pasos del flujo de la figura 11. La siguiente oración constituye el texto a analizar:

*El juez terminó el juicio en la sala Penal del Tribunal Provincial.*

**Paso 1:** teniendo en cuenta que la preposición (en) constituye la preposición sintácticamente ambigua identificada en el texto en análisis, se procede a determinar cuál de los términos escritos después de esta, constituye un sustantivo. La identificación del sustantivo se realiza, comparando cada elemento gramatical que precede la preposición, con los sustantivos que inician cada regla de la gramática definida para en esta segunda parte del método propuesto, hasta encontrar la coincidencia, ver figura 13.





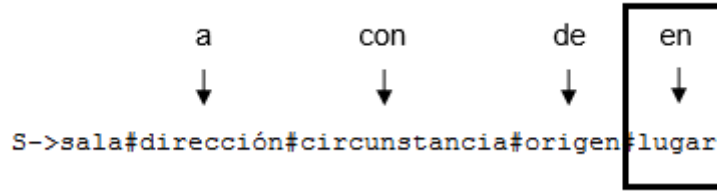
**Leyenda:**

1. Preposición posiblemente ambigua
2. Porción del texto a analizar
3. Término del texto que coincide con uno de los sustantivos de la gramática.

**Figura 13.** Proceso de identificación de un sustantivo que precede una preposición sintéticamente ambigua en un texto. Fuente: elaboración propia

**Paso 2:** una vez identificado el término (*sala*), como el sustantivo que precede la preposición (*en*), en el texto en análisis, se procede a determinar el sentido que ofrece este sustantivo a la preposición. Para encontrar el sentido, se analiza en la gramática la regla de producción que corresponde al sustantivo (*sala*).

Es importante señalar que, en todas las reglas de producción de la gramática, los términos que se encuentran a continuación del sustantivo, constituyen el sentido que estos brindan a cada preposición sintácticamente ambigua en el orden (a, con, de, en). Por tanto, en la oración analizada en el presente ejemplo, el sustantivo (*sala*) le atribuye sentido de lugar a la preposición (*en*), ver figura 14.



**Figura 14.** Sentidos que atribuye el sustantivo sala a las preposiciones sintácticamente ambiguas.  
Fuente: elaboración propia

**Paso 3:** una vez identificado el sentido de la preposición (*en*), todo queda listo para brindar una información detallada al usuario sobre el uso de esta preposición en el texto, sirviendo de ayuda para el logro de una mejor interpretación de este.

Un procedimiento similar al analizado en los tres pasos anteriormente descritos a través del ejemplo, es el que se debe seguir en el tratamiento de cualquiera de las otras preposiciones sintácticamente ambiguas (*a*, *con*, *de*), cuando están presentes en un texto.

### 2.2.3 Brindar información al usuario.

Una vez descrita la primera y segunda parte del método propuesto, se procede a describir la tercera. El propósito de esta parte, es brindar información detallada al usuario sobre las causas que originan problemas de ambigüedad en un texto, a partir de los resultados obtenidos con la aplicación de las dos primeras partes del método. La información brindada al usuario le permite conocer: los problemas estructurales que causan ambigüedad en el texto, el sentido con el cual es utilizada cada preposición sintácticamente ambigua y las posibles soluciones a problemas de ambigüedad originados por la existencia de conjunciones sintácticamente ambiguas. Toda esta información hace posible que el usuario sea asistido por una solución informática en la comprensión de textos ambiguos. En el siguiente ejemplo se demuestra cómo aplicando el tercer paso del método propuesto, se brinda información detallada sobre los posibles problemas de ambigüedad sintáctica presentes en un texto.

Teniendo en cuenta la oración: *En el proceso jurídico participaron acusado, fiscal y juez y abogado.*

Una vez aplicado el primer y segundo paso del método propuesto, se identifican como posibles causas para la ocurrencia de ambigüedades sintácticas, los elementos señaladas en la figura 15.

**En** el proceso jurídico participaron acusado, fiscal **y** juez **y** abogado.

**Figura 15.** Señalización de posibles causas de ambigüedades en un texto. Fuente: elaboración propia

Una vez aplicada la tercera parte del método propuesto, el usuario recibe información detallada sobre las posibles causas que pueden originar ambigüedades sintácticas en el texto en análisis, con el propósito de ser asistido en la interpretación de este, ver figura 16.

Esta oración es posiblemente ambigua por el uso consecutivo de la conjunción sintácticamente ambigua [y] y la existencia de la preposición sintácticamente ambigua [en].

- El sentido que tiene en el texto la preposición sintácticamente ambigua [en] es de lugar.
- Se recomienda sustituir la oración por: **En el proceso jurídico participaron acusado, fiscal, juez y abogado**

En el proceso jurídico participaron acusado, fiscal y juez y abogado.

**Figura 16.** Resultado de aplicar el tercer paso del método propuesto. Fuente: elaboración propia

A partir del resultado de la figura 16, el usuario conoce el sentido con el cual es utilizada la preposición sintácticamente ambigua [en] y recibe una propuesta de solución que elimina el uso consecutivo de la conjunción sintácticamente ambigua [y]. Por tanto, ya se tienen los elementos necesarios para lograr una mejor comprensión del texto, en caso de haber presentado alguna duda en su interpretación inicial, como resultado de los elementos señalados en la figura 15.

De forma similar al ejemplo de la figura 16, se procede en un texto cuando existen otras preposiciones o conjunciones sintácticamente ambiguas. En los casos donde los problemas de ambigüedad sintáctica se originan a partir de la incorrecta estructuración del texto, el usuario obtiene información específica sobre los problemas identificados con la aplicación del paso uno de la primera parte del método propuesto.

### **2.3 Conclusiones parciales**

El desarrollo del diagnóstico inicial, a partir de una encuesta aplicada a juristas vinculados a la Sociedad Cubana de Derecho e Informática, evidenció dificultades en la comprensión de textos jurídicos dadas por la existencia de ambigüedades, así como desconocimiento de la existencia de soluciones informáticas que contribuyan al tratamiento de estos problemas, corroborándose la necesidad de desarrollar la presente investigación.

El diseño de la solución propuesta utilizando gramáticas y recursos del PLN (reglas heurísticas y diccionario), favoreció la identificación de ambigüedades y su procesamiento, haciendo posible la adaptación del método a otros dominios de la ciencia.

La información que se brinda al usuario como resultado de la identificación y procesamiento de ambigüedades, contribuye a mejorar la comprensión de textos de la legislación cubana.

### CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA DE SOLUCIÓN

En el presente capítulo se describen los métodos científicos aplicados en la validación de la solución descrita en el capítulo anterior. Los métodos utilizados fueron Cuasi experimento, Grupo focal y Escalamiento de Osgood, los cuales fueron aplicados sobre una herramienta informática que responde a la implementación del método de desambiguación propuesto. Los resultados obtenidos con la aplicación de cada técnica de validación fueron comparados aplicando triangulación metodológica entre métodos.

#### 3.1. Cuasi Experimento

El experimento es una situación de control en la cual se manipulan, de manera intencional, una o más variables independientes (causas-antecedentes) para analizar las consecuencias que tal manipulación tiene sobre una o más variables dependientes (efectos-consecuentes). Un cuasi experimento se utiliza cuando no es posible la realización de un experimento verdadero. Los diseños cuasi experimentales no tienen garantizada la equivalencia inicial porque no hay asignación aleatoria, ni emparejamiento. Los grupos están formados antes del experimento, son grupos intactos (SAMPIERI et al, 2006).

Los cuasi experimentos se organizan en tres tipos: diseño con post prueba únicamente y grupos intactos, diseño con pre prueba-post prueba y grupos intactos y por último diseño cuasi experimental de series cronológicas (SAMPIERI et al, 2008).

El diseño con pos prueba únicamente y grupos intactos, utilizan dos grupos, uno que recibe el tratamiento experimental y el otro no. Ambos grupos serán comparados en la post prueba para analizar si el tratamiento experimental tuvo efecto sobre la variable dependiente (SAMPIERI et al, 2008).

