



UNIVERSIDAD NACIONAL DE COLOMBIA

---

# **TRANSFORMACIÓN AUTOMÁTICA DE TEXTO A GRAFOS CONCEPTUALES**

**Sonia Ordoñez Salinas**

Universidad Nacional de Colombia  
Facultad Ingeniería, Departamento de Sistemas e Industrial  
Doctorado en Ingeniería Sistemas y Computación  
Ciudad, Colombia  
2011



# **TANSFORMACIÓN AUTOMÁTICA DE TEXTO A GRAFOS CONCEPTUALES**

**Sonia Ordoñez Salinas**

Tesis o trabajo de investigación presentada como requisito parcial para optar al  
título de:

**Doctor en Ingeniería de Sistemas y Computación**

Director:

Ph.D., Alexander Gelbukh

Línea de Investigación:

Procesamiento de Lenguaje Natural y Lingüística Computacional

Universidad Nacional de Colombia  
Facultad Ingeniería, Departamento Sistemas e Industrial  
Doctorado en Ingeniería de Sistemas y Computación  
Ciudad, Colombia

2011



*A quienes me han apoyado incondicionalmente, mi compañero y amigo de vida Alvaro, mis adorados hijos Sebastian y Laura, a mi amada madre y hermanos.*

## **Agradecimientos**

Agradezco la asesoría y colaboración del profesor Alexander Gelbukh, director del proyecto, ya que sin su ayuda no hubiera sido posible la realización de esta investigación. Quiero de manera especial resaltar su empeño y dedicación hacia la investigación y la transmisión del conocimiento.

De manera muy especial quiero reconocer el gran apoyo brindado por el profesor Luis Fernando Niño, quien con su desinteresada colaboración a todos los estudiantes ha permitido la consolidación del doctorado y los procesos de investigación en la Universidad.

De igual forma agradezco el cuerpo de Docentes entre los que me gustaria destacar al profesor Fabio Gonzalez. Quiero hacer extensivo estos agradecimientos al cuerpo administrativo de la facultad y de la Universidad en General.

Por último quiero expresar mi sentido de gratitud a la Universidad Distrital F.J.C por la comisión de estudios que me fue otorgada y que me permitió dedicarme a desarrollar esta investigación y al Instituto Politécnico Nacional de México por permitirme realizar la pasantía.

## Resumen

Dentro de las estructuras computacionales utilizadas para el procesamiento del lenguaje natural, se encuentran los grafos conceptuales (GCs) que por sus ventajas se pueden convertir en un instrumento eficiente para la representación formal del significado del texto. Usualmente los GCs son creados utilizando una estructura sintáctica que es modificada hasta obtener los GCs, lo que involucra dos procesos costosos: el análisis sintáctico y la transformación al los GCs. Esta tesis propone dos soluciones más simples para la transformación directa de texto a grafos conceptuales, sin involucrar una estructura intermedia.

Una solución es a través de la construcción de una gramática especializada. Dicha gramática se crea de forma automática a partir de un recurso léxico existente construido para otro fin.

Otra solución propuesta, es a través del desarrollo de un analizador sintáctico basado en un método estadístico no supervisado y pobre en conocimiento. El método usa las estadísticas de coocurrencia de las palabras en un corpus grande (o en Internet) y permite obtener estructuras simplificadas de GCs. Cabe mencionar que para su entrenamiento no se requiere ningún trabajo manual previo ya que se basa en un corpus no preparado, más aún, sólo se requieren las estadísticas de coocurrencia y no se requiere el acceso al corpus mismo, lo que es importante en el caso del uso de los motores de búsqueda en Internet como la fuente de información estadística.

**Palabras clave:** Procesamiento de lenguaje natural, Lingüística computacional, Grafos conceptuales, Análisis sintáctico no supervisado, Gramática especializada.

## Abstract

One of the computational structures used in natural language processing is the Conceptual Graphs (CGs), which thanks to their advantages can become an efficient tool for formal representation of the meaning of the text. GCs are usually created using some syntactic structure that is modified to obtain the CGs. This involves two costly processes: parsing and transformation to the CGs. This thesis proposes two simpler solutions for converting text to conceptual graphs directly, without involving an intermediate structure.

One solution is through the construction of a specialized grammar. This grammar is automatically created from an existing lexical resource that has been built for another purpose.

The other proposed solution is through the development of a parser based on a knowledge-poor unsupervised statistical method. The method uses co-occurrence statistics of words in a large corpus (or Internet) and generates simplified CG-like structures. It should be mentioned that the training process does not require any previous manual work because it is based on a raw text corpus. Moreover, it only requires the co-occurrence statistics and does not require access to the corpus itself, which is important for the use of Internet search engines as the source of statistical information.

**Keywords:** Natural language processing, Computational linguistics, Conceptual graphs, unsupervised parsing, Specialized grammar.



# Contenido

<i>1.INTRODUCCIÓN</i>	<i>1</i>
1.1 HIPÓTESIS	2
1.2 ESTRUCTURA DEL DOCUMENTO	3
<i>2. ESTADO DEL ARTE Y MARCO TEÓRICO</i>	<i>4</i>
2.1 ESTRUCTURAS COMPUTACIONALES EN EL PLN	5
2.1.1 Estructuras básicas	5
2.1.2 Estructuras avanzadas	8
2.1.3 Estructuras conceptuales estandarizadas	11
2.2 GRAFOS CONCEPTUALES (GCs)	16
2.2.1 Características de los GCs	16
2.2.2 Herramientas para generar grafos conceptuales	17
2.2.3 Recuperación automática de grafos conceptuales	22
<i>3.MÉTODO BASADO EN REGLAS SINTÁCTICAS Y SEMÁNTICAS</i>	<i>25</i>
3.1 INTRODUCCIÓN	25
3.2 TRABAJO PREVIO	25
3.2.1 La tarea principal no corresponde a la Creación de los GC	26
3.2.2 La tarea es restringida a un vocabulario controlado	26
3.2.3 Se crean reglas sobre estructuras sintácticas	27
3.2.4 Uso de recursos	27
3.3 TRABAJO PROPUESTO	27
3.3.1 Los grafos conceptuales y las estructuras sintácticas	28
3.3.2 Gramática para los grafos conceptuales (gramática-GC)	33
3.3.3 Parsing cic-ipn	38
3.3.4 Validación los grafos conceptuales	40

<b>4.MÉTODO ESTADÍSTICO NO SUPERVISADO</b>	<b>47</b>
4.1      “PARSING” PROBABILÍSTICOS .....	47
4.1.1      Modelo de Caroll y Charniak.....	48
4.1.2      Modelo de Paskin y Yuret .....	49
4.1.3      Modelo de Klein y Manning.....	49
4.2      LAS ESTRUCTURAS DE DEPENDENCIA Y LOS GRAFOS CONCEPTUALES .....	49
4.3      ASOCIACIONES POR CO-OCURRENCIAS .....	51
4.4      ESTRUCTURA SIMPLIFICADA DE GRAFOS CONCEPTUALES .....	52
4.5      PARSING PARA GENERAR UNA ESTRUCTURA SIMPLIFICADA DE GC .....	56
4.5.1      Método.....	56
4.5.2      Algoritmo voraz.....	58
4.5.3      Resultados experimentales del parsing .....	58
<b>5.RESULTADOS DE LA INVESTIGACIÓN</b>	<b>60</b>
5.1      CONCLUSIONES Y TRABAJO FUTURO .....	60
5.2      APORTES .....	60
5.3      PRODUCTOS ENTREGABLES .....	61
5.4      PUBLICACIONES .....	62

## Lista de figuras

<b>Figura No. 1.</b> Generación de GC utilizando una estructura sintáctica	26
<b>Figura No. 2.</b> Generación de CG por medio de reglas gramaticales	28
<b>Figura No. 3.</b> Árbol de constituyente para la frase “Economic news had little effect on financial markets”	30
<b>Figura No. 4.</b> Árbol de Dependencia para la frase “Economic news had little effect on financial markets”	31
<b>Figura No. 5.</b> GC de “Economic news had little effect on financial markets”	32
<b>Figura No. 6.</b> Promedio de Precisión media (Mean Average Precision MAP). vs. $\beta$	45
<b>Figura No. 7.</b> Promedio de Precisión Media, sin normalizar los pesos, como una función relativa al peso de las relaciones y los resultados normalizada.	46
<b>Figura No. 8.</b> Estructura de dependencia para “From my infancy I was noted for the docility and humanity of my disposition”	50
<b>Figura No. 9.</b> Estructura de GC para “From my infancy I was noted for the docility and humanity of my disposition”	51
<b>Figura No. 10.</b> Estructura simplificada de GC para “From my infancy I was noted for the docility and humanity of my disposition”	52
<b>Figura No. 11.</b> Estructura simplificada de GC para “noted” y “docility” de la frase “From my infancy I was noted for the docility and humanity of my disposition”	53
<b>Figura No. 12.</b> Estructura de GC para “noted” y “docility” de la frase “From my infancy I was noted for the docility and humanity of my disposition”. Donde theme se assume como un rol semántico	53
<b>Figura No. 13.</b> Estructura simplificada de GC para “When reason returned with the morning - when I had slept off the fumes of the night's debauch - I experienced a sentiment half of horror, half of remorse, for the crime of which I had been guilty; but it was, at best, a feeble and equivocal feeling, and the soul remained untouched”	54
<b>Figura No. 14.</b> Estructura de dependencia para “My tenderness of heart was even so conspicuous as to make me the jest of my companions”	54

**Figura No. 15.** Estructura de GC simplificada para “My tenderness of heart was even so conspicuous as to make me the jest of my companions” 55

**Figura No. 16.** Estructura de dependencia (izquierda )vs GC simplificada (derecha) para “succession of causes and effects” 55

**Figura No. 17.** Comportamiento de la precisión para diferentes valores de alfa y beta 58

## Lista de tablas

<i>Cuadro No. 1. Clasificación palabras que enriquecen los nodos concepto de los GC.</i>	35
<i>Cuadro No. 2. Regla Noun_Like: de palabras que puede funcionar como un Sustantivo en grammar-GC.</i>	36
<i>Cuadro No. 3. Reglas para Atributo como nodo concepto</i>	37
<i>Cuadro No. 4. Ejemplo de Roles semánticos</i>	38
<i>Cuadro No. 5. Ejemplo se reglas para Verbo como nodo-concepto</i>	39
<i>Cuadro No. 6. Reglas que denotan un Atributo</i>	40
<i>Cuadro No. 7. Preguntas de la colección ImageClef2008</i>	41
<i>Cuadro No. 8. Ejemplo de salida del Analizador Sintáctico CIC-IPN (Grafo conceptual)</i>	43
<i>Cuadro No. 9. Valores de Alfa y beta para máxima precisión</i>	58
<i>Figura No. 17. Comportamiento de la precisión para diferentes valores de alfa y beta par</i>	58
<i>Cuadro No. 10. Resultados para diferentes Modelos de Parsing no supervisado.</i>	59



# 1. INTRODUCCIÓN

Dado que gran cúmulo de información digital se encuentra almacenada en lenguaje natural y que la gran necesidad de los usuarios de esta información es gestionarla (almacenarla, consultarla y actualizarla) de forma automática, los medios computacionales han permitido que dicha gestión sea posible a través de artificios estructurales. Sin embargo, ésta presenta aún grandes dificultades, dada la imposibilidad de que los medios computacionales puedan <<entender>> el conocimiento inmerso dentro de una expresión en lenguaje natural tal como lo hace un ser humano. Dicha insuficiencia ha repercutido en la búsqueda de alternativas que minimicen aquellos errores que se ocasionan en los diferentes procesos que involucran de una u otra forma el lenguaje natural, tales como recuperación de información, traducción, procesos de respuesta-pregunta, extracción de información, identificación de patrones, identificación de identidades nombradas y en general procesos que ameritan la comunicación entre hombre-máquina a través de lenguaje escrito.

Dentro de las causas del por qué de dicha imposibilidad, podemos enumerar, entre otras: a) que el computador solo puede procesar formalismos estructurados (llámese lenguaje estructurado, estructura conceptual o estructura matemática) que no pueden almacenar toda la semántica propia del lenguaje natural; b) se debe traducir del lenguaje natural a dicho formalismo; c) se debe escoger el formalismo más adecuado; d) el formalismo debe permitir realizar sobre él operaciones propias de la matemática como comparación, unión, entre otras; e) el usuario del común debe poder entender la respuesta del computador, es decir que se espera que el contenido del formalismo estructurado permita ser traducido de nuevo al lenguaje natural.

Además de las anteriores causas, no se puede dejar de lado que en cualquier formalismo escogido, se debe recurrir, generalmente, a ayudas (aplicativos) que permitan interpretar aspectos semánticos o sintácticos del lenguaje natural que dependen del contexto. Dichas ayudas incluyen desde procedimientos manuales a través de expertos (en lingüística o en un dominio específico) hasta sistemas de software complejos como las gramáticas, ontologías, tesauros, diccionarios. Es así que, las ayudas no solo aumentan la complejidad, el tiempo en procesamiento sino la pérdida de información (al descartar conocimiento relevante propio del lenguaje natural) tanto en los procesos principales como en los procesos necesarios para construir las ayudas.

Una gran variedad de formalismos estructurados han sido propuestos para representar el lenguaje natural, que van desde las representaciones más simples como los espacios vectoriales hasta las más complejas como los lenguajes estructurados, estructuras conceptuales y formalismos matemáticos como los grafos. Los formalismos más simples, en términos de que pueden capturar menos semántica, tienen la ventaja de que su procesamiento, incluida la traducción, es mucho más simple. Mientras que los formalismos más complejos, si bien incluyen más contenido semántico, el procesamiento (la traducción de lenguaje natural al formalismo y su proceso inverso así como todos los aspectos de la gestión) es mucho más complejo.

Los grafos conceptuales es uno de los formalismos que además de lograr incluir mayor cantidad de semántica, ya que permiten expresar diferentes tipos de conceptos y relaciones, cuenta con algunas características propias de la lógica matemática, de lo que se derivan las ventajas propias del uso del instrumental matemático. Es por lo que este trabajo involucra los grafos conceptuales, y pese a que no se va a ocupar de la instrumentación matemática propiamente dicha, sino de cómo transformar el lenguaje natural en dicho formalismo, sus resultados se orientan a facilitar y profundizar en el uso de este formalismo.

No son muchas las investigaciones que involucran el problema de cómo construir los grafos conceptuales a partir del lenguaje natural, por lo que en este trabajo se presenta una propuesta para dicha transformación.

## 1.1 HIPÓTESIS

- ✓ Dado que los GCs conservan muchos elementos de la estructura sintáctica y semántica, una gramática para la construcción de GCs puede incluir tanto la estructura sintáctica y semántica de los GCs.
- ✓ La anterior gramática se puede obtener automáticamente a partir de recursos léxicos desarrollados con otros fines.
- ✓ Los GCs generados por la anterior gramática son lo suficientemente buenos para mejorar alguna tarea práctica, como la recuperación de información.
- ✓ Se puede proponer una estructura de GCs simplificados que capture importantes aspectos semánticos de los GCs.
- ✓ Esta estructura se puede construir directamente, sin necesidad de construir primero la estructura sintáctica y utilizando un método no supervisado y pobre en conocimiento (corpus de las estadísticas de co-ocurrencia).