Los diseños de pre prueba-post prueba y grupos intactos, son similares al que incluye únicamente post prueba y grupos intactos, solo que a los grupos se les administra un pre prueba, la cual sirve para verificar la equivalencia inicial de los grupos (SAMPIERI et al, 2008).

El diseño cuasi experimental de series cronológicas se refiere a la obtención de un conjunto de mediciones u observaciones tomadas en diferentes periodos (SAMPIERI et al, 2008).

### CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA DE SOLUCIÓN

El autor de la presente investigación decide desarrollar un cuasi experimento de tipo, diseño con pre prueba-post prueba y grupos intactos, teniendo en cuenta las características del método de desambiguación sintáctica propuesto; siendo necesario comprobar su validez, evaluando con una misma muestra de usuarios, cómo estos se manifiestan en la comprensión de textos ambiguos, antes y después de aplicar el método. Por tanto, con el uso de este tipo de cuasi experimento se comprueba la relación causa-efecto entre la variable independiente “reducir la ambigüedad sintáctica presente en textos” y la variable dependiente “mejora en la comprensión”. Esta última es desglosada en los indicadores legibilidad y coherencia entre las ideas, con el propósito de comprobar cómo incide sobre ella el cumplimiento de la variable independiente.

La legibilidad según la (RAE, 2016), es la capacidad o posibilidad que tiene un texto para ser leído con claridad. En la validación de la presente investigación este término es comprobado en textos jurídicos, a través de la capacidad que muestren los usuarios en la comprensión clara de estos. Por otro parte la (RAE, 2016) define el término coherencia como la conexión, relación o unión lógica de varios elementos. Asimismo, en el presente trabajo este término se emplea con el fin de evaluar la correcta estructuración de textos jurídicos, de igual forma en que es comprobado el otro indicador.

Para aplicar el cuasi experimento con diseño pre prueba-post prueba y grupos intactos, se tomó una muestra de 20 especialistas del Centro de Gobierno Electrónico, centro de desarrollo de software adjunto a la Facultad 3 de la Universidad de las Ciencias Informáticas, con el fin de comprobar la relación causa-efecto de la variable independiente sobre la dependiente, a través de la evaluación de los indicadores antes definidos, utilizando como instrumento la aplicación de una encuesta, ver Anexo 2. La muestra fue seleccionada teniendo en cuenta que los especialistas de este centro son informáticos y tienen experiencia en el trabajo con textos jurídicos, a partir del perfil del área, siendo encuestadas las 20 personas con mayor tiempo trabajando allí.

Los resultados de la etapa de pre prueba se obtuvieron con la respuesta a las cuatro primeras preguntas de la encuesta, comprobándose el nivel de comprensión que tiene los usuarios sobre un texto jurídico antes de ser procesado por el método propuesto. Luego con las otras preguntas se comprueba en la post prueba cómo los usuarios mejoran en la comprensión del mismo texto, al ser procesado con la aplicación del método. En la tabla 3 se muestra en valores cuantitativos los resultados de aplicar ambas etapas del cuasi experimento.

### CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA DE SOLUCIÓN

**Tabla 3.** Resultados de aplicar el cuasi experimento. Fuente: elaboración propia

<b>PRE PRUEBA</b>			
<b>Muestra  20</b>	<b>Nivel de incoherencias</b>		
	<b>Alto</b>	<b>Medio</b>	<b>Bajo</b>
	14	4	2
	<b>Nivel de legibilidad</b>		
	<b>Alto</b>	<b>Medio</b>	<b>Bajo</b>
	3	4	13
	<b>Consideran afectada la comprensión del texto</b>		
	<b>Sí</b>		<b>No</b>
	17		3
<b>POST PRUEBA</b>			
<b>Muestra  20</b>	<b>Nivel de incoherencias</b>		
	<b>Alto</b>	<b>Medio</b>	<b>Bajo</b>
	0	1	19
	<b>Nivel de legibilidad</b>		
	<b>Alto</b>	<b>Medio</b>	<b>Bajo</b>
	17	2	1
	<b>Mejora la comprensión del texto</b>		
	<b>Sí</b>		<b>No</b>
	18		2

Los siguientes gráficos muestran una comparación entre los resultados del cuasi experimento en las etapas de pre pruebas y post prueba, expuestos en la tabla 3.