## 1.2 ESTRUCTURA DEL DOCUMENTO

Dado que la investigación se enfoca a la transformación automática de los Grafos conceptuales y que para su desarrollo se han considerado a los grafos como una estructura computacional que incluye aspectos semánticos y sintácticos del lenguaje natural, idioma inglés, el desarrollo del trabajo se ha enfocado en dos aspectos fundamentales, por un lado a la construcción de los grafos conceptuales vía un modelo de reglas sintácticas y semánticas que confluyen en una gramática y por el otro a la construcción de los grafos a partir de un enfoque puramente semántico con la utilización de un método no supervisado. El primer aspecto se focaliza en la creación de los Grafos conceptuales a través de reglas gramaticales. Como resultado de la investigación se ha obtenido una gramática que incluye tanto los aspectos estructurales y sintácticos, como los semánticos propios de los grafos conceptuales. El segundo caso se concentra en la creación de un analizador sintáctico a partir de un método no supervisado y pobre en conocimiento basado en estadísticas. Teniendo en cuenta estos dos aspectos el trabajo se ha estructurado de la siguiente forma: El capítulo 2, incluye el marco teórico y el estado del arte, donde se ubican a los grafos conceptuales y se presenta el estado de la investigación en la recuperación de los grafos conceptuales. El capítulo 3, presenta la construcción de grafos conceptuales de acuerdo a un método basado en reglas. Las reglas incluyen elementos estructurales, sintácticos y semánticos. El capítulo 4, incluye la construcción de los grafos a partir de un analizador sintáctico - semántico ("*parsing*"), basado en un método no supervisado y pobre en conocimiento. En el capítulo 5, se incluyen los resultados de la investigación, se presentan los aportes tanto teóricos como técnicos, las conclusiones, el trabajo futuro y las publicaciones resultado de la investigación.

## 2. ESTADO DEL ARTE Y MARCO TEÓRICO

La mayor parte de la información que posee la humanidad se encuentra almacenada en forma de lenguaje natural. La gran necesidad de los usuarios de esta información es gestionarla: almacenarla, consultarla, entenderla y actualizarla. Los medios computacionales han permitido que dicha gestión sea posible a través de artificios estructurales. Sin embargo, ésta presenta aún grandes dificultades, dada la imposibilidad de que los medios computacionales puedan «entender» de manera directa y total el conocimiento inmerso dentro de una expresión en lenguaje natural tal como lo hace un ser humano. Un texto expresado en lenguaje natural generalmente es entendible para el ser humano pero para una máquina computacional no es tan obvio como se quisiera.

Para el ser humano una frase expresada en lenguaje natural leída, escuchada, hablada o escrita adquiere inmediatamente un significado. Cada palabra o conjunto de palabras, el ser humano las asocia con un conjunto de experiencias e imágenes que al ser relacionadas y filtradas trae a la mente exactamente el significado correcto. Para una máquina computacional, un conjunto de palabras corresponde a una cadena de bits (símbolos) sin significado alguno. Para que una máquina pueda «entender» su significado, ésta debe ser programada para cumplir una serie de tareas complejas que incluyen la representación en una estructura computacional.

El procesamiento del lenguaje natural (PLN), es una disciplina que con el apoyo de la lingüística computacional se encarga de proveer soluciones para la interpretación y gestión del lenguaje natural. Dichas disciplinas se soportan en técnicas y métodos propios de la estadística, las matemáticas, la lingüística, la inteligencia artificial, entre otras. Lo correspondiente a la generación de herramientas y aplicativos que incluyen la gestión del lenguaje natural se puede enmarcar dentro de lo que se conoce como ingeniería lingüística.

En la actualidad ya no es solo el deseo de que las máquinas computacionales puedan establecer una comunicación utilizando el lenguaje natural, tal y como lo hacen las personas sino que se ha convertido en una exigencia. Responder a un sinnúmero de necesidades actuales como la utilización de la información almacenada a través de la Web o de aplicativos ya sea para buscar información específica, patrones de comportamiento o predicciones ha hecho que cada vez se busquen y optimicen alternativas para que de alguna manera las máquinas «entiendan» el contenido de los documentos digitales expresados en

lenguaje natural. Consecuentemente, entre las aplicaciones más comunes en las áreas en mención son la recuperación de información, la respuesta automática a preguntas, la traducción automática y la clasificación de documentos.

En este capítulo se presenta una revisión del primer aspecto mencionado, es decir de las estructuras computacionales utilizadas para el procesamiento del lenguaje natural y en se profundiza en los Grafos Conceptuales, resaltando algunas ventajas a la hora de realizar tareas de procesamiento de lenguaje natural o de creación de ontologías, tesauros o bases de conocimiento.

## **2.1 ESTRUCTURAS COMPUTACIONALES EN EL PLN**

Una gran variedad de formalismos estructurados han sido propuestos para representar los textos en el lenguaje natural. Estos formalismos van desde las representaciones simples como las vectoriales hasta más complejas como los lenguajes estructurados, estructuras conceptuales y formalismos matemáticos como los grafos. Los formalismos más simples, en el sentido de que pueden capturar menos cantidad de elementos que conduzcan a la interpretación del significado o semántica, tienen la ventaja de que su procesamiento, incluida la transformación de la cadena de caracteres a dicha estructura, es mucho más simple. Mientras que los formalismos más complejos, si bien permiten incluir más elementos que pueden contribuir a la interpretación del significado inmerso dentro de la cadena, el procesamiento —la transformación de lenguaje natural al formalismo y viceversa— es mucho más complejo y, lo más importante, mucho menos confiable con la tecnología actual.

En las siguientes secciones se presenta una revisión de las estructuras computacionales, empezando por las más básicas y continuando con más avanzadas.

### **2.1.1 ESTRUCTURAS BÁSICAS**

Dentro de esta categoría se clasifican aquellas estructuras que se usan cuando no se requiere mayor nivel de detalle. En ocasiones estas estructuras también son la base para las estructuras más avanzadas.

#### *2.1.1.1 Modelo de espacio vectorial*

Teniendo en cuenta el conjunto de palabras que hacen parte de una sentencia, una representación muy clásica es la propuesta por Salton [81], autor del trabajo pionero en representar los documentos como un vector de frecuencias, donde cada entrada (coordenada) del vector corresponde a una determinada palabra dentro del documento y su valor es la frecuencia de aparición de dicha palabra. Se puede decir que la mayoría de trabajos para la recuperación de información y clasificación utilizan dicha representación, como [22, 35, 44].

Trabajar con las palabras que pertenecen a un lenguaje conlleva al manejo de un gran número de variables y por ende el procesamiento de alta dimensionalidad. El procesar el lenguaje a través de alguna estructura no elimina, por sí solo, la problemática de la dimensionalidad, por lo que aparecen trabajos que estudian el comportamiento de las palabras con análisis estadístico [58] y los que permiten asimilar las palabras a sus raíces (método llamado stemming en inglés) como el presentado en [6] o el algoritmo propuesto por Porter, llamado Porter stemmer en inglés [78].

Otra representación basada también en vectores, es aquella donde cada entrada del vector significa la presencia o ausencia de una palabra en el documento, es decir que el documento se representa como un vector de entradas binarias [46].

La representación de documentos con la estructura vectorial, puede darse a través de tuplas como un conjunto de parejas (término, peso), donde el peso corresponde a un valor dado de acuerdo a la parte del documento donde aparece la palabra —por ejemplo, un título tendrá mayor peso que un subtítulo.

La colección de palabras puede ser extractada de todo el documento o de alguna parte específica: por ejemplo, en [98] las palabras se extraen únicamente de los enlaces para así cumplir con un trabajo de clasificación.

Con esta representación se pueden utilizar diferentes técnicas. Por ejemplo, en los trabajos de recuperación de información y de indexación se utilizan métodos de semántica latente que buscan visualizar las relaciones ocultas que existen entre las palabras a través de la aplicación de vectores y valores propios [17]. Estas técnicas se aplican sobre vectores de frecuencia o sobre vectores cuyo contenido corresponde a valores binarios [36].

#### *2.1.1.2 Listas*

Las colecciones de palabras representadas con estructuras básicas no solo se han tenido en cuenta de forma vectorial sino también como listas. Dentro de esta clasificación se pueden referenciar las estructuras presentadas en [79], donde cada documento se representa por medio de una lista de clases, con el fin de indexar los documentos.

#### *2.1.1.3 Grafos*

Los grafos, y específicamente los árboles, son estructuras matemáticas que permiten representar, entre otras cosas, un documento.

Existen varios métodos para representar los documentos como grafos. En [83], se clasifican los métodos en: estándares, simples, distancia  $n$ , distancia  $n$ simple, frecuencia absoluta y frecuencia relativa. Cada método se basa en examinar los términos en cada documento y sus adyacencias. Al igual que en otros métodos, los términos se extraen del documento y se

realiza un preprocesamiento, que generalmente consiste en eliminar las palabras de más alta frecuencia (palabras basura, o stopwords en inglés) y así disminuir la dimensionalidad.

A continuación brevemente se expondrá en qué consiste cada una de las representaciones.

Estándar. Cada palabra corresponde a la etiqueta de un nodo y si una palabra *a* inmediatamente precede a una palabra *b* en una sección *s*, entonces existe una arista que comienza en *a* y termina en *b* etiquetada con *s*. En este caso se tiene en cuenta la puntuación y las secciones propias de un documento como el título o resumen, entre otros.

Simple. A diferencia de la estándar, no se etiquetan las aristas con las secciones y no se tienen en cuenta todas las secciones sino aquellas que sean más visibles.

Distancia *n*. Esta representación depende de un parámetro *n*, que lo provee el usuario. Teniendo en cuenta dicho parámetro, se buscan las *n* palabras siguientes a partir de un término dado, y las aristas se etiquetan con la distancia al punto inicial.

Distancia simple. Es similar a la anterior, con la diferencia de que las aristas no son etiquetadas y lo único que se sabe es que la distancia entre dos nodos conectados es menor que *n*.

Frecuencia absoluta. Es similar a la representación simple, pero cada nodo y arista son etiquetados con una medida de frecuencia. Para un nodo esto indica cuántas veces los términos aparecen en el documento; para las aristas esto significa el número de veces que dos términos conectados aparecen en el orden específico. Bajo esta representación, el tamaño del grafo se define como la suma de las frecuencias de los nodos más la suma de las frecuencias de las aristas.

Frecuencia relativa. Es similar a la frecuencia absoluta con la diferencia de que las frecuencias son normalizadas entre 0 y 1. Los nodos se normalizan por el valor máximo de frecuencia de los nodos y las aristas similarmente por el máximo valor de frecuencia en las aristas.

Varias propuestas utilizan la representación con grafos. Por ejemplo, en [6] se representa un documento por un grafo que tiene en cuenta la frecuencia de ocurrencia de las palabras. En [82] se incluyen varias formas de representar los documentos a través de los grafos. Además, en [6] se propone una metodología para la construcción de grafos: a partir de análisis del dominio se presentan objetos o entidades que son agrupados dentro de clases o tipos.

#### *2.1.1.4 Estructuras estadísticas*

Con base a la teoría de la información se han realizado investigaciones sobre el comportamiento de las palabras y la información que más aporta a un documento.

El primer modelo estadístico del lenguaje fue propuesto por Claude Shannon [85]. De acuerdo a la teoría de la información, el lenguaje se modela como una fuente estadística de información. La estadística se ha constituido en una herramienta fundamental para el análisis de lenguaje natural. De igual forma se han propuesto estructuras que incluyen funciones probabilísticas para representar el contenido de un texto. Dentro de estas estructuras están, por ejemplo, los modelos de Markov, las gramáticas probabilísticas, los analizadores probabilísticos [55]. En general, cualquier estructura (un vector, un grafo o lista, entre otros) puede ser marcada con probabilidades o funciones de probabilidad.

En los modelos probabilísticos, se construyen distribuciones de los documentos que pertenecen a una clase. En el caso de recuperación de información, por ejemplo, se asume una distribución para los documentos relevantes y otra para los no relevantes. Se establecen suposiciones sobre las distribuciones, como independencia, y se aplican procedimientos como Naïve Bayes [55]. En los trabajos como [70], la representación de los documentos se logra a través de funciones probabilísticas, y se asume que un documento fue producido por la generación de una función de densidad. En esta investigación, se utiliza un método de agrupamiento (clustering) discriminativo distribucional para extraer las características relevantes de los documentos y así representar los documentos como una distribución de probabilidad.

En otros trabajos se mezclan algunas de las técnicas previamente ya expuestas con modelos del lenguaje. En [54], por ejemplo, se presenta un modelo probabilístico para representar el grafo de un documento, con la esperanza de que en el marco de la recuperación de información, un modelo de grafo pueda generar o estimar un grafo de la consulta.

Los modelos estadísticos diferentes a los modelos probabilísticos incluyen las dependencias directas que se presentan por la proximidad o adyacencia entre palabras como en los bi-gramas y en general los n-gramas [54]. En el trabajo referenciado en [101] utilizan una representación basada en palabras y n-gramas.

## 2.1.2 ESTRUCTURAS AVANZADAS

Dentro de esta categoría se encuentran las representaciones a través de las gramáticas, a través de la lógica, así como los lenguajes orientados a objetos y aquellos usados para la creación de repositorios de datos como el Lenguaje de Definición de Datos (DDL, por sus siglas en inglés: *Data Definition Language*).

### 2.1.2.1 Representación Sintáctica

Una forma que se podría utilizar para representar un documento es a través de la descripción morfológica y sintáctica. La descripción morfológica establece la clase gramatical de cada una de las palabras del texto. Dicha clase se define de acuerdo a la función gramatical en el proceso que se denomina etiquetamiento de categorías gramaticales (*part of speech tagging en inglés*).

El análisis sintáctico se encarga de analizar la sentencia teniendo en cuenta tanto la función que ejerce cada palabra así como las relaciones existentes entre las mismas. Para dicho análisis se construyen estructuras sintácticas mediante las técnicas de constituyentes, de dependencias o de enlaces, entre otras.

La técnica de análisis de constituyentes consiste en construir un árbol mediante un proceso iterativo de segmentación y clasificación gramatical de la oración, hasta que las partes constituyentes sean indivisibles.

Desde el punto de vista de la teoría de las dependencias [97] se puede construir una jerarquía —esquemática en un árbol— de acuerdo al papel que ejercen las palabras dentro de la frase como la encabezada o raíz del árbol, las subordinadas y rectoras [12]. Las relaciones se marcan con flechas y varias palabras pueden depender de una sola, aunque cada palabra excepto la raíz depende de exactamente una otra palabra.

Las gramáticas de enlace (*link grammars*) introducidas en [89] se definen como un conjunto de palabras que requieren de un enlace. Una sucesión de palabras equivale a una frase del lenguaje natural, si existe una forma de dibujar los enlaces entre las palabras que la conforman. Dichos enlaces no se cruzan y satisfacen los requerimientos de cada palabra.