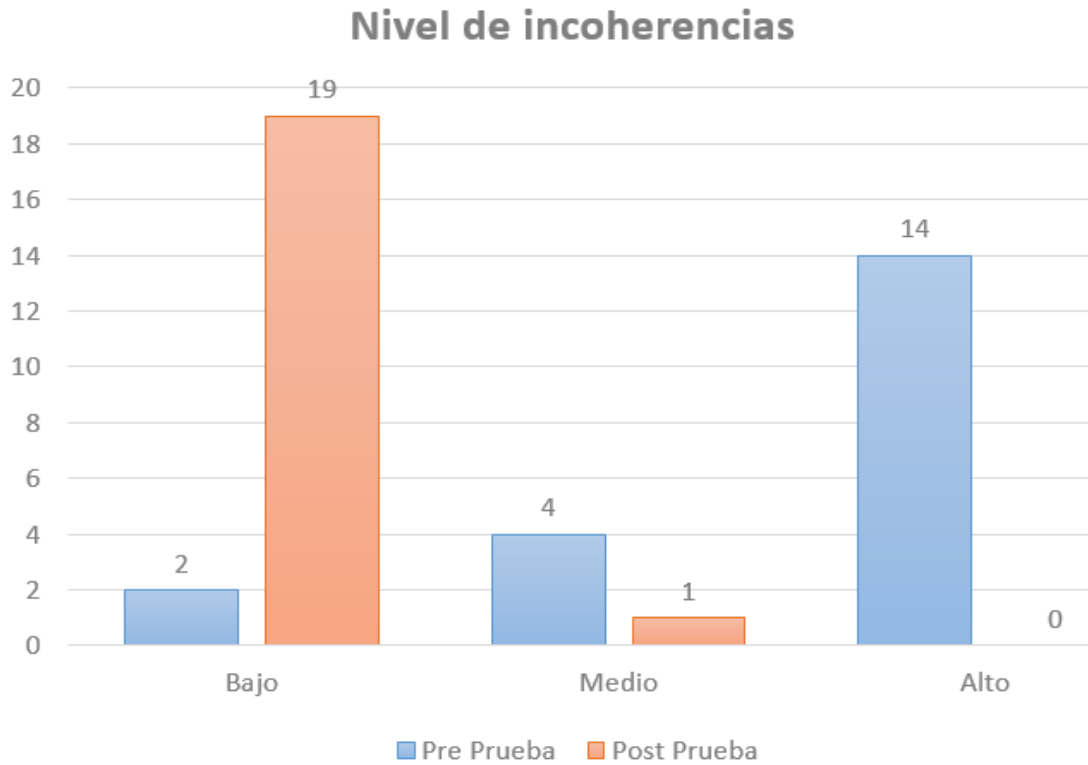


Figura 17. Nivel de incoherencia. Fuente: elaboración propia

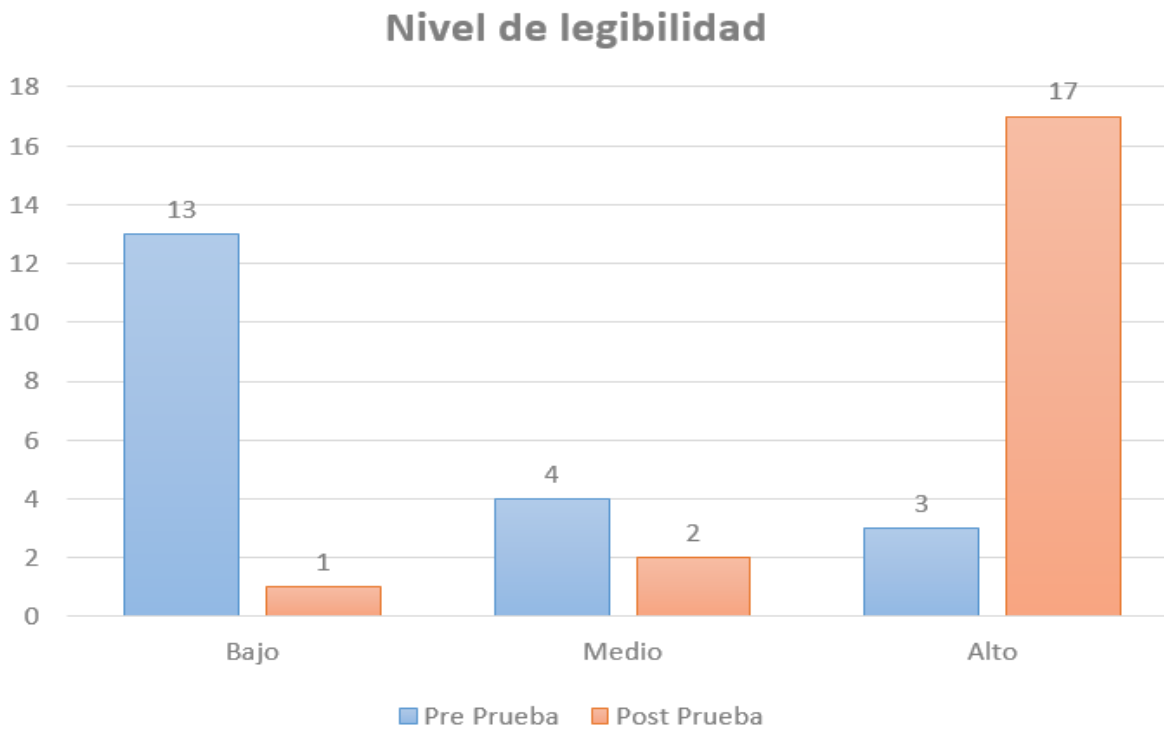
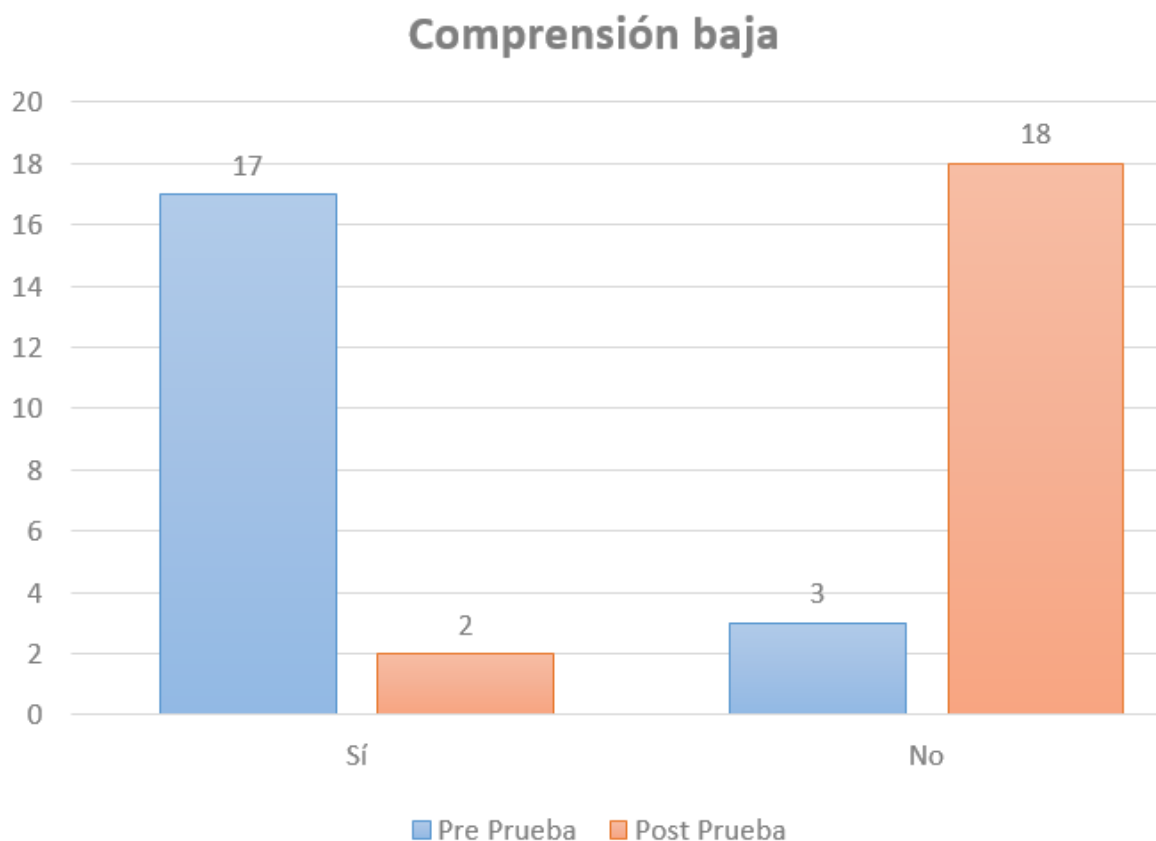


Figura 18. Nivel de legibilidad. Fuente: elaboración propia





**Figura 19.** Evaluación de la comprensión del texto. Fuente: elaboración propia

Los resultados de las gráficas anteriores demuestran en todos los casos que en la etapa de post prueba (con la aplicación del método) cada uno de los indicadores definidos para aplicar el cuasi experimento indican una mejora significativa. Esta es expresada en la evaluación brindada por los usuarios seleccionados para analizar la comprensión que obtienen de un texto ambiguo, antes y después de aplicar el método de desambiguación sintáctica propuesto. De esta forma se puede afirmar que la solución propuesta permite reducir la ambigüedad sintáctica presente en textos de la legislación cubana, contribuyendo a la mejora en la comprensión de estos, cumpliéndose así el objetivo de la presente investigación.

### 3.2. Grupo focal

Grupo focal es un método empírico de investigación, rentable y rápido para obtener información cualitativa y retroalimentación desde un grupo específico de expertos. Este se puede utilizar en varias fases y tipos de investigaciones (MENDOZA-MORENO, et al, 2013).

La aplicación del método resulta adecuada para: obtener retroalimentación de los participantes sobre preguntas de investigación o nuevos conceptos; explorar experiencias pasadas que puedan estudiarse con mayor detalle empleando otros métodos; realizar la evaluación inicial de potenciales soluciones, basado en los practicantes o usuarios; recopilar recomendaciones de lecciones aprendidas o generar ideas; identificar o priorizar potenciales raíces que derivan un fenómeno; obtener retroalimentación sobre la manera en que los modelos o conceptos son presentados o registrados; así como descubrir importantes motivaciones (Kontio, et al, 2008).

En el caso de la presente investigación, Grupo focal se utiliza con el propósito de evaluar la utilidad del método de desambiguación sintáctica propuesto y obtener recomendaciones que contribuyan a su perfeccionamiento. A continuación, se describe la aplicación de este método de validación, como elemento de valoración colectiva de la propuesta de solución, organizado en las cuatro partes del proceso propuesto por (MENDOZA-MORENO, et al, 2013) para aplicar Grupo focal en Ingeniería de Software.

#### **Fase de planeamiento de la investigación:**

Los objetivos definidos en la presente investigación para establecer el debate a través del Grupo focal son: conocer las opiniones de los expertos sobre la validez y utilidad del método propuesto, así como medir la pertinencia e impacto de este.

En esta primera parte de la aplicación del Grupo focal se definieron mediante una guía, ver Anexo 3, los aspectos a evaluar a través de la opinión de los expertos. Esta guía fue elaborada por el autor de la presente investigación y antes de ser presentada a los expertos fue avalada por las Máster en Ciencias Yarina Amoroso Fernández e Isabel González Flores, especialistas en informática jurídica y Ciencias Informáticas respectivamente. Otros de los elementos precisados en la fase inicial fueron: la decisión de presentar el método de desambiguación a través de su ejecución práctica y llevar por escrito las opiniones de los expertos sobre los criterios de evaluación definidos en la guía, para luego ser procesadas y obtener un criterio general que brinde una valoración cualitativa sobre la validez y utilidad de la solución propuesta.

#### **Fase de definición de grupos de discusión**

Una vez definidos los objetivos por los cuales se va a validar a través de este método, la guía de elementos a evaluar y la forma de presentar la solución, se procede a la selección de los expertos. Los criterios tenidos en cuenta en la elección fueron: más del 75 % tener

### CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA DE SOLUCIÓN

experiencia en roles como analista, programador o arquitecto de software, al menos el 25 % presentar grado científico y existir representación de varias áreas. Bajo estas condiciones se seleccionaron 10 personas. En la tabla 4 se realiza un desglose de los expertos seleccionados:

**Tabla 4.** Expertos seleccionados para aplicar Grupo focal. Fuente: elaboración propia

No.	Nombre y Apellidos	Área	Rol o profesión	Grado Científico
1.	Juan P. Febles Rodríguez	Dir. Postgrado	Profesor	Doctor
2.	Mailen Edith Escobar Pompa	Facultad 3	Analista	Máster
3.	Isabel González Flores	CEGEL <sup>22</sup>	Metodóloga	Máster
4.	Yarina Amoroso Fernández	CEGEL	Jurista	Máster
5.	Danae Castillo Domínguez	CESIM <sup>23</sup>	Analista	
6.	Reinier Silverio Figueroa	CEGEL	Arquitecto	
7.	Juan David Gómez Amador	CEGEL	Programador	
8.	Yulia Fustiel Álvarez	CEDAE <sup>24</sup>	Analista	
9.	Robin Sencial Terrero	CEGEL	Arquitecto	
10.	Héctor Luis Duran Cruzata	DESOFT	Programador	

Al quedar seleccionado el grupo de expertos, se procede a la segmentación de estos por subgrupos de debate. Fueron creados tres subgrupos, aplicando como criterio de agrupamiento los roles de cada uno, ver más detalles en la tabla 5. Este criterio se utiliza con el fin de evaluar desde cada rol la validez del método propuesto en cuanto a secuencia lógica de los pasos, elementos empleados en su estructuración, así como la forma de ser implementado para demostrar su utilidad a través de un software.

---

<sup>22</sup> Centro de Gobierno Electrónico

<sup>23</sup> Centro de Informática Médica

<sup>24</sup> Centro de Consultoría y Desarrollo de Arquitecturas Empresariales

## CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA DE SOLUCIÓN

**Tabla 5.** Subgrupos de debate para aplicar Grupo focal. Fuente: elaboración propia

<b>Subgrupo 1</b>				
No.	Nombre y Apellidos	Área	Rol o profesión	Grado Científico
1.	Yulia Fustiel Álvarez	CEDAE	Analista	
2.	Mailen Edith Escobar Pompa	Facultad 3	Analista	Máster
3.	Danae Castillo Domínguez	CESIM	Analista	
<b>Subgrupo 2</b>				
No.	Nombre y Apellidos	Área	Rol o profesión	Grado Científico
1.	Reinier Silverio Figueroa	CEGEL	Arquitecto	
2.	Juan David Gómez Amador	CEGEL	Programador	
3.	Héctor Luis Duran Cruzata	DESOFT	Programador	
4.	Robin Sencial Terrero	CEGEL	Arquitecto	
<b>Subgrupo 3</b>				
No.	Nombre y Apellidos	Área	Rol o profesión	Grado Científico
1.	Juan P. Febles Rodríguez	Dir. Postgrado	Profesor	Doctor
2.	Yarina Amoroso Fernández	CEGEL	Jurista	Máster
3.	Isabel González Flores	CEGEL	Metodóloga	Máster

### **Fase de conducción de la sesión de debate**

Luego de la definición de los subgrupos de trabajo se procede al inicio de las sesiones de debate. Teniendo en cuenta que se crearon 3 subgrupos, se organiza igual cantidad de sesiones. Para cada sesión se fija una fecha y hora de realización. Cada debate es conducido por un moderador, responsabilidad asumida en este caso por el autor del presente trabajo, con el propósito de garantizar efectividad en la aplicación de cada punto de la guía definida en la fase inicial. También se selecciona un especialista que no forma parte del grupo de expertos, para llevar la relatoría de los debates en cada sesión, garantizando así imparcialidad en el proceso. Todos los debates se realizaron con alto nivel profesional, lo que favoreció arribar a consideraciones importantes para la validación cualitativa del resultado de esta investigación. Los criterios emitidos por cada experto fueron claramente relatados por la persona designada para esta tarea.

### Fase de análisis de información y reporte de resultados

Al concluir todas las sesiones de debate, se analizaron las relatorías generadas en cada subgrupo y se elaboró un reporte con todos los resultados. A continuación, se lista un resumen de estos:

- Existió total coincidencia con respecto a la secuencia lógica de los flujos de actividades que describen las partes del método.
- Hubo concordancia con la aprobación de los elementos utilizados en la estructuración del método.
- Se valoró de novedosa la gramática propuesta para identificar los sentidos de las preposiciones en un texto.
- Todos los criterios fueron positivos respecto al cumplimiento del objetivo de la investigación, al demostrarse a través de un ejemplo práctico la contribución de la aplicación del método propuesto a mejorar la comprensión de textos jurídicos ambiguos.
- Las opiniones formuladas en las sesiones de debate, posibilitaron al autor de la presente investigación reconsiderar algunos aspectos relacionados con la descripción del método y la forma de ser implementado, a partir de la valoración constructiva realizada por algunos analistas y arquitectos en aras de elevar la calidad del resultado de la investigación.

Entre las recomendaciones recibidas se encuentran:

- Restructurar la gramática libre de contexto en función de hacerla más flexible ante actualizaciones y cambios, en cuanto a las reglas que la definen.
- Extender el método al análisis de otros tipos de ambigüedades.

### 3.3. Escalamiento de Osgood

Las escalas son instrumentos de recolección de información con base en una lista de ítems, reactivos o frases cuidadosamente seleccionadas, de forma que constituyen un criterio sistemático, confiable, válido y específico para medir cuantitativamente alguna forma de fenómeno, particularmente, actitudes y aquellas relacionadas con sentimientos, opiniones y creencias (Ortega, 2016).

### CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA DE SOLUCIÓN

El escalamiento de Osgood, denominado también diferencial semántico, constituye un conjunto de escalas bipolares que sirven para la evaluación directa de conceptos. Es un método que permite medir el significado que tiene un objeto para un individuo (OSSGOOD, et al, 1982).

La aplicación del diferencial semántico se realiza en el presente trabajo con el propósito de evaluar las opiniones de los usuarios en relación al resultado de la investigación. El instrumento fue aplicado a una muestra de 20 usuarios. Estos fueron seleccionados teniendo en cuenta los mismos criterios de la selección hecha para el cuasi experimento. A continuación, se explica la escala utilizada para evaluar cada uno de los ítems definidos en la encuesta aplicada, ver Anexo 4.

#### Muy satisfactorio 5 4 3 2 1 Insatisfactorio

- 5 Muy satisfactorio
- 4 Satisfactorio
- 3 Neutro
- 2 Poco satisfactorio
- 1 Insatisfactorio

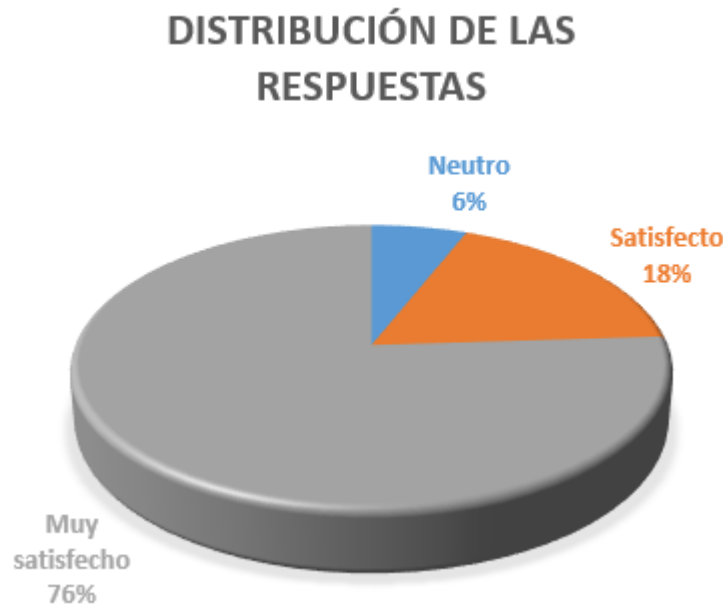
La tabla 6 muestra los 5 ítems evaluados y un resumen con la distribución de las respuestas brindadas por los 20 usuarios sobre cada uno de estos, en función del diferencial semántico utilizado en la escala, una vez aplicada la encuesta.