Las gramáticas de adjunción de árboles [3] son una extensión de las gramáticas formales e independientes del contexto que permiten la recursión. Las reglas son definidas no reescribir símbolos sino para reescribir árboles. La derivación tiene su origen en un árbol inicial que se combina con otros mediante las operaciones de sustitución y adjunción.

#### 2.1.2.2 Lógica de primer orden y otros métodos basados en lógica

En la lógica de primer orden (FOL por sus siglas en inglés: *First Order Logic*), con pocos símbolos básicos se pueden representar elementos del mundo real y a través de predicados se pueden establecer las relaciones entre los elementos. Es así que usando elementos, proposiciones y operadores simples se puede representar el texto [9].

La lógica del primer orden es un sistema deductivo formal usado en las matemáticas, filosofía, lingüística e informática. Se conoce también como cálculo de predicados de primer orden (FOPC por sus siglas en inglés: *First Order Predicate Calculus*), el más bajo cálculo de predicados, el lenguaje lógico de primer orden o lógica de predicados.

Con esta técnica se pueden representar texto y por ende documentos. Por ejemplo, en [15] un documento se presenta como una sentencia lógica proposicional de la forma:

$$d = (\text{recuperación} \wedge \text{información}) \vee (\text{recuperación} \wedge \text{datos}).$$

Siguiendo la representación a través de la lógica, se encuentran investigaciones como [13], donde a través de la lógica difusa se representa el documento para implementar posteriormente una forma de recuperación de documentos.

### 2.1.2.3 Lenguaje de Representación de Marcos

Los lenguajes de tipo FRL (por sus siglas en inglés: *Frame Representation Language*) clasificados bajo esta categoría se definen como metalenguajes basados en el concepto de marco (*frame*), orientado al reconocimiento y descripción de objetos y clases. La idea principal del razonamiento basado en estos lenguajes es la simplificación, ya que la unidad primaria de organización es el marco. Un marco tiene un nombre y puede tener varios atributos. Cada atributo tiene a su vez un nombre y un valor. Los diferentes tipos de valores pueden ser de una amplia variedad de acuerdo al sistema de marcos. Los valores más comunes son las cadenas y los símbolos [42].

En los FRLs, la herencia es el mecanismo de inferencia central basado en la organización jerárquica. Muchos sistemas que se basan en este lenguaje, permiten la herencia múltiple, como los lenguajes de programación y en particular los orientados a objetos. Los FRLs no se catalogan como ontologías, ya que solo se ocupan de representar el conocimiento como objetos, contrario a las ontologías que buscan representar el conocimiento en detalle [9]. Sin embargo, en algunas implementaciones y lenguajes basados en FRLs, se asumen como un lenguaje ontológico.

Dentro de las implementaciones basadas en FRLs destacan los siguientes:

Protégé [69], es una plataforma de desarrollo de ontologías definida en marcos y bajo el estándar del lenguaje ontológico para la Web (OWL: *Web Ontology Language*), desarrollada en Java. Las ontologías definidas pueden ser exportadas en una amplia variedad de formatos.

Capa de inferencia ontológica. La Capa de Inferencia Ontológica o lenguaje de intercambio ontológico (OIL: *Ontology Inference Layer o Ontology Interchange Language*) [62] se define como una estructura ontológica para la semántica en la Web.

Lógica de marcos. La lógica de marcos (*Frame logic o F-logic*) [45] es un lenguaje ontológico que permite representar el conocimiento. Su estructura está basada tanto en el FRL como en la orientación a objetos.

Sistema para representación de conocimiento. El sistema para representación de conocimiento (*Knowledge Representation One o KL-ONE*) [10] es un lenguaje muy similar al mismo FRL (aunque los marcos en este lenguaje se llaman conceptos) con la diferencia de que se permite el manejo de subclase, superclase y múltiple herencia. Es muy utilizado para representar las redes semánticas.

Lógica de descripciones. La lógica de descripciones (*DL: Descriptions Logics*) [5] corresponde a una familia de lenguajes definidos para representar el conocimiento. Su nombre se refiere por un lado a la descripción de conceptos para describir dominios y por otro a la semántica basada en lógica de predicados de primer orden. Las DLs fueron diseñadas como una extensión de los FRLs y las redes semánticas con el fin de fortalecer la parte semántica formal.



La lógica descriptiva no solamente representa el conocimiento como objetos sino justifica el proceso con un razonamiento estrictamente lógico basado en conceptos, roles y relaciones. Los conceptos corresponden a clases de elementos y son tratados como subconjuntos del universo. Las relaciones corresponden a vínculos entre elementos y son tratados como relaciones binarias.

### 2.1.3 ESTRUCTURAS CONCEPTUALES ESTANDARIZADAS

Las estructuras conceptuales para la representación de conocimiento son modelos (artefactos) que representan una realidad percibida. Con base en técnicas como la inteligencia artificial, estos modelos además de representar el conocimiento, pueden lograr reconstruir conocimiento a través de métodos como la inferencia [66].

Dentro de esta clasificación se encuentran las redes semánticas, los grafos conceptuales, el formato de intercambio de conocimiento (*KIF: Knowledge Interchange Format*) [30], el Lenguaje de Descripción de Recursos (*RDF: Resource Description Framework*) [51] del consorcio “*World Wide Web Consortium (W3C)*” y los diferentes lenguajes ontológicos para la Web (*OWL: Web Ontology Language*) propuestos por el W3C [18].

Otra estructura conceptual que se puede mencionar, es la Lógica Común (*CL: Common Logic*) que no hace parte del grupo W3C pero se posiciona al lado de RDF y de OWL. CL es un marco (“*framework*”) para una familia de lenguajes basados en lógica. La lógica de primer orden tiene como objetivo el intercambio y la transmisión de información. El objetivo de este framework es proveer la sintaxis y semántica abstracta de las sintácticas concretas o dialectos. El CL preserva el modelo teórico de primer orden, pero cuenta con características muy particulares como la de permitir construcciones de orden superior tales como cuantificaciones sobre clases o relaciones.

#### 2.1.3.1 Redes Semánticas

Las redes semánticas aparecen a partir de trabajos lingüísticos presentados en el año 1968, y de ahí en adelante los diferentes aportes llevaron a que a finales de los años 70, se puedan observar dos tendencias: por un lado, las redes estructuradas y los sistemas de representación del conocimiento y por el otro, las multiredes orientadas a las ciencias cognitivas.

De forma general se puede definir una red semántica como un modelo matemático que consiste de una estructura conceptual formada por un conjunto de conceptos y relaciones cognitivas entre éstos. Son representadas por un grafo generalizado donde los conceptos corresponden a los nodos y las relaciones entre los conceptos corresponden a los arcos [34] y, desde el punto de vista semántico, como una colección de las diferentes relaciones que los conceptos tienen entre sí [96]. Generalmente, la construcción de una frase se logra con la ayuda de los analizadores sintácticos. Sin embargo, esto es aún un problema para aquellos idiomas que no cuentan con un orden estricto de palabras, como el español [27]. Las redes semánticas se catalogan en tres categorías [39]:

Red de relaciones es-un. Se construye desde los conceptos más generalizados hasta más específicos, que se representan de forma jerárquica. Dado que la filosofía de estas redes gira en torno a la herencia y la explicación de los conceptos mediante otros conceptos, son generalmente complejas. Entre las investigaciones que utilicen una red semántica del tipo red de relaciones es-un (is-a en inglés) se puede mencionar, por ejemplo, [52], donde se construye una red en forma de árbol; sus atributos y palabras se seleccionan a través del método de entropía condicional.

Red de marcos. Este tipo de red se representa a través de estructuras de datos llamadas marcos. Cada marco se corresponde a una clase o a una instancia. Las clases describen los conceptos mediante un conjunto de propiedades y los marcos son las instancias de las clases. Ellos describen objetos concretos y heredan propiedades de los marcos clase. Los atributos y valores se esquematizan a través de ranuras (*slots*).

Grafos conceptuales. Durante los años 60 la representación semántica basada en grafos fue popular tanto a nivel teórica como en la lingüística computacional. Esta representación se conoce como grafo conceptual. Esta estructura, propuesta por Sowa [96] está basada en los grafos existenciales de Pierce [87].

El estándar para los grafos conceptuales especifica la sintaxis, la semántica y la representación de cadenas de caracteres en el formato de intercambio de grafos conceptuales (CGIF: *Conceptual graph interchange form*). CGIF ha sido diseñado para intercambio de información entre los sistemas que hacen parte del estándar «Formato para el modelamiento de esquemas conceptuales» (CSMF: *Conceptual Schema Modeling Facilities*) especificado por el estándar ISO/IEC 14481. El estándar de los GC provee una guía para implementar sistemas que usan los grafos conceptuales en la representación interna o externa. Las representaciones externas se definen para la comunicación humano—máquina, y la interna, para la comunicación entre las máquinas [91].

Los grafos conceptuales para representar texto fueron introducidos por Sowa [91] Los grafos conceptuales manejan dos tipos de nodos: conceptos y relaciones conceptuales. Los conceptos tienen un tipo (clase de concepto) y un referente (la instancia de este tipo de objeto). Las relaciones conceptuales señalan la manera en que los conceptos se relacionan [68]. Cada relación conceptual tiene uno o más (usualmente dos) arcos, cada uno de los cuales debe estar enlazado a un concepto [91].

Dado que la representación por medio de un grafo conceptual denota los términos que contribuyen a la semántica de la sentencia y que cada término se escoge de acuerdo a la posición dentro de la sentencia [86], los grafos conceptuales cuentan con una serie de características que hacen que sean muy ricos semánticamente y se utilicen no solo para el intercambio de información sino para la creación de bases de conocimiento y ontologías.

### 2.1.3.2 *Formato de Intercambio de Conocimiento*

El Formato de Intercambio de Conocimiento (KIF: *Knowledge Interchange Format*) se basa en caracteres que pueden ser combinados en lexemas; los lexemas a su vez pueden ser combinados en expresiones. La sintaxis del KIF [30] generalmente se presenta con una modificación de la notación de las formas BNF (*Backus-Naur forms*).

El alfabeto de KIF consiste de seis bloques de datos para referenciar las mayúsculas, las minúsculas, los dígitos, los caracteres alfa —caracteres especiales que se usan de la misma forma que las letras—, los caracteres especiales y otros caracteres como el espacio.

El proceso de convertir los caracteres a lexemas se llama análisis léxico. Al proceso entra una cadena de caracteres y se obtiene una cadena de lexemas. Este proceso de análisis léxico es cíclico. En este proceso se leen las cadenas de caracteres hasta que se encuentra un carácter que no puede ser combinado con los caracteres previos y dentro del lexema actual. Cuando esto ocurre, el proceso se vuelve a empezar con el nuevo carácter y otro lexema.

KIF maneja cinco clases de lexemas: lexemas especiales, palabras, referencias al carácter, cadenas de caracteres y bloques de caracteres.

Los lexemas se forman de acuerdo a una serie de reglas. Se presentan tres tipos de expresiones: términos, sentencias y definiciones. Los términos son usados para denotar objetos, las sentencias para expresar hechos y las definiciones para definir constantes. Las definiciones y las sentencias se llaman formas; una base de conocimiento es un conjunto finito de formas.

La base de la semántica de KIF es la conceptualización del mundo en términos de objetos y relaciones entre los objetos. El universo del discurso es el conjunto de todos los objetos que hipotéticamente existen en el mundo. La noción de objeto es amplia. Los objetos pueden ser concretos o abstractos, primitivos o compuestos y pueden ser de ficción.

Además de permitir incluir listas, el lenguaje permite incluir sentencias matemáticas, de control, de relaciones y lógicas.

### 2.1.3.3 *Infraestructura para la Descripción de Recursos*

El lenguaje RDF (*Resource Description Frame-work*) [51] ha sido definido para representar información sobre recursos en la Web. En particular, intenta representar metadatos sobre los recursos de la Web como el título, el autor, fechas, derechos y en general cualquier información relevante. Por otro lado, proporciona interoperabilidad entre las diferentes aplicaciones que intercambian información en la Web. Su desarrollo se ha basado no solo en la necesidad de la Web sino en los demás estándares que definen las diferentes comunidades, tales como los presentados a continuación.

Lenguaje de metadatos para publicar hipertexto en Internet. El lenguaje HTML (*HyperText Markup Language*) es estandarizado por el grupo W3C y es el más popular para escribir las páginas web. Permite describir la estructura y el contenido en forma de texto, incluir imágenes, tablas, vínculos y muchos otros aspectos de presentación y diseño.

Plataforma para la selección de contenido en Internet. Esta especificación (PICS: *Platform for Internet Content Selection*) [49] habilita los metadatos que pueden ser asociados con el contenido de Internet. PICS fue inicialmente diseñada para ayudar al control de contenido que acceden los menores de edad en la Web; sin embargo actualmente es ampliamente utilizada en los filtros. En general este estándar permite etiquetar el contenido propio o relacionado, creando así el principal parámetro de control.

Lenguaje de marcado generalizado. El lenguaje SGML (*Standard Generalized Markup Language*) se define como un sistema para la organización y etiquetado de documentos. Al igual que el HTML, fue normalizado por la Organización Internacional de Estándares (ISO) en 1986.

Lenguaje de marcado extensible. El Lenguaje de Marcado Extensible (XML: *Extensible Markup Language*) es un metalenguaje extensible de etiquetas desarrollado por el W3C. Es una simplificación y adaptación del SGML. Permite definir la gramática de lenguajes específicos, como HTML. En general XML no es un lenguaje en particular, sino una manera de definir lenguajes para diferentes necesidades. XML estandariza el intercambio de información estructurado entre las diferentes plataformas computacionales.

#### 2.1.3.4 Lenguaje ontológico

El Lenguaje de Ontologías para la Web (OWL: *Web Ontology Language*) [52] es un lenguaje de marcado desarrollado por el grupo W3C para publicar y compartir ontologías en la Web. Fue desarrollado como una extensión del RDF y de DAML + OIL [62], incluyendo capacidad de distribución, escalabilidad y accesibilidad.

Una ontología OWL es un grafo RDF que permite expresar relaciones complejas entre las diferentes clases de RDFs. Provee los recursos para determinar propiedades y elementos y para la construcción de nuevas clases a partir de una u otras.

#### 2.1.3.5 Lógica común

Lógica Común (CL: *Common Logic*) es una estructura definida para una familia de lenguajes lógicos basados en lógica de primer orden. Define estándares para el intercambio de información basados en formas sintácticas llamadas dialectos. Un dialecto puede usar cualquier sintaxis que conforme una semántica abstracta CL. Todos los dialectos son equivalentes, es decir que pueden automáticamente ser traducidos entre ellos aunque algunos

pueden ser más expresivos que otros, en cuyo caso se pueden traducir solo a menos expresivos. El estándar ISO 24707 para la Lógica Común especifica tres tipos de dialectos:

- ✓ Formato de Intercambio de Lógica Común (CLIF: *Common Logic Interchange Format*),
- ✓ Formato de Intercambio de Grafos Conceptuales (CGIF: *Conceptual Graph Interchange Format*),
- ✓ Notación basada en XML para la Lógica Común (XCL: *XML Based notation for Common Logic*).

Son muchos los lenguajes que hereden de una sintaxis abstracta de la CL, entre los cuales destaca el que se presenta en [33].

Lenguajes naturales controlados. Los lenguajes naturales controlados son subconjuntos de los lenguajes naturales restringidos en la gramática y el vocabulario con el fin de reducir o eliminar la ambigüedad y la complejidad [7]. Los lenguajes controlados pueden ser desarrollados con dos objetivos: aquellos que mejoran la legibilidad para los lectores humanos y aquellos que permiten el análisis semántico automático confiable del texto. Dentro de estos lenguajes están el inglés, el chino y el español controlados [33].

Diagramas FLIPP. Son diagramas para representar la lógica. Cada diagrama consiste de un bloque dividido en bloques más pequeños. El diagrama total represente un grafo a-cíclico. Cada sub-bloque puede contener información en lenguaje declarativo, natural o matemático.

Mapas de tópicos y mapas conceptuales. Los mapas conceptuales son artefactos para la organización y representación del conocimiento. Tienen su origen en teorías sobre psicología [4]. El objetivo de estos mapas es de representar relaciones entre conceptos en forma de proposiciones. Los conceptos están incluidos en cajas o círculos, mientras que las relaciones entre ellos se explicitan mediante líneas que unen las cajas respectivas. Las líneas, a su vez, tienen palabras asociadas que describen la naturaleza de la relación que liga los conceptos.

Lenguaje de modelamiento unificado. Es un lenguaje de modelado de software (UML: *Unified Modeling Language*). Su estándar es definido por el grupo *Object Management Group* (OMG). El estándar define el modelamiento estático y dinámico de todos los componentes que pueden hacer parte de un sistema de software.

Otros lenguajes. El lenguaje de consulta estructurado (SQL: *Structured Query Language*) es un lenguaje declarativo que permite recuperar información estructurada de las bases de datos relacionales. El lenguaje de restricción de objetos (OCL: *Object Constraint Language*) ha sido definido para describir las reglas que aplican al UML. Prolog es un lenguaje de programación lógico e interpretativo. Datalog, una derivación del Prolog, es un lenguaje de consulta para bases de datos deductivas. Esquema RDF, una extensión semántica de RDF, es un lenguaje primitivo para la descripción de vocabulario ontológico. Finalmente, el ya

mencionado Lenguaje de Ontologías para la Web (OWL: *Web Ontology Language*) es un lenguaje de marcado para gestionar y compartir datos en la Web.

## 2.2 GRAFOS CONCEPTUALES (GC)

Durante los años 60 la representación semántica basada en grafos fue popular tanto a nivel teórica como en la lingüística computacional, y es así que aparecen las redes semánticas [34] redes de correlaciones, grafos de dependencias y los grafos conceptuales con Sowa [96] basados en los grafos existenciales de Pierce [87].

Un grafo conceptual (GC) es una representación gráfica para la lógica basada en redes semánticas y grafos existenciales. Varias versiones de los GC se han implementado en los últimos 30 años y a partir de la investigación de los GC, se han explorado nuevas técnicas de razonamiento, representación de conocimiento y semántica del lenguaje natural.

El estándar “Formato para el modelamiento de esquemas conceptuales” (ISO/IEC 14481) [40] especifica la sintaxis, la semántica y la representación de cadenas de caracteres para el intercambio de grafos conceptuales (Conceptual graph interchange form C([50])), a nivel interno y externo. Las representaciones externas se definen para la comunicación humano máquina y la interna para la comunicación máquina [91].

### 2.2.1 CARACTERÍSTICAS DE LOS GC

Los grafos conceptuales para representar texto son introducidos por Sowa [93] y a diferencia de los grafos tradicionales, los grafos conceptuales manejan dos tipos de nodos, conceptos y relaciones conceptuales. Los conceptos tienen un tipo (clase de concepto) y un referente (la instancia de este tipo de objeto). Las relaciones conceptuales señalan la manera en que los conceptos se relacionan. La relación se establece a través del tipo relacional y la valencia. El tipo relacional indica el rol semántico que realizan los conceptos adyacentes a la relación y la valencia indica el número de estos [91]. Cada relación conceptual tiene uno o más arcos, cada uno de los cuales debe estar enlazado a un concepto.

Dado que la representación por medio de un grafo conceptual, denota los términos que contribuyen a la semántica de la sentencia y que cada término se escoge de acuerdo a la posición dentro de la sentencia [86], los grafos conceptuales cuentan con una serie de características que hacen que sean muy ricos semánticamente y se utilicen no solo para el intercambio de información sino para la creación de bases de conocimiento y ontologías. Dentro de las características más relevantes de los GC se pueden resaltar [65] y [92]: a) Los conceptos como las relaciones se pueden categorizar por tipos lo que da una posibilidad de combinaciones e interpretaciones; b) Los conceptos pueden ser especificados por referentes que a través de un cuantificador, un apuntador y un descriptor, determinan las conexiones entre el formalismo del GC y las entidades a que este se refiere; c) Los GC permiten involucrar el contexto a través de un concepto cuyo apuntador es un grafo conceptual; d) Los

GC permiten incluir una base de conocimiento cuyo apuntador es un GC consistente de cuatro conceptos: un tipo jerárquico, una relación jerárquica, un catalogo de individuos y un contexto externo.

## 2.2.2 HERRAMIENTAS PARA GENERAR GRAFOS CONCEPTUALES

Existen en el mercado gran variedad de herramientas que ayudan en la tarea de la creación de grafos conceptuales. A continuación se hace una pequeña descripción de algunas de ellas.

### 2.2.2.1 *Prolog+cg*

Prolog+cg fue desarrollada por A. Kabbaj [41] y nace como una implementación en Java del estándar Prolog, pero con extensiones para la construcción de grafos conceptuales bajo la teoría de Jhon Sowa. La herramienta permite utilizar conceptos de programación orientada a objetos, procesamiento de lenguaje natural y lenguaje lógico para proveer a través de los grafos conceptuales un lenguaje de representación de conocimiento que incluye entre otras características redes semánticas, manejo de jerarquías y operaciones de grafos conceptuales.

En las versiones iniciales de Prolog+CG, los grafos conceptuales simples y compuestos se tratan como estructuras básicas y primitivas tales como listas y términos. Posteriormente, Prolog+cg se integra dentro de la plataforma Amine, y en recientes versiones permite incluir los grafos conceptuales dentro de proyectos java. Como resultado de dicha integración, los programas de Prolog+CG son interpretados de acuerdo a una ontología que debe ser cargada primero. Los tipos jerárquicos y las estructuras conceptuales pueden buscarse directamente desde la ontología y cada identificador se puede buscar a partir de un léxico de la ontología. Se introduce la noción de proyecto que pueden compartir los mismos recursos, entre ellos la ontología. A través del lenguaje y los métodos propios de la herramienta se puede crear una ontología que responde a preguntas formuladas por el usuario.

### 2.2.2.2 *Amine*

Amine es una plataforma de código abierto que permite el desarrollo de diferentes tipos de sistemas inteligentes. Es el resultado de la integración de grafos conceptuales y sistemas inteligentes. Entendida la ontología como aquella que permite describir los conceptos y las relaciones entre los conceptos en un dominio particular. En los GC, una ontología consiste de una jerarquía que contiene tipos que representan grupos de entidades con similares rasgos [1]. Amine está basado en una estructura multinivel que ha progresado en el tiempo. En la actualidad cuenta con nueve capas. A través de la herramienta se pueden crear y editar ontologías. Las ontologías definidas como estructuras conceptuales, se le pueden asociar varios léxicos conceptuales (inglés, francés, español, árabe, etc.).

### 2.2.2.3 *Charger*

CharGer permite crear ontologías definidas como GC en forma gráfica y permite generar su forma lineal. Las instancias son creadas haciendo un enlace al formato de texto y son mostradas en una caja rectangular con el nombre del tipo de concepto. Dentro de las características con las que cuenta esta herramienta se pueden destacar [19]: a) Almacena grafos en diferentes formatos incluido XML; b) Copia y pega los grafos usando un editor interno; c) Los tipos y relaciones jerárquicas pueden ser editadas y almacenadas; d) Soporta los contextos y acceso a bases de datos a través de los actores; e) Soporta algunas operaciones entre grafos como empalme y uniones; f) Permite enlazar a tesauros como Wordnet [53] o glosarios genéricos para conceptos y tipos; g) Permite manejar parámetros de estilo y colores; h) Dentro de las limitaciones que se mencionan en el manual es la falta de mecanismos de validación, de copia y pegue.

### 2.2.2.4 *Get*

Esta herramienta (*Graphs Editor and Tools Get*) [100], está definida como una implementación parcial de las estructuras conceptuales de Sowa para sentencias en portugués. GET es un editor de grafos que permite mediante una colección de predicados Prolog implementar operaciones. Permite analizar y generar la notación lineal de los GCs como una notación extendida de la presentada por Sowa, especialmente en lo concerniente a tipos y campos en contextos anidados. Como característica especial de esta herramienta, admite incorporar una base de datos que puede hacer referencia a un tipo jerárquico, a un conjunto de grafos y a varios esquemas para un tipo de concepto.

### 2.2.2.5 *Cogitant*

Es una herramienta desarrollada en C++ y nace como extensión de la librería Conceptual Graphs Integrated Tools (CoGITO) desarrollada en 1994 [29]. Permite la construcción de bases del conocimiento u ontologías basadas en GC. Provee las clases necesarias para gestionar cada uno de los elementos que hacen parte de los GC, como los mismos grafos, conceptos, relaciones, reglas, restricciones y los métodos para las principales operaciones del modelo de grafos como la proyección. Admite construir los grafos en memoria, los grafos simples o anidados, tipos de conceptos y relaciones. La librería cuenta con los métodos para el acceso a la aplicación desde un aplicativo Java y XML.

### 2.2.2.6 *Cogui*

Es una herramienta gráfica libre, desarrollada en Java para la construcción de ontologías basadas en grafos conceptuales [29]. La construcción de los grafos conceptuales se logra a través de un árbol o de una interfaz gráfica. Permite incluir jerarquías de conceptos, relaciones, reglas y grafos. Cada uno de los elementos que hacen parte del grafo se les puede editar, incluir información adicional, borrar y arrastrar. Para el manejo de la herramienta se



crean proyectos y módulos. El editor permite navegar a través de la ontología y editar gráficamente su estructura y contenido. Esta herramienta es heredera de COGUITANT con un enfoque a procesos educativos.

#### 2.2.2.7 *Corese*

Corese (Conceptual Resource Search Engine), es un motor que habilita el procesamiento de instrucciones RDFS, OWL y RDF basado en GC. CORESE está desarrollado en Java y cuenta con un API que permite a los desarrolladores adicionar semántica a las aplicaciones [15]. La principal funcionalidad de la herramienta está dirigida a recuperar recursos Web anotados en (*Resource Description Framework* RDF en (W3 RDF/XML) [99]) RDFs, usando un lenguaje de consulta basado en (Query Language for RDF SPARQL) SPARQL y un motor de reglas de inferencia.

El API cuenta con los métodos necesarios para crear y gestionar ontologías, conceptos, propiedades, reglas, anotaciones, instancias, etiquetas, entre otros. Es así, que por ejemplo, A través de la unión de varios grafos, se puede inferir un solo grafo conceptual que permite responder las preguntas formuladas por el usuario. Las transformaciones se logran gracias a las similitudes que existen entre las páginas Web basadas en RDFs y los GCs.

#### 2.2.2.8 *Cpe*

Cpe (*Conceptual Programming Environment*). CP fue originalmente desarrollado como un aplicativo para el desarrollo de sistemas de conocimiento con base en grafos conceptuales y con la posibilidad de manejar información temporal y espacial. A partir de CP surge CPE, con un diseño más modular, flexible y con la posibilidad de interactuar con otras aplicaciones a través lenguajes como el C y el C++. Dentro de las características se pueden resaltar: a) Aplica conocimientos de estructuras de bases de datos a la forma de estructurar y analizar los grafos conceptuales; b) Divide la base de conocimiento en dos capas lógicas: el conocimiento del mundo y el conocimiento del prototipo y c) Permite realizar operaciones básicas de los GCs incluidas la proyección y la de máxima unión [71].

#### 2.2.2.9 *Notio*

Es un API desarrollado en Java que contienen un conjunto de primitivas para la construcción y manipulación de GCs y provee una plataforma para el desarrollo de herramientas y aplicaciones de GC. Es desarrollada en el 1999, por lo que se presenta como una de las primeras herramientas orientadas a objetos y con la posibilidad de contar con una capa de aplicación que interopera con las demás capas, entre las cuales está la de gestión de los GC [90].

#### 2.2.2.10 Webkb

Esta herramienta cuenta con una interfaz en línea que permite a través de sentencias especializadas crear o compartir una base de conocimiento. La base de conocimiento fue inicializada con el contenido de la base léxica WordNet [64], sin tener en cuenta la información con respecto a los verbos, adverbios y adjetivos. Se complementaron las categorías de mayor nivel del WordNet pasando a más de 100 tipos de conceptos y 140 tipos de relaciones básicas [61]. Incluye las ontologías como parte de una base de conocimiento, definida como una lista de categorías y declaraciones que dan significado formal a las categorías. Para la herramienta una categoría es una referencia a una clase de objeto (tipo de concepto), relación entre objetos (tipo de relación) u objeto particular (individuo). Cada categoría se asocia con un identificador único que permite distinguir una de otra categoría, y uno o varios nombres. Se maneja un nombre de categoría que puede tener varios significados y que puede estar asociado a varias categorías.

La herramienta integra una serie de lenguajes para su funcionalidad tales como: a) Lenguaje de comandos (*For Structuration FS*); b) Formato lineal de GC (Conceptual Graph Linear Format CGLF); c) Formato de intercambio de GC (*Conceptual Graph Interchange Format C* en [60]) y d) Lenguaje de descripción de recursos (*Resource Description Format RDF*).

Dentro de las características del WebKB-2 que se presenta en [61] y que se pueden destacar están; a) Es mucho más amplio y consistente que el WordNet; b) Es expandible en cualquier momento por cualquier usuario, c) Por estar en línea, permite cooperación asíncrona entre los usuarios ya que los usuarios pueden rehusar, complementar o corregir el conocimiento de otros usuarios pero no tienen que estar de acuerdo entre sí; d) La base de conocimiento es única y al ser compartida en línea, permite al máximo la interconexión de conocimiento, recuperación y descubrimiento de inconsistencia; e) Los lenguajes de representación de conocimiento utilizados orientan al usuario a expresar y leer conocimiento; f) Limita la forma de expresar el conocimiento estableciendo un lenguaje para tal fin; g) Permite la construcción compartida de páginas amarillas, como listas de catálogos; h) Puede recuperar documentos que incluye palabras claves y permite extraer conocimiento, es decir recuperar e interrelacionar información precisa.

#### 2.2.2.11 Ontoseek

Ontoseek (*Content-Based Access to the Web Ontoseek*) utiliza el WordNet y los grafos conceptuales existenciales para crear páginas amarillas o catalogo de productos. En el trabajo [32] demuestran que con el uso de la estructura propuesta, lenguaje de expresiones limitadas y la explotación de ontologías lingüísticas como el WordNet, se incrementa el índice de precisión y de recuperación. Permite realizar búsquedas clásicas, por ejemplo, por nombre en cambio de categorías.