**Tabla 6.** Resumen de las valoraciones de escala realizadas por los usuarios sobre la aplicación del método de desambiguación sintáctica propuesto. Fuente: elaboración propia

Ítems a evaluar	1	2	3	4	5
La forma en que se evalúa la correcta estructuración sintáctica de los textos.			2	4	14
La identificación de los sentidos de cada preposición sintácticamente ambigua en el texto.				4	16
La propuesta de desambiguación brindada cuando existe una misma conjunción sintácticamente ambigua escrita de forma consecutiva en el texto.				3	17

La información que visualizan los resultados al usuario.			4	5	11
La utilidad de aplicar el método.				2	18

En el gráfico de la figura 20 se muestra el por ciento de satisfacción reflejado por los usuarios, a partir de la aplicación del escalamiento de Osgood, teniendo en cuenta los resultados resumidos en la tabla anterior.



**Figura 20.** Niveles de satisfacción de los usuarios en la escala de Osgood. Fuente: elaboración propia

Estos valores en por ciento revelan que de manera general los usuarios muestran una alta satisfacción sobre los resultados de la investigación propuesta.

### 3.4. Triangulación de resultados

La triangulación es la aplicación y combinación de varias metodologías de la investigación en el estudio de un mismo fenómeno (Sabiote, et al, 2006). Es una técnica utilizada para el análisis de datos recopilados a través de diferentes métodos, lo cual permite analizar una situación desde diversos ángulos. Es un procedimiento de control implementado para garantizar la confiabilidad en los resultados de cualquier investigación debido a que los métodos cualitativos y cuantitativos se consideran complementarios, a partir de que las

### CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA DE SOLUCIÓN

debilidades de un método individual van a ser compensadas por la fortaleza contra balanceadora del otro (Riverol, 2014).

Existen cuatro tipos básicos de triangulación: triangulación de datos con tres subtipos: tiempo, espacio y persona (el análisis de persona, a su vez, tiene tres niveles: agregado, interactivo y colectivo); triangulación de investigador que consiste en el uso de múltiples observadores, más que observadores singulares de un mismo objeto; triangulación teórica que consiste en el uso de múltiples perspectivas, más que de perspectivas singulares; por último está la triangulación metodológica que puede implicar triangulación dentro de métodos y triangulación entre métodos (Valencia, 2012).

El autor de la presente investigación decide aplicar la triangulación metodológica entre métodos, con el propósito de validar los datos recolectados tanto cualitativos como cuantitativos a través de la aplicación de los métodos: cuasi experimento, grupo focal y escalamiento de Osgood. En la tabla 7 se recogen los resultados de la triangulación.

**Tabla 7.** Resumen de la triangulación de métodos. Fuente: elaboración propia

<b>Objetivo a evaluar:</b> desarrollo de un método que permita reducir la ambigüedad sintáctica en textos de la legislación cubana.		
<b>Métodos cuantitativos</b>	Cuasi experimento	Demuestra que los resultados obtenidos en la etapa de post prueba son superiores a la pre prueba, en cuanto a la reducción de la ambigüedad sintáctica presente en los textos y la contribución al mejoramiento en la comprensión de estos.
	Escalamiento de Osgood	La aplicación del diferencial semántico con el propósito de evaluar las opiniones de los usuarios en relación al resultado de la investigación, reveló que la mayoría de los usuarios muestran un por ciento alto de satisfacción sobre los resultados del trabajo.
<b>Método cualitativo</b>	Grupo focal	Con el propósito de conocer las opiniones de los expertos sobre la validez y utilidad del método propuesto, así como medir la pertinencia e impacto de este, se demostró que:



		<ul style="list-style-type: none"> <li>• Existe total coincidencia con respecto a la secuencia lógica de los flujos de actividades que describen las partes del método y se concuerda con la validez de los elementos utilizados en su estructuración.</li> <li>• La propuesta de solución fue valorada de novedosa y tuvo una aceptación favorable.</li> </ul>
<p><b>Conclusiones:</b> Los resultados obtenidos con la aplicación de los métodos cualitativos y cuantitativos demuestran que:</p> <ul style="list-style-type: none"> <li>• Existe una relación causa efecto entre la variable independiente con la dependiente.</li> <li>• La satisfacción de los usuarios respecto a la estructuración y funcionamiento del método propuesto.</li> </ul> <p>Por tanto, queda demostrada la validez del método propuesto para la desambiguación sintáctica de textos de la legislación cubana.</p>		

Luego de aplicar cada uno de los métodos de validación antes descritos se concluye que la hipótesis planteada en el diseño metodológico de la investigación es positiva, cumpliéndose así con el objetivo general de la investigación.

### 3.5. Conclusiones parciales

La utilización del diseño con pre prueba-post prueba y grupos intactos en el desarrollo del cuasi experimento, demostró la relación causa-efecto de la variable independiente sobre la dependiente, evaluando la ambigüedad sintáctica en los textos seleccionados, antes y después de ser procesados a través del método propuesto.

La aplicación de Grupo focal facilitó la evaluación de la utilidad del método de desambiguación sintáctica, corroborando la secuencia lógica de sus flujos de actividades y su contribución en la mejora de la comprensión de textos jurídicos ambiguos.

La aplicación del escalamiento de Osgood, a partir de un conjunto de ítems definidos sobre el diferencial semántico en la escala Muy satisfactorio – Insatisfactorio, permitió evaluar el nivel de aceptación de los expertos sobre la aplicación del método de

### CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA DE SOLUCIÓN

desambiguación sintáctica propuesto, evidenciándose que los usuarios muestran una alta satisfacción sobre los resultados de la investigación.

La integración de los resultados de la investigación aplicando la técnica de triangulación metodológica entre métodos, hizo posible combinar los resultados obtenidos con la aplicación de los métodos científicos: grupo focal, escalamiento de Osgood y cuasi experimento; con el propósito de nivelar las debilidades de unos con las fortalezas de otros, para confirmar que, reduciendo la ambigüedad sintáctica presente en textos de la legislación cubana, se contribuye a mejorar la comprensión de estos.

### CONCLUSIONES GENERALES

El estudio realizado sobre las categorías gramaticales del idioma español, los métodos del PLN como vía para combinar la lingüística y la informática, así como las características de las soluciones existentes para la desambiguación de textos, constituyen la base para la estructuración del método de desambiguación sintáctica propuesto.

El desarrollo del diagnóstico inicial de la investigación constató que las soluciones informáticas existentes para el análisis de ambigüedades en textos del idioma español son pocas y que los problemas de ambigüedad en textos jurídicos son frecuentes, lo que justifica la necesidad e importancia de desarrollar el método propuesto.

La aplicación del método de desambiguación sintáctica propuesto constituye una solución informática que auxilia la redacción de textos jurídicos, haciendo que estos sean promulgados con una mejor calidad de redacción y sin ambigüedades.

El desarrollo del trabajo es importante para la evolución de la informática jurídica en Cuba, pues aporta una solución práctica a un problema que es consustancial al Derecho, toda vez que el lenguaje jurídico es un segmento especializado del lenguaje natural.

El empleo de métodos cualitativos y cuantitativos en la validación del resultado de la presente investigación, demostró la validez y utilidad del método de desambiguación sintáctica propuesto, corroborando que, reduciendo la ambigüedad sintáctica en un texto, se mejora la comprensión de estos.