#### 2.2.2.12 *Ontolingua*

Es una herramienta que soporta más de 150 usuarios activos en línea y en ambiente distribuido. Permite buscar, crear, revisar, modificar y usar ontologías [24]. Admite que aplicaciones se conecten remotamente al servidor a través de una interfaz especial. La página de la herramienta presente una fecha del año 95 y algunos de los enlaces están deshabilitados, por lo que se expondrá muy brevemente en qué consiste.

De acuerdo a la documentación, proporciona un conjunto de herramientas y librerías modulares que permiten la administración de las ontologías. Soporta la inclusión de grafos cíclicos y habilita a los usuarios para extender las ontologías en múltiples formas. Los instrumentos en Ontolingua se orientan hacia la creación, edición y actualización de ontologías.

#### 2.2.2.13 *Care*

En [72] se hace referencia a las herramientas NETCARE y CARE donde la primera es predecesora de la segunda. La diferencia entre las dos radica en que la primera está desarrollada en Prolog y la segunda es desarrollada como un applet de Java. Sin embargo la filosofía de las dos es la misma, es decir que son herramientas que asisten a la construcción gráfica de grafos conceptuales. Permiten modelar bases de datos por medio de GCs, e involucran algunas operaciones del lenguaje de consulta estructurado (*Structured Query Language Sql*).

#### 2.2.2.14 *Grafitos*

De acuerdo al documento citado en [74], es un editor didáctico para grafos conceptuales que hace parte de un proyecto de grado en el programa Ingeniería de Sistemas y Computación, de la Universidad Tecnológica de Pereira. El objeto de Grafitos es facilitar el aprendizaje de la naturaleza de los grafos conceptuales. Cuenta con una interfaz para la composición y la edición gráfica de los mismos, así como la transformación automática a su representación en formato lineal y viceversa. Según [74] la herramienta permite: a) Definir y editar de una base de conocimiento (ontología de conceptos y relaciones); b) Representar visualmente la base de conocimiento; c) Componer grafos conceptuales; d) Transformar los grafos conceptuales de la forma gráfica a la forma lineal y viceversa y e) trabajar de modo libre, es decir con restricciones semánticas definidas por el propio usuario.

#### 2.2.2.15 *Somed*

Esta herramienta [11] se enmarca dentro del dominio de la medicina y actualmente se encuentra en su versión 3. Permite esquemas de códigos que incluyen modificadores y referencias de tiempo. Permite manejar grafos conceptuales de forma consistente ya que permite relacionar los códigos con herramientas para modelos relacionales de bases de datos y de cálculo de predicado de primer orden.

## 2.2.3 RECUPERACIÓN AUTOMÁTICA DE GRAFOS CONCEPTUALES

No son muchos los trabajos cuyo objeto es la construcción de los grafos conceptuales sin embargo de los encontrados siguen claramente dos tendencias: Los que utilizan un procedimiento manual o determinísticos en el sentido que no requieren entrenamiento de entrada y que siguen paso a paso ciertas órdenes y reglas y los que siguen un procedimiento estadístico que por medio de aprendizaje de máquina intentan detectar una o varias partes del grafo conceptual. En la mayoría de los trabajos se utiliza una gramática para obtener una notación estructural inicial y fácil de procesar, se ayudan de léxicos como el WordNet [64] u ontologías y generalmente se incluyen un conjunto de reglas que permiten seguir los pasos secuenciales o realizar el aprendizaje automático.

### 2.2.3.1 Método determinístico

Dentro de esta clasificación se han incluido aquellos trabajos que muestran cómo construir un grafo conceptual desde el punto de vista metodológico, estructural o procedimental. En este primer trabajo [36], se muestra cómo construir un grafo a partir de una frase desde el punto de vista metodológico. Mientras que en [66], se presenta un caso de la creación de estos desde el punto de vista de modelamiento ya que se crean los grafos conceptuales similar a como se crean las clases utilizando el lenguaje unificado de modelamiento (UML).

Como ya se ha notado se utilizan diferentes gramáticas para conseguir una primera aproximación a los grafos. Dentro de los primeros trabajos que utilizan un analizador gramatical para posteriormente obtener los GC, está el de [95], presentado por el mismo creador de los grafos Sowa. El árbol sintáctico es modificado asociando cada elemento con un árbol canónico para así obtener un gran grafo conceptual por cada sentencia.

En los trabajos que se mencionan a continuación, con algunas variaciones se construye el grafo conceptual siguiendo un procedimiento lineal. En [16] se presenta un sistema que convierte un texto libre en español a grafos conceptuales. El texto libre pasa por un analizador morfológico y un analizador sintáctico. A partir de los árboles se crean los grafos. En el trabajo [27] se trabajan textos en francés y los grafos conceptuales se construyen a partir de dos estructuras iniciales. La primera modela la representación semántica que permite a través de sistemas primitivos expresar patrones cognitivos, mientras que la segunda es un diagrama temporal que representa las restricciones temporales y que son exploradas a través de los tiempos de los verbos. Uniendo estas estructuras obtienen el modelo de grafos conceptuales a un nivel básico.

Sobre el dominio específico de la medicina se encuentran trabajos [77, 76], que buscan la estructuración de reportes médicos y los GC se construyen a partir de anotaciones sobre el texto que se consiguen con base en reglas de proximidad, información semántica y sintáctica. En [84], se construye un sistema llamado Metaxa que se enmarca en el dominio de las anotaciones sobre radiologías y se representa el contenido de los reportes en grafos conceptuales. En [20] a pesar de estar en desarrollo vale la pena mencionarlo ya que espera

lograr la representación de la semántica para reportes de diagnóstico con base en la nomenclatura multiaxial y la indexación de sustantivos, para luego transformar dichos índices en grafos conceptuales. En [13] presenta algunos aspectos de cómo se logró transcribir registros de asistencia médica a una estructura de grafos conceptuales, con el fin de realizar mejor consultas y recuperación. El trabajo se centra en el modelamiento de una base de conocimiento centrada en la estructura de grafos conceptuales, muy similar a como se modela bajo un paradigma de entidad/relación.

### 2.2.3.2 Método estadístico

En esta categoría se incluyen los trabajos que en alguno de sus procedimientos incluyen procesos estadísticos, incluso si el procedimiento no es directamente para crear los grafos conceptuales. En [36,37,38], utilizan el parser Charniak<sup>1</sup> para obtener la estructura sintáctica y con base en esta y el VerbNet [45] y/o el WordNet identifican los roles semánticos. Vale la pena resaltar el [33], un prototipo de un procesador semántico de sentencias italianas que usa un léxico especial manualmente adquirido y que extrae parcialmente conocimiento de los textos a través de plantillas o por la generación de un conjunto de patrones lingüísticos.

El problema planteado en [90] es el de construir automáticamente a partir de texto una base de conocimiento léxico expresada en grafos conceptuales. Para lo cual, etiquetan las palabras, utilizan un analizador gramatical para construir los árboles gramaticales y a partir de una serie de reglas transforman dichos árboles en grafos conceptuales. Una vez transformados se optimizan los grafos al desambiguarlos estructural y semánticamente. Para la construcción de las reglas tanto del analizador gramatical como de las reglas que permiten pasar los árboles gramaticales en grafos conceptuales se utilizan diferentes técnicas incluidos métodos heurísticos.

Utilizando las gramáticas de enlace [50, 89], se pueden citar la investigación presentada en [101], donde se construyen los GC a partir de dichas gramáticas, el WordNet y las máquinas de aprendizaje como un problema de clasificación que puede ser entrenado para diferentes dominios y un conjunto pequeño de reglas. En el trabajo se aprovecha la similitud que existe entre las gramáticas de enlace y los grafos conceptuales, para formular un problema de clasificar las estructuras conceptuales. Una implementación del sistema fue desarrollada para el laboratorio de investigación de la International Business Machines IBM en China. El trabajo [42], también utiliza estas gramáticas para producir las relaciones sintácticas entre las palabras de una sentencia e incorpora las condiciones necesarias para proporcionar las necesidades especiales del dominio. Para generar el grafo conceptual, la estructura generada

---

<sup>1</sup> Este analizador gramatical probabilístico trabaja con máxima entropía y produce un estilo de árbol *Penn tree-bank*. Fue Probando en el Wall Street Journal y logró un 90.1% en promedio de exactitud para sentencias que no excedían de 40 palabras y 89.5% para sentencias con longitud mayor a 100 palabras.

la cruzan con las raíces. El generador de grafos conceptuales consistió de un algoritmo que va identificando los diferentes elementos de la sentencia; El determinante, el adjetivo, el sustantivo, la preposición, el verbo, el sustantivo preposicional, las relaciones, etc.

En [48] se trata de desarrollar un sistema de inducción gramatical que pueda asignar sentencias descriptivas a modelos ontológicos representados por grafos conceptuales. Para esto el sistema aprende las reglas de asociación entre las sentencias y representaciones ontológicas a través de un enfoque de clasificación. Los elementos del modelo son agrupados bajo clases: conceptos, relaciones y contenedores o módulos estructurales. Los conceptos y las relaciones se dividieron adicionalmente de acuerdo a varios aspectos. El grafo conceptual se obtiene a partir de una estructura de dependencia.

El propósito en [22] es el de explorar la estructura de la frase para generar hipótesis de conceptos, calificar estos de acuerdo a la probabilidad y seleccionar aquellas más creíbles a través agrupamientos (método no supervisado). La adquisición del conocimiento inicial se logra a través de un analizador basado en la gramática de árbol adjunto y la extracción de dependencias [97] se logra con un analizador de dependencias.

Por último vale la pena resaltar las investigaciones relacionadas con las operaciones sobre los grafos conceptuales, como la presentada en [67], donde se definen algunas medidas para la semejanza entre los grafos conceptuales.

## **3. MÉTODO BASADO EN REGLAS SINTÁCTICAS Y SEMÁNTICAS**

### **3.1 INTRODUCCIÓN**

De acuerdo con el estado del arte, una de las formas más comunes de generar automáticamente los grafos conceptuales es lograda incluyendo como entrada una estructura sintáctica. Con base en esta estructura se incluye uno o varios procedimientos que permiten transformar dicha estructura a la de grafos conceptuales (ver Figura No. 1).

Las estructuras sintácticas generalmente son generadas a través de los analizadores sintácticos o gramaticales (“parsing”). Estos analizadores pueden constituirse a partir de reglas o de procedimientos estadísticos. Los que se basan en reglas pueden incluir las reglas como parte del proceso de análisis (analizador) o las pueden incluir como una entrada a este analizador. En este último caso, el conjunto de reglas es conocido como una gramática.

Por último, es de aclarar que tanto los analizadores como los procesos que transforman la estructura sintáctica a los grafos, generalmente utilizan recursos adicionales que permiten enriquecer la estructura en cuestión o resolver ambigüedades.

Con el fin de describir el trabajo desarrollado, en este capítulo se incluirán los aspectos más relevantes del trabajo previo, se describirá el método propuesto y se mostrarán las ventajas de este último.

### **3.2 TRABAJO PREVIO**

Excluyendo la creación de grafos conceptuales de forma manual o como parte de otra estructura como un modelo entidad-relación y analizando el estado del arte se puede apreciar que son pocas las investigaciones que se dedican a la creación automática de los grafos conceptuales a partir de texto. Considerando las encontradas se pueden mencionar algunas características.





En un vocabulario o dominio contralado, algunas sentencias ocurren solamente en el dominio específico y se asume que las sentencias de dichos dominios tienen las siguientes características:

- El conjunto de vocabulario es relativamente pequeño.
- Los términos y las jergas del dominio aparecen frecuentemente.
- La gramática es elegante.
- La semántica ambigua es rara y las ambigüedades se pueden resolver rodeando el contexto.

### 3.2.3 SE CREAN REGLAS SOBRE ESTRUCTURAS SINTÁCTICAS

La utilización de estructuras sintácticas como entrada al proceso de generación de los grafos, es un procedimiento muy utilizado. Las estructuras sintácticas son generadas utilizando Analizadores Sintácticos (*parsing*), por lo que son variadas las estructuras que se utilizan. Por ejemplo Zhang [102] y Kamaruddin [42] utilizan las estructuras de enlace (*Link Grammar*). En el primer caso identifican los modificadores para convertir la estructura de enlace a los Grafos conceptuales, en el segundo se identifican las reglas que permiten modificar la estructura. Hensman en [36,37,38] utiliza una estructura de constituyentes generada por el analizador Charniak. Fernandez [23], utiliza una estructura de adjunción de árboles generada por una meta-gramma. Cruz [16] utiliza estructuras de dependencia para luego crear un código que paso a paso transforma dicha estructura en un grafo.

### 3.2.4 USO DE RECURSOS

Tanto en el proceso de análisis sintáctico encargado de generar las estructura sintácticas que se mencionan en numeral anterior, como en la transformación de dicha estructura al Grafo conceptual se utilizan recursos léxicos con el fin de resolver roles semánticos, homónimos, sinónimos, entre otros. Dentro de estos recursos el más utilizado es el WordNet.

## 3.3 TRABAJO PROPUESTO

De acuerdo a lo expuesto previamente se puede concluir que no existe una gramática que directamente genere los Grafos conceptuales. La generación hasta el momento se ha logrado modificando estructuras sintácticas existentes, por lo que cabe preguntar si se puede crear una gramática que genere los grafos conceptuales, partiendo del hecho de que los Grafos conceptuales son estructuras más semánticas que sintácticas. En este capítulo el trabajo se direcciona a la creación de una gramática que al ser procesada por un analizador permite generar los Grafos conceptuales (ver Figura No. 2).

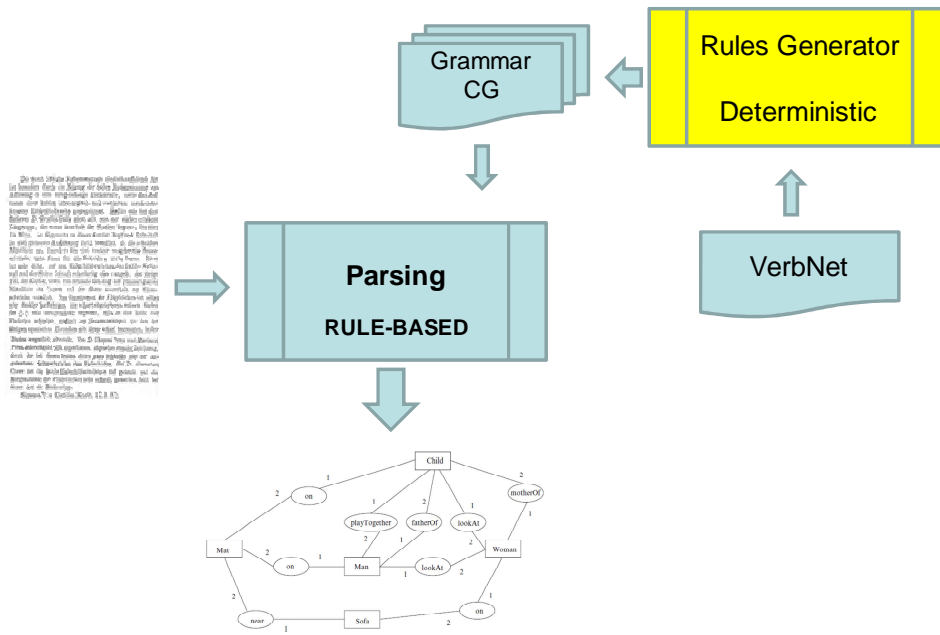


Figura No. 2. Generación de CG por medio de reglas gramaticales

### 3.3.1 LOS GRAFOS CONCEPTUALES Y LAS ESTRUCTURAS SINTÁCTICAS

Para la construcción de la gramática para los GCs, se hizo necesario asumir dicha construcción desde el punto de vista sintáctico, aun cuando se espera como resultado una estructura más semántica que sintáctica. Lo anterior obedece a que toda gramática está dirigida esencialmente a la construcción de estructuras sintácticas.