### RECOMENDACIONES

Agregar a las gramáticas y el lexicón la mayor cantidad de reglas y términos posibles, con el propósito de hacer efectiva la aplicación del método de desambiguación a cualquier texto jurídico.

Convertir el método en un componente informático que pueda ser incorporado al visor de documentos utilizado por algunos softwares jurídicos, con el fin de validar la correcta estructuración de los textos generados a través de estos visores y evaluar el nivel de ambigüedad que se pueda generar.

Extender el alcance del método, al tratamiento de otros tipos de ambigüedades como la léxica y la semántica.

## REFERENCIAS BIBLIOGRÁFICAS

1. Aarnio, A. (1987). Sobre la ambigüedad semántica en la interpretación jurídica. *Repositorio Institucional Universidad de Alicante*. doi:10.14198/DOXA1987.4.07
2. Agirre, E. (2000). *Enriching very large ontologies using the WWW*. Berlin: In Proceedings of ECAI Workshop on Ontology Learning. Obtenido de <https://www.cs.cmu.edu/~hovy/papers/00ECAI-ontol-enrich.pdf>
3. Agirre, E., & Edmonds, P. (2006). Word Sense Disambiguation: Algorithms and Applications. En N. I. Véronis (Ed.). Netherlands: University of the Basque Country and Sharp Laboratories of Europe.
4. Alvarez, G. I. (2013). *Compiladores: Análisis Sintáctico*. Cali: Pontificia Universidad Javeriana Cali.
5. Arte, C. E. (1998). *Arte e identidades culturales*. Oviedo: Universidad de Oviedo. Obtenido de <https://dialnet.unirioja.es/servlet/libro?codigo=2106>
6. Ascaño, M. B. (2015). *La Similaridad de términos en el Lenguaje Jurídico. Similaridad semántica y correspondencia de nombres*. X CONFERENCIA INTERNACIONAL DE DERECHO E INFORMATICA DE LA HABANA.
7. Bisbal, E., Molina, A., Moreno, L., Pla, F., Saiz-Noeda, M., & Sanchis, E. (2003). 3LB-SAT: Una herramienta de anotación semántica. *Procesamiento del lenguaje natural* (Nº 31), pp. 193-200. Obtenido de <http://rua.ua.es/dspace/handle/10045/1510>
8. Bosque, I., & Guitiérrez-Rexach, J. (2009). *Fundamentos de Gramática Formal*. Departamento de Tecnologías de la Información. Madrid: Universidad de Huelva.
9. Calvo, R. A. (2009). Lexicalización y colocaciones: una introducción a su estudio diacrónico. *Redalyc* (núm. 19), pp. 33-56. Obtenido de <http://www.redalyc.org/pdf/1345/134513178002.pdf>
10. Campos, M. d., Simón Cuevas, A., & Olivas, J. A. (2014). Recuperación de conceptos similares de un Corpus de Mapas Conceptuales. *Redalyc*, pp. 37 - 44. Obtenido de [http://www.redalyc.org/pdf/1814/Resumenes/Resumen\\_181433733006\\_1.pdf](http://www.redalyc.org/pdf/1814/Resumenes/Resumen_181433733006_1.pdf)
11. Coutiño, M. A. (2006). PASCQA : búsqueda de respuestas con base en anotación predictiva de contextos léxico-sintácticos. *Procesamiento del lenguaje natural* (n. 36), pp. 59-60.
12. Cruzata, H. L. (2011). *Procedimiento para el diseño de un Sistema de Información Basado en Ontologías*. La Habana.

13. Cucatto, M. (2013). El lenguaje jurídico y su “desconexión” con el lector especialista. El caso de a mayor abundamiento. *Letras de Hoje*, pp. 127-138.
14. Cueto, A. S. (2004). *Resolución de la ambigüedad semántica de las palabras mediante modelos de probabilidad de máxima entropía*. Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante. Obtenido de <https://rua.ua.es/dspace/bitstream/10045/4070/1/Su%C3%A1rez%20Cueto,%20Armando.pdf>
15. Gavilán, C. M. (2009). *Lenguajes documentales. Principales tipos de clasificación. Encabezamientos de materia, descriptores y tesauros*. Obtenido de <http://eprints.rclis.org/14817/1/lendoc.pdf>
16. Gil, M. J. (1990). *Fundamentos del Análisis semántico*. Santiago de Compostela: Universidad de Santiago de Compostela.
17. Haro, S. N., & Gelbukh, A. (2007). *INVESTIGACIONES EN ANÁLISIS SINTÁCTICO PARA EL ESPAÑOL*. (PRIMERA EDICIÓN ed.). México: Instituto Politécnico Nacional.
18. Hernández, C. P. (2002). El corpus en los estudios lingüísticos. En *Estudios de Lingüística del Español (ELiEs)* (Vol. 18). Universidad de Málaga. Obtenido de <http://elies.rediris.es/elies18/index.html>
19. Hernández, C. U. (30 de Septiembre de 2000). Procesamiento Computacional del Lenguaje Natural. Victoria, Tamaulipas, México.
20. Hopcroft, J. E., Motwani, R., & Ullman, J. (2001). *Automata Theory, Language and Computation*. (Tercera edición ed.). Addison-Wesley .
21. Kontio, J., Bragge, J., & Lehtola, L. (2008). *The Focus Group Method as an Empirical Tool in Software Engineering*. London: Springer London. doi:1007/978-1-84800-044-5\_4
22. Leal, E. T. (17 de Abril de 2009). *La Desambiguación del Sentido de las Palabras: revisión metodológica*. Obtenido de No solo usabilidad: revista sobre personas, diseño y tecnología: <http://www.nosolousabilidad.com/articulos/desambiguacion.htm>
23. Llorach, E. A. (1994). *Unidades de relación: las preposiciones*. Madrid: Espasa Calpe. Obtenido de [http://www.geocities.ws/rufi\\_metal/preposiciones](http://www.geocities.ws/rufi_metal/preposiciones)
24. López, L. A. (2002). *Resolución de la ambigüedad léxica en tareas de clasificación automática de documentos*. Alicante: Club Universitario. Obtenido de [www.sepln.org/wp-content/uploads/2011/02/monografiaUrena.pdf](http://www.sepln.org/wp-content/uploads/2011/02/monografiaUrena.pdf)