La generación de una estructura sintáctica a partir de una oración depende tanto de la gramática o especificación formal de las estructuras permisibles en el lenguaje como del método utilizado para analizar la oración y construir la estructura escogida. Los árboles son las estructuras sintácticas más utilizadas, más estandarizadas y más investigadas y dado que un árbol es considerado como una forma especial de un grafo, los GCs contienen a la estructura árbol y pueden ser tratados como un caso especial de las estructuras sintácticas. Para lograr dicho tratamiento, se asumieron una serie de supuestos que se explicarán a lo largo del presente capítulo.

La generación de las estructuras sintácticas ha evolucionado bajo distintas corrientes. En especial dos de estas corrientes han sido muy investigadas y se han desarrollado casi que simultáneamente. Por un lado los árboles de constituyentes y por el otro los árboles de dependencia. Los árboles de constituyentes son el producto de las gramáticas generativas propuestas por Chomsky [14] y los árboles de dependencia son producidas por las gramáticas de dependencia propuestas por Tesnière [97]. Tanto la una como la otra, han dado lugar a una variedad de teorías que sustentan el entendimiento del lenguaje natural y la creación de las estructuras sintácticas.

### 3.3.1.1 Elementos Gramaticales

En este ítem se definen dos de las características o rasgos lingüísticos que se deben tener en cuenta para la definición de una gramática generativa.

- **Categoría gramatical.** Corresponde a la categoría de tipo semántico y no-funcional. La gramática tradicional distingue nueve partes de la oración: determinante, sustantivo, pronombre, verbo, adjetivo, adverbio, preposición, conjunción e interjección. El llamado etiquetado gramatical (*Part-of-speech tagging*, *POS tagging* o *POST*), es el proceso de etiquetar cada una de las palabras de un texto con su categoría gramatical.
- **Sintaxis.** La sintaxis estudia las formas en que se combinan las palabras, así como las relaciones sintagmáticas y paradigmáticas existentes entre ellas. Es la parte de la gramática que estudia las reglas que gobiernan la combinatoria de constituyentes sintácticos y la formación de unidades superiores a estos, como los sintagmas y oraciones gramaticales.

### 3.3.1.2 Árbol de Constituyentes

Un constituyente es un conjunto de palabras que funcionan como una unidad dentro de una estructura jerárquica de una oración. Cada una de las partes de un constituyente puede ser a su vez un constituyente. El sintagma es un tipo de constituyente formado por al menos una palabra, una de las cuales se comporta como núcleo sintáctico, ya que el sintagma asume la función de este núcleo.

Los árboles de constituyentes son producto de las gramáticas generativas transformacionales propuestas por Noam Chomsky [14] y permiten definir un conjunto de reglas o principios que predicen las combinaciones que aparecen en las oraciones gramaticalmente correctas para un determinado lenguaje. Estas gramáticas asumen que una oración está organizada jerárquicamente en bloques o constituyentes.

El modelo de gramática generativa establece tres componentes que identifican los aspectos lingüísticos de una oración.

- **Componente Fonológico.** Encargado de asignar una realización fonética a la oración, permitiendo que las estructuras se hagan perceptibles.
- **Componente Sintáctico.** Encargado de la generación de la sentencia y formado por reglas formales que se enmarcan dentro de las gramáticas formales libres de contexto. Una gramática formal consta de un conjunto finito de símbolos terminales, un conjunto finito de símbolos no terminales, un conjunto de reglas de producción con un lado izquierdo y otro derecho, y un símbolo inicial. Las reglas se aplican sustituyendo la parte de la izquierda por la parte de la derecha. Una derivación es una secuencia de aplicaciones de

reglas. Los símbolos terminales corresponden al conjunto de palabras del idioma en cuestión, en este caso inglés y no terminales al conjunto de partes sintácticas de una oración.

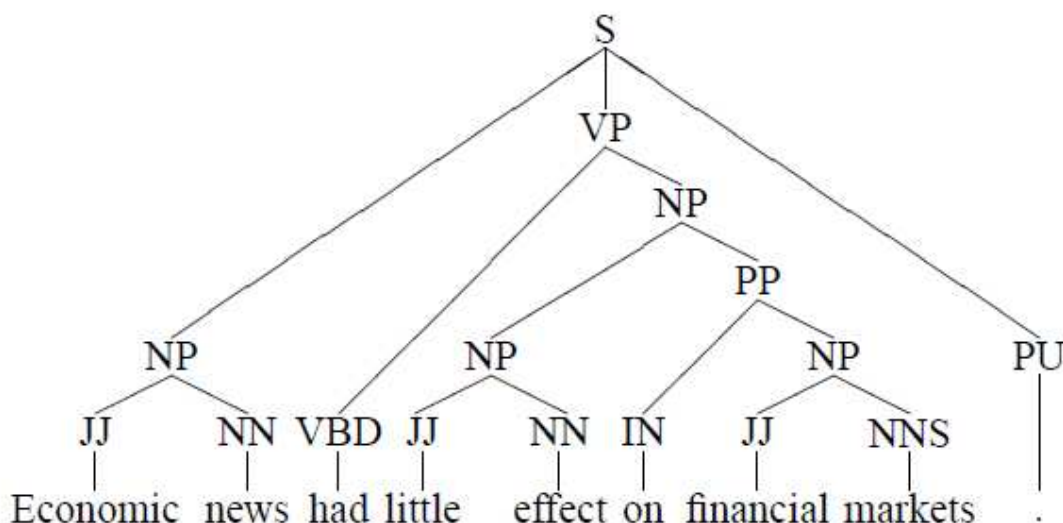


Figura No. 3. Árbol de constituyente para la frase “Economic news had little effect on financial markets”<sup>2</sup>

Por ejemplo, reglas que hacen parte de una gramática generativa para inglés podrían ser:

S: sentencia; PN: sintagma sustantiva (*noun phrase*); PV: sintagma verbal (*verb phrase*);  
det: determinante; S: sustantivo (*noun*).

S -> PN + PV

PV -> V + PN

PN -> det + N

det -> {the, a, an, ..}

N->{car, cat, dog, ..}

V->{run, eat, .}

*Componente Semántico.* Encargado del procesamiento semántico de la oración y por ende el encargado de interpretar el significado de una oración. Este componente debe recurrir a la estructura profunda de la frase para poder interpretar, por ejemplo, cuando dos frases o palabras significan lo mismo.

<sup>2</sup> \* Ejemplo de “Dependency Grammar and Dependency Parsing” by J. Nivre

Para la frase “*The economic news had little effect on financial markets*” la estructura de árbol de constituyente se puede visualizar en la Figura No. 3. Dicha estructura es generada bajo la teoría de gramáticas generativas de constituyentes. La estructura incluye todas las palabras que hacen parte de la sentencia, las categorías gramaticales, así como los sintagmas (S: oración (“sentence”), VP: frase verbal (verbal phrase), NP: frase nominal, PP: frase preposicional, NN: sustantivo, jj: adjetivo VBD: verbo principal), la constituyentes y sub-constituyentes. En esta estructura, no existe la relación directa entre las palabras de la oración.

### 3.3.1.3 Árbol de Dependencia

Estas estructuras son el producto de aplicar las teorías de Tesnière [97] sobre las gramáticas de dependencia. Esta gramática se fundamenta en la asignación y combinación de valencias dadas a las palabras. De acuerdo a las valencias se pueden combinar y formar estructuras sintácticas principales (actuales) y secundarias (modificadoras). La estructura del árbol de dependencia es completamente simétrica y supone que existe una palabra de mayor jerarquía (palabra rectora) y otras que dependen de esta.

Los elementos están directamente relacionados y las marcas sintácticas se identifican mediante roles sintácticos. En la frase de ejemplo “*The economic news had little effect on financial markets*” se puede apreciar (ver Figura No. 4) la relación directa entre las palabras a diferencia de la estructura de constituyentes donde se relacionan las constituyentes. Igualmente se visualiza que las palabras rectora principal corresponde al verbo. De esta palabra le siguen en jerarquía los sustantivos. Y los sustantivos se unen a las preposiciones para continuar con la jerarquía.

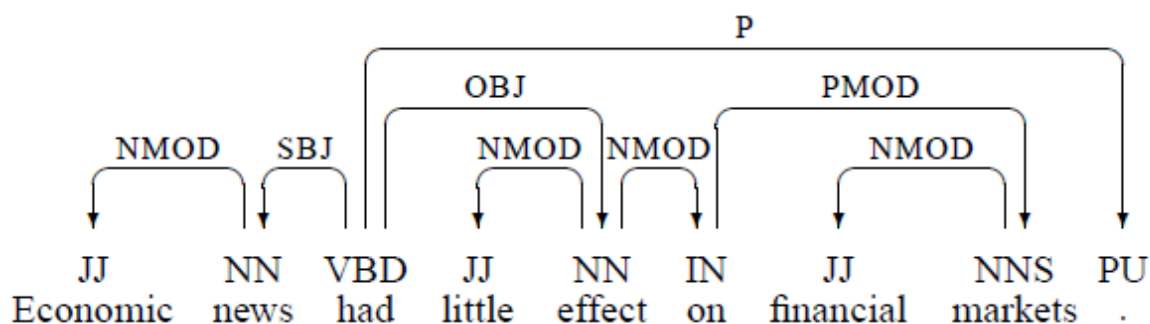


Figura No. 4. Árbol de Dependencia para la frase “Economic news had little effect on financial markets”

### 3.3.1.4 Grafo Conceptual

Crear una gramática específica para los GC, nace de la necesidad de incluir las particularidades de los GC. Teniendo en cuenta la frase “*The economic news had little effect on financial markets*” y el GC de la Figura No. 5, se pueden explicar dichas particularidades.

Los GC se representan a través de dos tipos de nodos. Nodos concepto y nodos relación. En la Figura No. 5, cada una de las palabras-contenido (verbos, sustantivos, adjetivos y adverbios) corresponde a un nodo concepto (cuadrados). Los nodos relación dependen de las palabras-funcionales (determinantes, conjunciones, adverbios, preposiciones e interjecciones) así:

- Con palabras como “*on*”, “*in*” “*between*”, entre otras, se convierten en un nodo relación.
- Con palabras como los determinantes y los artículos, enriquecen a los nodos concepto. Como la palabra “*the*” que enriquece al nodo concepto *News*, cuantificándolo con el símbolo “#”.
- Las palabras de los nodos relación como *agnt*, *rcpt* y *attr* corresponden a abreviaturas de roles semánticos que no hacen parte de la oración pero que están muy relacionados con las palabras funcionales.
- El tiempo del verbo como *Past* pasa a ser un elemento externo del grafo y el verbo que es representado en un concepto se conjuga en presente.

Las estructuras de los GCs tienen en común con las estructuras de dependencia que relacionan directamente las palabras. Pero se diferencian en que no incluyen todas las palabras.

Las relaciones para los GCs se dan en términos semánticos, mientras que para las estructuras de dependencia son de orden sintáctico.

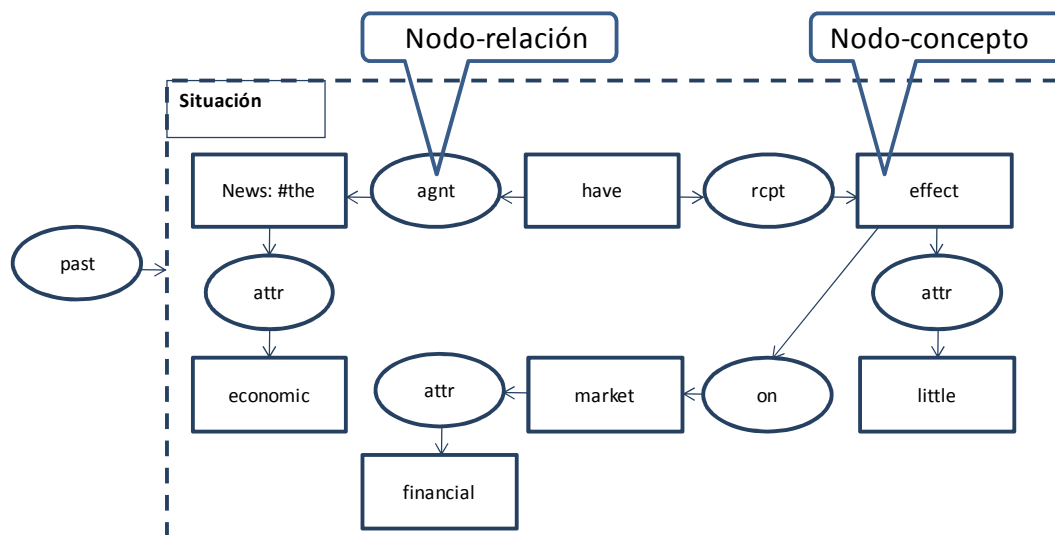


Figura No. 5. GC de “Economic news had little effect on financial markets”

Para la creación de la estructura de los GC en esta investigación se desarrolló una gramática que toma algunos elementos de las gramáticas de constituyentes y las gramáticas de dependencia.

### 3.3.2 GRAMÁTICA PARA LOS GRAFOS CONCEPTUALES (GRAMÁTICA-GC)

La gramática para los Grafos Conceptuales se enmarca como una mezcla entre las gramáticas generativas propuesta por Chomsky [14] y las gramáticas de dependencia propuestas por Tesnière [97]. De las constituyentes se mantuvo la generación de reglas transformacionales, así como la direccionalidad y de las gramáticas de dependencia se conservó la relación directa entre las palabras, los encabezados y el etiquetado a través de roles, con la diferencia de que en la gramática desarrolla, gramática-CG, los roles no corresponden a bloques sintácticos sino a roles semánticos.

#### 3.3.2.1 *Automatización*

La tarea de crear gramáticas para la construcción de las estructuras sintácticas, generalmente se logra con la ayuda de expertos, quienes deben contar con el conocimiento del lenguaje en particular, las sintaxis, la morfología, la semántica, la forma como se deben expresar las reglas, entre otros. Lo anterior ha conducido a que se busquen alternativas que por lo menos mitiguen la tarea del experto. La gramática-GC fue construida de manera automática con la ayuda del léxico VerbNet. Considerar en crear gramáticas que no impliquen la ayuda directa y dedicada para tal fin del experto es el trabajo que se ha desarrollado.

VerbNet es un léxico construido por expertos pero con un fin muy distinto al que se le ha dado en este trabajo. El VerbNet surge como un recurso léxico que complementa a recursos como el WordNet [64] y las clases de Levin base del VerbNet. VerbNet es un léxico hecho para el idioma inglés e independiente del contexto. Utiliza la clasificación de los verbos de Levin e incluye para cada clase, información sobre los roles semánticos. Actualmente cuenta con 23 roles y 274 clases verbales. Este recurso se presenta como un diccionario léxico de verbos organizado jerárquicamente, que aporta información sintáctica y semántica para los verbos del inglés. Los verbos son agrupados en clases según criterios semánticos, permitiendo que un verbo pueda estar en más de una clase si éste cuenta con diferentes acepciones. Cada una de estas clases describe las características sintácticas y semánticas.

A pesar de que el VerbNet no incluye todos los verbos que hacen parte del idioma inglés y su diseño o forma de presentar el léxico no permite una fácil sistematización, es un recurso ampliamente usado para el procesamiento de lenguaje natural.

#### 3.3.2.2 *Asignación de palabras*

Cada uno de los elementos de la gramática-GC se estableció de acuerdo a las premisas establecidas, las características del VerbNet y la función que puede ejercer cada palabra sobre la estructura del grafo.

La función de las palabras se establece si la palabra pertenece a una categoría Léxica o a una categoría funcional. A las categorías léxicas pertenecen aquellas palabras con contenido semántico y que permiten la creación de nuevas palabras a través de funciones como la composición y derivación. A esta categoría pertenecen los sustantivos, los adjetivos, los verbos y los adverbios. A la categoría funcional pertenecen las palabras que tienen una función principalmente gramatical y sus propiedades dependen de su ubicación física. Las palabras que pertenecen a la categoría funcional por si solas no poseen un significado. A esta última categoría pertenecen los determinantes, las conjunciones, los auxiliares y los nexos subordinados.

### 3.3.2.3 Premisas

Teniendo en cuenta la definición del un grafo conceptual, en la gramática-GC se incluyeron las siguientes premisas.

- Se definieron dos tipos de nodo (ver Figura No. 5): nodo-concepto (cuadrados) y nodo-relación (redondos).
- Los nodos-relación son incluidos dentro de la gramática como un rol semántico. Esto es que para que aparezcan los nodos-relación dentro de la estructura, se incluyeron los roles: atributo (attr), agente (agnt), destinatario (rcp), entre otros, como un rol semántico.
- Se asumió que entre dos palabras siempre debe existir un rol semántico.
- Las palabras de la oración que pertenecen a la categoría léxica (palabras-contenido) pasan a ser nodo-concepto del grafo conceptual.
- Las palabras funcionales de la oración pueden cumplir una de tres funciones. 1) que desaparezcan; 2) que se conviertan en nodo-relación (rol semántico) y 3) que pasen a enriquecer el contenido de un nodo-concepto.
- Se tuvo en cuenta una cabeza rectora para incluir la direccionalidad de las aristas.
- Los nodos-relación son definidos automáticamente de acuerdo al VerbNet.
- La gramática no detecta Entidades nombradas
- La gramática no incluye tiempos verbales como el *past* de la Figura No. 5.
- La gramática no incluye situaciones como la etiqueta de *situation* de la Figura No. 5.

### 3.3.2.4 Nodo-concepto

Como ya se especificó en las premisas los nodos concepto están formados por las palabras-contenido. Pero dado que, algunas de estas palabras ejercen una función adicional sobre el grafo, se les dio un tratamiento diferente a algunas de ellas.

- *Sustantivos como Nodo-concepto*

Los nodos concepto formados por un sustantivo son generalmente enriquecidos por la palabra funcional que preceda al sustantivo. De acuerdo al estándar de los GC se hizo una clasificación basada en el símbolo, con excepción del símbolo vacío ( $\emptyset$ ), definido por este



trabajo. Es de aclarar que en el estándar se definen los símbolos de acuerdo a algunos ejemplos y no se establece en que situaciones se debe aplicar uno u otro símbolo.

En este trabajo se clasifican dentro de cada símbolo aquellas palabras que se pueden incluir como un determinante del sustantivo (ver Cuadro No. 1). Se clasificaron bajo el símbolo #, los artículos, los pronombres demostrativos y los adjetivos posesivos que, aun cuando no son palabras funcionales algunas veces si se comportan como un determinante del sustantivo.

Las palabras que indican unos, algunos, como *any*, *every* se incluyeron bajo el símbolo {\*}. Las palabras que denotan números se clasificaron bajo el símbolo @ y las que denotan *todo* con el símbolo  $\forall$ . Las que significan vacío se incluyeron bajo el símbolo  $\emptyset$ .

Entrada Gramática	Palabras	Símbolo	Funcion Gramatica	Ejemplo	Nodo GC
ART	a, an, the	#	DET	the children is fat	[childre:#the]
PRON_DEMO	this, that, these, those	#	DET	This house is big	[house:#this]
PRON_INDEF_SOME	any, anybody, anyone, anything, both, each, either, everybody, everyone, everything, few, many, another, several, some, somebody, someone, somethinge	{*}	DET	Any kid would have behaved like that	[kid:{*}]
PRON_INDEF_ALL	all	$\forall$	DET	all people get	[people: $\forall$ ]
PRON_INDEF_NONE	no	$\emptyset$	DET	no people came	[people: $\emptyset$ ]
ADJ_POS	my, your, his, her, our, their	#	DET	Their house is similar to ours	[house:#their]
NUMER_ORD	first, second, ..	@	DET	first place	[pleace:@first]
NUMER_CAR	1, 2, ..	@	DET	44 children arrived	[children:@44]

Cuadro No. 1. Clasificación palabras que enriquecen los nodos concepto de los GC.

En el Cuadro No. 1 se incluyó un ejemplo de cada una de las situaciones. Por ejemplo en la frase “*the children is fat*”, el nodo que denota al niño se especifica como [childre:#the].

- *Palabras-funcionales como Nodo-concepto*

Algunas palabras funcionales cumplen la función de un sustantivo en una sentencia, tal es el caso de pronombres subjetivos “*I*”, “*he*”. Dentro de la gramática, a esta categoría de palabras se le dio el nombre de *NOUN\_LIKE* y se puede visualizar en el Cuadro No. 2. Cada una de las categorías que hacen parte de esta la regla, a su vez, están compuestas por un conjunto de símbolos terminales.

El símbolo @ es parte de la definición de la gramática e indica que el nombre que precede a este símbolo, será el nodo que encabeza la dirección de la flecha.

REGLA	
NOUN_LIKE	
	-> @:PRON_SUBJ
	-> @:PRON_OBJ
	-> @:PRON_INDEF_SOME
	-> @:PRON_INDEF_ALL
	-> @:PRON_INDEF_NONE
	-> @:NUMER_ORD
	-> @:NUMER_CAR
	-> @:PRON_POSS
	-> @:PRON_DEMO

Cuadro No. 2. Regla Noun\_Like: Palabras que pueden funcionar como un Sustantivo en la gramática-GC (El símbolo @ es parte de la definición de la gramática e indica que el nombre que precede a este símbolo será el nodo que encabeza la dirección de la flecha).

- *Atributo como Nodo-concepto*

En los grafos conceptuales los atributos se establecen como un nodo-concepto. Un atributo puede ser definido dentro de una oración de diferentes modos. La forma más común es definida a través de un adjetivo seguido por un sustantivo. Puede presentarse por un sustantivo seguido por un sustantivo, entre otras. En el Cuadro No. 3, se presentan las formas incluidas dentro de la gramática. La primera regla incluye el caso cuando a un adjetivo lo precede el verbo is. La segunda regla especifica el caso en que el atributo es expresado por una frase preposicional. En el ejemplo “attributes of the database”, la palabra “attributes” ejerce el papel de atributo en la frase, mientras que “database” ejerce el papel de sustantivo. La última regla incluye los casos en que de una lista de sustantivos, algunos ejercen la función de adjetivos o atributos. Es el caso de “wood house” y de “black canvas chair”. Los nodos-relación son incluidos a través del rol “attr”

- *Verbo como Nodo-concepto*

Las reglas para los verbos son incluidas de manera automática. A partir de los clases definidas por VerbNet se incluyeron no solo las reglas sino los roles semánticos correspondientes. Para esta tarea se desarrolló el software que permitió la creación de dichas reglas. En el Cuadro No. 5 se incluyeron las primeras ocho reglas con su correspondiente ejemplo y grafo.

Para la primera regla “-> agent:LIS\_NP @:V\_ACCOMPANY-51-7 theme:LIS\_NP” se incluyeron los roles agent y theme definidos en VerbNet para los verbos que están clasificados bajo V\_ACCOMPANY-51-7. El ejemplo correspondiente “Jackie accompanied Rose” cuenta con el verbo accompanied que hace parte del grupo V\_ACCOMPANY-51-7. El grafo para este ejemplo es [Jackie]<-(agent)<-[accompanied]->(theme)->[Rose]. El nombre de V\_ACCOMPANY-51-7 fue definido por el software desarrollado y depende del grupo asignado en VerbNet. A partir del VerbNet se incluyeron un total de 1443 reglas. En las

reglas de la gramática se incluyó la clasificación verbal, la semántica, los ejemplos y las restricciones sintácticas de clase. La gramática en total cuenta con 1.911 reglas base y 6054 diccionario. Reglas que al expandirlas resultaron en un total de 17 Mg de líneas de código (reglas). El diccionario se obtuvo con base en los archivos de WordNet y en las colecciones de trabajo.

REGLA	EJEMPLO	GRAFO CONCEPTUAL	
SENTENCE	-> LIS_NP @V_IS attr:ADJ	the children is fat	[children:#the]->(attr)->[fat]
NP	-> @:NP attr:PP	attributes of the database	[database:#the]->(attr)->[attribues]
	-> @:LIST_N_WITH_LIST_ADJ_WITH_DET		
LIST_N_WITH_LIST_ADJ		wood house	[house]->(attr)->[wood]
	-> @:LIST_N		
	-> attr:ADJ @:LIST N WITH LIST ADJ	black canvas chair	[chair]->(attr)->[canvas]->(attr)->[black]

Cuadro No. 3. Reglas para Atributo como nodo concepto

- *Rol Semántico como Nodo-Relación*

Según la teoría de Gramática de caso (“case grammar”), desarrollada por Fillmore, la relación entre los verbos y los sintagmas nominales está definida por un rol semántico o caso. Cada verbo permite un determinado número de casos, dando lugar las clases verbales. Para los grafos conceptuales estos roles semánticos son definidos como algunos de los nodos relación. En el cuadro se pueden ver algunos de los roles utilizados.

- *Preposiciones como Nodo-relación:*

Generalmente las preposiciones forman un nodo-relación. Por ejemplo, *in*, *between*, *on*, *about*

- *Atributo como Nodo-relación*

El rol atributo se presenta siempre que haya una descripción (adjetivación) de un objeto (Un sustantivo). Dicha adjetivación de un objeto puede darse a través de varias formas (Ver Cuadro No. 6). En inglés un atributo puede presentarse cuando existe un adjetivo que precede a un sustantivo “*fat cat*”. Cuando un sustantivo precede a otro sustantivo o lista de sustantivos, ejemplo “*color box*”, cuando el adjetivo es definido por un sustantivo o lista de sustantivos precedido por una preposición, como en “*house of wooden*”. Adicionalmente es de tener en cuenta que así como un sustantivo puede ejercer la función de adjetivos, un verbo puede ejercer la función de sustantivo.

Nombre Rol	Descripción
Agent	El que causa un evento
Attribute	Atributo
Beneficiary	El beneficiario de un evento
Destination	Destino de un objeto en un evento de traslado
Experiencer	El que experimenta un evento
Instrument	El instrumento utilizado en un evento
Location	Lacalización
Source	Origen del obojto en un vento de traslado
Theme	El afectado en un vento en forma directa
Force	El causante involuntario de un evento
Result	El resultado de un evento
Content	Proposición o contenido de un evento proposicional
Goal	Destino de un objeto en un evento de traslado

Cuadro No. 4. Ejemplo de Roles semánticos

### 3.3.3 PARSING CIC-IPN

Para analizar la gramática-GC se uso la herramienta desarrollada [28] en el Laboratorio de procesamiento de lenguaje natural de Instituto politécnico Nacional de México CIC-IPN y que se en encuentra disponible en [nlp.cic.ipn.mx/tools/parser](http://nlp.cic.ipn.mx/tools/parser). La herramienta con algunas modificaciones permitió compilar y generar todas las posibles combinaciones de reglas de la gramática desarrollada.

Como resultado de este compilador se obtuvieron un total de 17 Mg en reglas. Con base en este conjunto de reglas y el texto correspondiente a la colección de ImageClef, la herramienta generó una serie de aproximaciones a los grafos conceptuales. Un ejemplo se ilustra en el Cuadro No. 8 donde aparece el Grafo conceptual para el texto “*Thymic asthma. Sagittal illustration of tracheal narrowing (arrows) due to the enlarged thymus. In this case, tracheal narrowing is exaggerated with neck extension. Th = thymus. (Reprinted, with permission, from reference 6.) The right place at the wrong time: historical perspective of the relation of the thymus gland and pediatric radiology*”, correspondiente al documento número 33 de la colección ImageClef2008. La sentencia del ejemplo está formada por 6 oraciones:

- *Thymic asthma.*
- *Sagittal illustration of tracheal narrowing (arrows) due to the enlarged thymus*
- *In this case, tracheal narrowing is exaggerated with neck extension*
- *Th = thymus*
- *Reprinted, with permission, from reference 6*
- *The right place at the wrong time: historical perspective of the relation of the thymus gland and pediatric radiology*

SENTENCE	
Regla	-> agent:LIS_NP @:V_ACCOMPANY-51-7 theme:LIS_NP
Ejemplo	Jackie accompanied Rose
GC	[Jackie]<-(agent)<-[accompanied]->(theme)->[Rose]
Regla	-> agent:LIS_NP @:V_ACCOMPANY-51-7 theme:LIS_NP [spatial:SPATIAL] destination:PP
Ejemplo	Jackie accompanied Rose to the store
GC	[Jackie]<-(agent)<-[accompanied]->(theme)->[Rose]->(destination)->[store:#the]
Regla	-> actor1:LIS_NP @:V_ACQUIESCE-95 [dest_dir:DEST_DIR] actor2:LIS_NP
Ejemplo	The enemy soldiers submitted to us
GC	[enemy:#the]<-(attr)<-[soldiers]<-(actor1)<-[submitted]->(tactor2)->[us]
Regla	-> agent:LIS_NP @:V_ADDICT-96 patient:LIS_NP
Ejemplo	Mary addicted him
GC	[Mary]<-(agent)<-[addicted]->(patient)->[him]
Regla	-> agent:LIS_NP @:V_ADDICT-96 patient:LIS_NP [dest_dir:DEST_DIR] stimulus:PP
Ejemplo	I addicted him to Douglas
GC	[I]<-(agent)<-[addicted]->(patient)->[him]->(stimulus)->[Douglas]
Regla	-> agent:LIS_NP @:V_ADDICT-96 patient:LIS_NP [dest_dir:DEST_DIR] stimulus:V_G [PP]
Ejemplo	Mary addicted him to going on long journeys
GC	[Mary]<-(agent)<-[addicted]->(patient)->[him]->(stimulus)->[going]->(destination)->[long]<-(attr)<-[journeys]
Regla	-> agent:LIS_NP @:V_ADJUST-26-9 patient:LIS_NP
Ejemplo	He adapted himself
GC	[He]<-(agent)<-[adapted]->(patient)->[hinself]
Regla	-> agent:LIS_NP @:V_ADJUST-26-9 patient:LIS_NP [dest_dir:DEST_DIR] destination:PP
Ejemplo	He adapted himself to the situation
GC	[He]<-(agent)<-[adapted]->(patient)->[hinself]->(destination)->[situation]

Cuadro No. 5. Ejemplo se reglas para Verbo como nodo-concepto

En el grafo del Cuadro No. 8 en la primera columna, aparece el número de la oración y número de la palabra que describe al nodo-concepto 1; en la segunda columna aparece el número de la oración y el número de la palabra que describe al nodo-concepto 2; en la tercera columna aparece el nodo-relación el rol semántico que está entre los dos nodos-concepto. Y en la cuarta columna titulada presencia de la relación equivale al número de variantes donde aparece esta relación. En la frase “yo veo gato con telescopio”, existen dos variantes: 1) yo veo; veo gato; veo con telescopio y 2) yo veo; veo gato; gato con telescopio. Entonces, por ejemplo la relación yo veo, aparece en 100% de las variantes, mientras que gato con telescopio en el 50% de las variantes. En la primera columna en algunas ocasiones aparece el “ROOT” que significa el inicio del grafo.