25. Manterola, I., Diaz de Ilarraza, A., Gojenola, K., & Sarasola, K. (2010). Recursos en euskera para la herramienta NLTK para enseñanza de procesamiento del lenguaje natural. *Procesamiento del Lenguaje Natural* (n. 45), pp 305-306.
26. Maña, J. P. (2014). *Ambigüedad del lenguaje jurídico*. Obtenido de <http://es.slideshare.net/lireh/ambigüedad-del-lenguaje-juridico-11840148>
27. Marco, A. M. (2004). *Desambiguación en procesamiento del lenguaje natural mediante técnicas de aprendizaje automático*. Valencia. Obtenido de [www.dsic.upv.es/docs/bib-dig/tesis/etd-11262003.../TesisAntonioMolina.pdf](http://www.dsic.upv.es/docs/bib-dig/tesis/etd-11262003.../TesisAntonioMolina.pdf)
28. Marin, A., & Sifontes, A. (15 de Marzo de 2013). *Matemática discreta última evaluación*. Recuperado el 2016, de <http://losgruposdematematicadiscretaseccion2.blogspot.com/2013/03/lenguaje-gramatica.html>
29. Martínez, E. A. (2014). *Gramáticas libres de contexto*. Obtenido de <http://computacion.cs.cinvestav.mx/~efranco>
30. MENDOZA-MORENO, M., GONZÁLEZ-SERRANO, C., & PINO, F. (2013). FOCUS GROUP COMO PROCESO EN INGENIERÍA DE SOFTWARE: UNA EXPERIENCIA DESDE LA PRÁCTICA. *Dyna*, pp. 51-60.
31. Mitkov, R. (2012). *The Oxford Handbook of Computational Linguistics*. doi:10.1093/oxfordhb/9780199276349.001.0001
32. Miyao, Y. (2011). *A model of syntactic disambiguation based on lexicalized grammars*. Edmonton. Obtenido de [www.aclweb.org/anthology/W03-0401](http://www.aclweb.org/anthology/W03-0401)
33. Monteagudo, M. A. (2013). *La ambigüedad en el lenguaje jurídico: ¿Amplitud o distorsión semántica?* Perú. Obtenido de [www.realiter.net/wp-content/.../Journee\\_Realiter\\_2008\\_Larrauri\\_Monteagudo-2.pdf](http://www.realiter.net/wp-content/.../Journee_Realiter_2008_Larrauri_Monteagudo-2.pdf)
34. Ortega, C. E. (2016). Estrategia sustentada en las Tecnologías de la Información y las Comunicaciones para contribuir al desarrollo de habilidades blandas en la enseñanza básica elemental del Liceo Panamericano de la Ciudad de Guayaquil. *Revista Cubana de Educación Superior*, v. 35 (n. 2). Obtenido de [http://scielo.sld.cu/scielo.php?pid=S0257-43142016000200003&script=sci\\_arttext&tlng=es](http://scielo.sld.cu/scielo.php?pid=S0257-43142016000200003&script=sci_arttext&tlng=es)
35. OSSGOOD, C., SUCI, G., & TANNENBAUM, P. (1982). *Medición de actitudes*. México.
36. Paredes, V. E. (2011). MÉTODOS DE INTERPRETACIÓN JURÍDICA. *UNAM*, V. 16 (n. 6). Obtenido de <https://revistas-colaboracion.juridicas.unam.mx/index.php/quid-iuris/article/view/17406/15614>

37. Pedersen, T. (2011). *A decision tree of bigrams is an accurate predictor of word sense*. Minnesota. Obtenido de <http://www.aclweb.org/anthology/N01-1011>
38. Pérez, S. V. (2009). *Resolución de la ambigüedad semántica mediante métodos basados en conocimiento y su aportación a tareas de PLN*. Alicante: Universidad de Alicante. Obtenido de [http://rua.ua.es/dspace/bitstream/10045/11456/1/Tesis\\_vazquez.pdf](http://rua.ua.es/dspace/bitstream/10045/11456/1/Tesis_vazquez.pdf)
39. Phillips, S. J., Anderson, R., & Schapire, R. (2006). MAXIMUM ENTROPY MODELING OF SPECIES GEOGRAPHIC DISTRIBUTIONS. *Ecological Modelling*, v. 190, pp. 231-259. Obtenido de <http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026>
40. Pólya, G. (1945). *How to solve it. A new Aspect of Mathematical Method*. New Jersey: Princeton University Press.
41. RAE. (2016). Obtenido de <http://www.rae.es/>
42. Riverol, Y. B. (2014). Estrategia para estructurar las evaluaciones de un laboratorio de pruebas de software desde la perspectiva de la acreditación. *Revista Cubana de Ciencias Informáticas*, v. 8 (n. 4). Obtenido de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2227-18992014000400009](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992014000400009)
43. Sabiote, C. R., Pozo Llorente, T., & Gutiérrez Pérez, J. (2006). La triangulación analítica como recurso para la validación de estudios de encuestas recurrentes e investigaciones de réplica en Educación Superior. *RELIEVE*, v. 12 (n. 2), p. 289-305. Obtenido de [http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2\\_6.pdf](http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_6.pdf)
44. SAMPIERI, R. H., & COLLADO, C. (2006). *Metodología de la Investigación. 4ta edición*. México: Mc Graw Hill. Obtenido de [https://competenciashg.files.wordpress.com/2012/10/sampieri-et-al-metodologia-de-la-investigacion-4ta-edicion-sampieri-2006\\_ocr.pdf](https://competenciashg.files.wordpress.com/2012/10/sampieri-et-al-metodologia-de-la-investigacion-4ta-edicion-sampieri-2006_ocr.pdf)
45. SAMPIERI, R. H., & COLLADO, C. (2008). *Metodología de la investigación (5ta edición ed.)*. Colombia: Mcgraw-Hill. Obtenido de [https://www.esup.edu.pe/descargas/dep\\_investigacion/Metodologia%20de%20la%20investigaci%C3%B3n%205ta%20Edici%C3%B3n.pdf](https://www.esup.edu.pe/descargas/dep_investigacion/Metodologia%20de%20la%20investigaci%C3%B3n%205ta%20Edici%C3%B3n.pdf)
46. Sassi, M., Catarsi, M., Ratti, D., & Saba, A. (1981). *Ordenadores y Lengua Española (Vol. 1)*. Pisa: Giardani Editori e Stampatori.
47. Slager, E. (2004). *Diccionario de uso de las preposiciones españolas (Vol. 1)*. Madrid: Espasa Calpe.
48. Soares, F. (2008). Legimática : a tecnologia da informação aplicada à qualidade da produção legislativa. *Senatus : cadernos da Secretaria de Informação e Documentação*, v. 6 (n.2), p. 18-32. Obtenido de <http://www2.senado.leg.br/bdsf/item/id/141095>



49. Sznur, S. (2012). *Análisis de consistencia de la legislación de defensa del consumidor mediante métodos formales*. Mar del Plata, Argentina: 1er Congreso Iberoamericano de Investigadores y Docentes de Derecho e Informática - CIIDDI 2012. Obtenido de <http://ciiddi.org/congreso2014/images/documentos/anlisis%20de%20consistencia%20de%20la%20legislacin%20de%20defensa%20del%20consumidor%20mediante%20mtodos%20formales%20sznur.pdf>
50. Sznur, S., Giaccaglia, M., & Belén Ruffa, M. (2015). Guía de buenas prácticas en el análisis de una ley mediante métodos formales. *V Congreso Iberoamericano de Investigadores y Docentes en Derecho e Informática*.
51. Torrego, L. G. (2007). *GRAMÁTICA DIDÁCTICA DEL ESPAÑOL*. (Vol. 1). Madrid: EDICIONES SM.
52. Torres Ramos, S. (2009). *Optimización global de coherencia en la desambiguación del sentido de las palabras*. México, D.F.: INSTITUTO POLITÉCNICO NACIONAL, CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN. Obtenido de <http://www.gelbukh.com/thesis/Sulema%20Torres%20Ramos%20-%20PhD.pdf>
53. Valdivia, T. M., García Vega, M., & Ureña López, L. (2011). *Resolución de la ambigüedad mediante redes neuronales*. Madrid. Obtenido de [http://www.dlsi.ua.es/projectes/srim/publicaciones/jaen/WSD\\_LVQ\\_SEPLN2002\\_DEF.pdf](http://www.dlsi.ua.es/projectes/srim/publicaciones/jaen/WSD_LVQ_SEPLN2002_DEF.pdf)
54. Valencia, M. M. (2012). *LA TRIANGULACIÓN METODOLÓGICA: SUS PRINCIPIOS, ALCANCES Y LIMITACIONES*. Obtenido de <http://www.uv.mx/mie/files/2012/10/Triangulacionmetodologica.pdf>
55. Vázquez, S., Montoyo, A., & Rigau Claramunt, G. (2003). Método de desambiguación léxica basada en el recurso léxico : dominios relevantes. *Procesamiento del lenguaje natural*.(31), pp. 141-148.
56. Verdejo, F., Gonzalo, J., Fernández, D., Peñas, A., & López, F. (2012). *ITEM: Un Motor de Búsqueda Multilingüe Basado en Indexación Semántica*. España: Dpto. de Lenguajes y Sistemas Informáticos UNED. Obtenido de [http://infonautica.net/docs/jbidi/jbidi2000/13\\_2000.pdf](http://infonautica.net/docs/jbidi/jbidi2000/13_2000.pdf)
57. Wijnakker, P. (2011). *El latinismo en español*. Universidad de Gante. Obtenido de [http://lib.ugent.be/fulltxt/RUG01/001/786/600/RUG01-001786600\\_2012\\_0001\\_AC.pdf](http://lib.ugent.be/fulltxt/RUG01/001/786/600/RUG01-001786600_2012_0001_AC.pdf)
58. Zapata, C., PALOMINO, K., & ROSERO, R. (noviembre de 2008). UN MÉTODO PARA LA DESAMBIGUACIÓN SINTÁCTICA DE TIPO COORDINATIVO Y PREPOSICIONAL. *Dyna*, 75 (n. 156), pp. 29-42. Obtenido de <http://www.redalyc.org/articulo.oa?id=49612071004>

**ANEXOS**

**Anexo 1:** Encuesta aplicada para el desarrollo del diagnóstico de la investigación.

**Encuesta sobre la existencia de soluciones informáticas para la reducción de la ambigüedad sintáctica en textos de la legislación cubana**

Con el propósito de contribuir al desarrollo de una investigación que responde a una tesis de Maestría en Informática Avanzada, le pedimos que responda de forma anónima el siguiente cuestionario:

1. ¿Usted ha presentado dificultades en la comprensión de textos jurídicos por causa de ambigüedades existentes en estos?

Sí \_\_\_\_ No \_\_\_\_

1. ¿Usted conoce de soluciones informáticas desarrolladas en Cuba, dirigidas a la detección y/o reducción de ambigüedades en textos jurídicos?

Sí \_\_\_\_ No \_\_\_\_

2. En caso de ser positiva la respuesta anterior. ¿A qué nivel han llegado estos trabajos?

Nivel teórico \_\_\_\_ Nivel práctico \_\_\_\_\_

3. En caso de ser positiva la respuesta de la pregunta 2, mencione algunos de estos trabajos.

---



---



---



---



---

4. ¿Considera usted necesaria la existencia de una solución informática que le auxilie en la comprensión de un determinado texto jurídico con problemas de ambigüedad?

Sí \_\_\_\_ No \_\_\_\_

5. ¿Qué nivel de importancia le atribuye usted al desarrollo de una solución informática que le auxilie en la comprensión de un determinado texto jurídico con problemas de ambigüedad?

Alto \_\_\_\_ Medio \_\_\_\_ Bajo \_\_\_\_

**Anexo 2:** Encuesta aplicada para el desarrollo del cuasi experimento.

**Encuesta para evaluar cómo los usuarios aumentan la comprensión sobre un determinado texto jurídico, al reducirse las ambigüedades presentes en este.**

Con el propósito de contribuir a la validación de una investigación que responde a una tesis de Maestría en Informática Avanzada, le pedimos que responda de forma anónima el siguiente cuestionario.

Antes de comenzar, es necesario que lea el texto que le estamos entregando.

1. ¿Usted considera ambigua alguna de las partes del texto?

Sí \_\_\_\_ No \_\_\_\_

2. En caso de ser positiva la respuesta anterior. ¿Considera usted que estos problemas de ambigüedad lo afectan en la comprensión del texto?

Sí \_\_\_\_ No \_\_\_\_

3. En caso de ser positiva la respuesta a la pregunta 1. ¿Cuáles de las siguientes causas le afectan en la comprensión del texto?

\_\_\_\_ Incoherencias en las ideas

\_\_\_\_ Sus ideas no son legibles (claras)

\_\_\_\_ Otras

4. En correspondencia a las respuestas de la pregunta anterior diga:

- a) ¿Cómo califica el nivel de incoherencias identificadas en la lectura del texto?

\_\_\_\_ Alto \_\_\_\_ Medio \_\_\_\_ Bajo

- b) ¿Cómo califica la legibilidad (claridad) de las ideas en el texto?

\_\_\_\_ Alto \_\_\_\_ Medio \_\_\_\_ Bajo

En caso de ser positiva la respuesta a la pregunta 1, le pedimos que procese el texto a través de la herramienta informática que le ofrecemos, la cual responde a la implementación del método de desambiguación sintáctica propuesto. Una vez procesado el texto, responda:

5. ¿Considera que la información que brinda la herramienta al concluir el proceso, contribuye a reducir los problemas de ambigüedades identificadas por usted en el texto?

Sí \_\_\_\_ No \_\_\_\_

6. En caso de ser positiva la respuesta a la pregunta 5. La reducción de los problemas de ambigüedad se logra en cuanto a:

\_\_\_\_ Eliminar incoherencias entre las ideas

\_\_\_\_ Hacer las ideas más legibles (claras)

\_\_\_\_ Otras

7. En correspondencia a las respuestas de la pregunta anterior diga:

a) ¿En qué nivel usted considera que se eliminan los problemas de incoherencias?

\_\_\_\_ Alto    \_\_\_\_ Medio    \_\_\_\_ Bajo

b) ¿En qué nivel usted considera que las ideas en el texto se hacen más legibles (claras)?

\_\_\_\_ Alto    \_\_\_\_ Medio    \_\_\_\_ Bajo

8. En correspondencia con las respuestas anteriores, ¿considera usted que con la aplicación del método se mejora la comprensión del texto?

Sí \_\_\_\_ No \_\_\_\_

**Anexo 3:** Guía de moderación aplicada en el Grupo focal.***Guía de preguntas para Grupo focal***

Viernes, 25 de noviembre 2016.

Buenos días, queremos comenzar dándoles las gracias por haber aceptado formar parte de este grupo de expertos seleccionado para aplicar Grupo focal sobre la investigación “Método para la desambiguación sintáctica de textos de la legislación cubana”. Mi nombre es Yordanis Garcia Leiva, soy matrícula de la maestría en Informática Avanzada y la investigación a valorar por ustedes constituye mi trabajo de tesis.

El encuentro tendrá una duración aproximada de 30 a 45 minutos y cada uno de los criterios expuestos por ustedes, serán relatados en un informe para su posterior análisis. Todos los comentarios que tengan sobre el método son importantes, aquí no hay respuestas correctas o incorrectas, lo que importa es su opinión sincera.

El debate se debe centrar en los siguientes puntos:

1. Luego de presentada la estructura del método, ¿qué opinión tienen sobre la secuencia lógicas de los flujos de actividades utilizados en su descripción?
2. ¿Cómo valoran la completitud de las reglas heurísticas definidas?
3. ¿Consideran que debe definirse alguna otra regla siguiendo los criterios por los cuales fueron formuladas las demás?
4. ¿Cómo valoran el diseño del lexicón propuesto?
5. ¿Qué opinión tienen sobre el diseño de las gramáticas utilizadas?
6. ¿Consideran que con el desarrollo del método se da cumplimiento al objetivo general de la investigación?
7. ¿Qué beneficios le atribuyen a la aplicación del método?
8. ¿Qué mejoras consideran que se le pueden hacer a la solución propuesta desde cada uno de los roles que ustedes representan?
9. Finalmente, ¿desean agregar algún otro punto o comentario al debate?

¡Muchas Gracias!

**Anexo 4:** Encuesta aplicada para el desarrollo del escalamiento de Osgood.**Encuesta aplicado diferencial semántico para evaluar la opinión de los usuarios en relación a la aplicación del método para la desambiguación sintáctica de textos de la legislación cubana**

Con el propósito de obtener su valoración sobre la aplicación del resultado de una investigación que responde a una tesis de Maestría en Informática Avanzada, le pedimos que nos colabore de forma anónima con la siguiente encuesta.

A continuación, se presentan 5 resultados a evaluar, le solicitamos que marque con una X la evaluación que le daría usted a cada uno, en función de lo observado con la aplicación del método. Si marca 1 significa que usted ha quedado insatisfecho con el resultado y si marca 5 es porque está muy satisfecho con el mismo. Usted puede marcar cualquiera de los números intermedios que mejor representen su opinión.

Ítems a evaluar	1	2	3	4	5
La forma en que se evalúa la correcta estructuración sintáctica de los textos.					
La identificación de los sentidos de cada preposición sintácticamente ambigua en el texto.					
La propuesta de desambiguación brindada cuando existe una misma conjunción sintácticamente ambigua escrita de forma consecutiva en el texto.					
La información que visualizan los resultados al usuario.					
La utilidad de aplicar el método.					