REGLA	
SENTENCE	-> LIS_NP @V_IS [PRON_INDEF_NO] attr:ADJ
	-> @:NP attr:PP
LIST_N_WITH_LIST_ADJ	-> @:LIST_N
	-> attr:ADJ @:LIST_N_WITH_LIST_ADJ
LIST_N	-> [attr:LIST_N] @:N # book shelf store management
	-> attr:N @:LIST_N # top store management
	-> [attr:LIST_N] @:V_ING # Random Cooking Games
	-> attr:V_ING @:LIST_N # Cooking book

Cuadro No. 6. Reglas que denotan un Atributo

### 3.3.4 VALIDACIÓN LOS GRAFOS CONCEPTUALES

Dado que no se encontró una colección de prueba con grafos conceptuales, que permitiera probar la validez de la gramática y con ello la estructura obtenida, fue necesario recurrir a validar la estructura a través de una de las tareas de procesamiento del lenguaje natural, como es recuperación de información. Para el proceso de recuperación de información se utilizó la colección de ImageClef200 [63]. Esta colección consta de un conjunto de anotaciones en texto para imágenes y radiografías médicas. Incluye anotaciones para varios idiomas y permite trabajar con varias tareas, entre ellas recuperación de información a partir de texto. Para este experimento se trabajó con las anotaciones definidas bajo el idioma inglés. La colección consta de 68.715 documentos (24 Mg) y 10 consultas. Se puede agregar que la tarea de recuperación de información sobre esta colección no es fácil, dada la especificidad del dominio y la clase de preguntas. Las preguntas a demás de ser muy cortas, tienen muy poca información semántica. Como se puede ver en el Cuadro No. 7, el texto “*show me photographs*” en las preguntas 21, 22, 23 y 24 no aporta nada a la respuesta y al quitar “*show me*”, por ejemplo en la pregunta 21 queda conformada por dos palabras “*of tumors*”

Dado que la hipótesis a probar corresponde a:

La tarea de recuperación de información en textos cortos, puede mejorar los resultados del modelo espacio vectorial si se utiliza una estructura de grafo conceptual donde se incluye tanto la información de los nodos-concepto como de los nodos-relación.

La hipótesis se valida comparando los resultados de aplicar la tarea de recuperación de información bajo el modelo espacio vectorial contra el modelo de grafos conceptual.

PREGUNTA	TEXTO PREGUNTA
21	Show me photographs of tumours.
22	Show me images of muscle cells.
23	Show me x-ray images of bone cysts.
24	Show me images containing a Budd-Chiari malformation.
25	Merkel cell carcinoma
26	gastrointestinal neoplasm
27	tuberous sclerosis
28	mitral valve prolapse
29	pulmonary embolism all modalities
30	microscopic giant cell

Cuadro No. 7. Preguntas de la colección ImageClef2008

#### 3.3.4.1 Modelo espacio Vectorial

Para construir el modelo espacio vectorial se extrajeron las palabras (*Tokenization*) de cada uno de los documentos y consultas. Se representó a cada documento y consulta con la frecuencia inversa de aparición de las palabras, se eliminaron los “stopWords” y se utilizó el algoritmo de “*Porter stemmer*” [73] para reducir la dimensionalidad. La selección de documentos relevantes se hizo a través del coeficiente Coseno definida en (1).

$$\cos ine = \frac{\sum_j \sum_i d_i q_j}{(\sum_i d_i^2 \sum_j q_j^2)^{1/2}} \quad (1)$$

#### 3.3.4.2 Medida de Similitud entre Grafos

La medida de similitud entre dos grafos G1 (documento) y G2 (consulta) se aplicó como una medida relativa al tamaño de su máxima superposición, es decir que se tomó el máximo común sub-grafo. Para encontrar el máximo común sub-grafo, se encontraron todos los máximos común sub-grafos y de estos se encontró el mayor.

Con el fin de encontrar el máximo común sub-grafo se siguió el siguiente procedimiento.

Se definió a un conjunto de vértices mapeados entre dos grafos etiquetados G1 y G2 como una función  $\varphi: S_1 \leftrightarrow S_2$ , donde  $S_i$  es un subconjunto de vértices o nodos-concepto de  $G_i$ , tal que, las etiquetas (las raíces pertinentes) en los correspondientes vértices (nodos-concepto) coincidan. De acuerdo a las siguientes frases:

*a fat cat sat on a mat and a fat dog slept*

*a fat cat slept and a fat dog sat on a mat*

el primer *fat* de la primera sentencia puede ser mapeada con la primera o segunda ocurrencia de *fat* en la segunda sentencia y el segundo *fat* será mapeado a la segunda ocurrencia; Similarmente existen 6 posibles formas de mapear las *a*'s, dando un total de 12 posibles ocurrencias.

A cualquiera de estos conjuntos isomorfos  $S1 \cong S2$  pertenece el conjunto de vértices del máximo común sub-grafo. Los arcos de este común sub-grafo son aquellos que están presentes con las mismas etiquetas en ambos grafos entre los correspondientes vértices de  $S1$  y  $S2$ , es decir, los arcos tales que  $u \xrightarrow{x} v$  donde si  $u, v \in S_1$ , es un arco en  $G_1$ , y si  $\phi(u) \xrightarrow{x} \phi(v)$ ,  $f(u), f(v) \in S_2$ , es un arco en  $G_2$ .

Para encontrar el máximo común sub-grafo se utilizó una medida de similitud muy similar al a la medida de similitud estándar pero que incluye las tanto las palabras como las relaciones de manera conjunto y separada a la vez.

$$sim(d_1, d_2) = \frac{\alpha \sum_w idf_w + \beta \sum_r idf_r}{\exp(\alpha \log \sum_w f_{w,1}^2 \sum_w f_{w,2}^2 + \beta \log \sum_r f_{r,1}^2 \sum_r f_{r,2}^2)} \quad (2)$$

$w$  recorre los vértices mapeados. Es decir, las palabras en común entre dos documentos (nodos de  $G_{12}$ ).

$r$  recorre los arcos (nodo-relación) entre los nodos comunes de los dos documentos (arcos de  $G_{12}$ ).

*idf*. La frecuencia inversa del documento (*inverse document frequency*) es una medida de la importancia del término en el conjunto del corpus. Consiste en calcular el logaritmo del inverso de la proporción de documentos del corpus que contienen el término. En este caso  $idf_w$ , corresponde a la medida calculada sobre los vértices e  $idf_r$ , la medida calculada sobre los arcos. De esta forma la frecuencia para un arco es una medida de tres elementos, las 2 etiquetas sobre los vértices y la etiqueta sobre la relación. Es así que por ejemplo,  $love \xrightarrow{agent} John$  es la unidad para contabilizar el *idf*.

El denominador de la expresión (2) corresponde al factor de normalización del modelo espacio vectorial, modificado para reflejar tanto los vértices (palabras) y arcos (relaciones).

$\alpha$ . Corresponden al peso dado a los vértices (palabras).

$\beta$ . Corresponde al peso dado a los arcos (relaciones).



Nodo-concepto	Nodo-concepto	Nodo-relación	Presencia de Relación (%)
1:0={ROOT}	1:2=asthma	-	100%
1:2=asthma	1:1=thymic	attr	100%
2:0={ROOT}	2:2=illustration	-	33%
2:0={ROOT}	2:9=due	-	67%
2:10=to	2:13=thymus	prep	67%
2:13=thymus	2:11=the	det	67%
2:13=thymus	2:12=enlarged	attr	67%
2:2=illustration	2:1=sagittal	attr	33%
2:2=illustration	2:3=of	attr	33%
2:3=of	2:5=narrowing	prep	33%
2:5=narrowing	2:4=tracheal	attr	33%
2:9=due	2:10=to	attr	67%
3:0={ROOT}	3:1=in	-	50%
3:0={ROOT}	3:7=is	-	25%
3:0={ROOT}	3:9=with	-	25%
3:11=extension	3:10=neck	attr	25%
3:1=in	3:2=this	in	50%
3:7=is	3:5=tracheal	attr	25%
3:7=is	3:6=narrowing	attr	25%
3:9=with	3:11=extension	prep	25%
4:0={ROOT}	4:3=thymus	-	100%
4:2==	4:1=th	attr	100%
4:3=thymus	4:2==	attr	100%
5:0={ROOT}	5:4=with	-	33%
5:0={ROOT}	5:7=from	-	67%
5:4=with	5:5=permission	prep	33%
5:7=from	5:8=reference	prep	67%
6:0={ROOT}	6:4=place	-	100%
6:10=perspective	6:11=of	attr	100%
6:10=perspective	6:14=of	attr	40%
6:10=perspective	6:6=the	det	100%
6:10=perspective	6:7=wrong	attr	100%
6:10=perspective	6:8=time	attr	100%
6:10=perspective	6:9=historical	attr	100%
6:11=of	6:13=relation	prep	100%
6:13=relation	6:12=the	det	100%
6:13=relation	6:14=of	attr	60%
6:14=of	6:17=gland	prep	100%
6:17=gland	6:15=the	det	100%
6:17=gland	6:16=thymus	attr	100%
6:20=radiology	6:19=pediatric	attr	100%
6:4=place	6:18=and	-	100%
6:4=place	6:20=radiology	attr	100%
6:4=place	6:2=the	det	100%
6:4=place	6:3=right	attr	100%
6:4=place	6:5=at	attr	100%
6:5=at	6:10=perspective	prep	100%

Cuadro No. 8. Ejemplo de salida del Analizador Sintáctico CIC-IPN (Grafo conceptual)

### 3.3.4.3 Colección de Prueba

Los resultados obtenidos se encuentran sobre la línea base del modelo espacio vectorial. Para las pruebas, se utilizó un sub-conjunto de la colección de imágenes anotadas de imageClef 2008. Uniendo la parte textual de los títulos de las imágenes y las anotaciones a las mismas, se obtuvo un total de 67.115 registros o documentos.

Se experimentó con un sub-conjunto de los primeros 1000 documentos (de 0000003 al 0001684) y con los primeros 15.630 documentos (de 0000003 al 0026687). Se utilizaron 9 de las 10 consultas (22 a la 30), se descarto la consulta 21 en virtud a su contenido se reduce a una sola palabra, y la estructura de grafos requiere mínimo 2 palabras.

Una vez obtenida la estructura de los grafos, con fines experimentales se redujo la dimensionalidad a través del procedimiento de Porter stemmer [73].

La muestra de 1000 documentos contiene 60 documentos relevantes para todas las preguntas y en la muestra de 15.603, 1.187 son documentos relevantes.

Los textos tanto de los documentos como de las consultas son muy cortos. Por ejemplo el texto para la consulta No 25 corresponde a: *Merkel cell carcinoma*” y el texto del primer documento (No 0000079) relevante para esta consulta aparece “*Eight single-level dynamic CT scans (A H) of the abdomen of a 32-year-old woman with abdominal pain. Scans were obtained during injection of 150 mL of nonionic contrast medium (iohexol) at 5.0 mL/sec. Scans show that the pancreas reaches peak enhancement before the liver. Effect of injection rate of contrast medium on pancreatic and hepatic helical CT*”.

Teniendo en cuenta nuestra hipótesis inicial de que para textos cortos el modelo de espacio vectorial puede mejorar al incluir información semántica adicional, efectivamente se obtuvo un mejor desempeño.

### 3.3.4.4 Medida de Desempeño

Para evaluar nuestro sistema contra la línea de base, se usó la precisión promedia media [56] (Mean Average Precision MAP). Dicha medida definida como en (3).

$$MAP = \frac{1}{Q} \sum_{q=1}^Q \frac{1}{m_q} \sum_{d=1}^{m_q} P_{qd} \quad (3)$$

Donde:

$q$  es el número total de consultas

$m_q$  es el número de documentos relevantes para una consulta.

$P_{qd}$  es la precisión en el conjunto de documentos calificados por el sistema, para la consulta  $q$ , mayores o iguales que el documento  $d$ .

La calificación (orden) fue definido de acuerdo a (2). Sin embargo, y dado que se encontraron bastantes documentos con calificaciones iguales, se aplicó la formula especificada en (4).

$$P_{qd} = \frac{P(R_{qd}^>) + P(R_{qd}^{\geq})}{2} \quad (4)$$

Donde:

$P$  es la precisión.

$R_{qd}^>$  Conjunto de documentos calificados mayores que  $d$ .

$R_{qd}^{\geq}$  Conjunto de documentos calificados mayores o iguales que  $d$ .

#### 3.3.4.5 Recuperación de información con los grafos conceptuales

Para cada documento y consulta, se construyó la representación semántica, variando  $\alpha$ ,  $\beta$  en (2). Después de considerar varias opciones incluyendo la posibilidad de no incluir el denominador, es decir considerar la medida de similitud consistente solamente el numerador. En la Figura No. 6 y Figura No. 7 se pueden apreciar los resultados obtenidos para la medida de evaluación de desempeño, para el parámetro  $\alpha = 1$  (coincidencia de palabras) y variando  $\beta$  (coincidencia de de arcos).

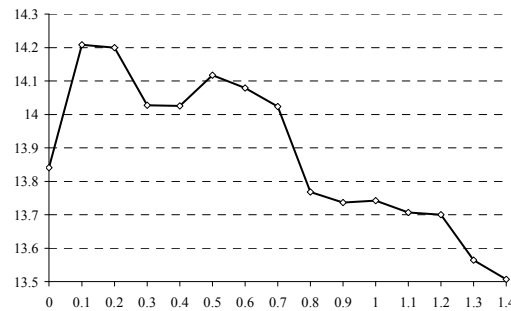


Figura No. 6. Promedio de Precisión media (Mean Average Precision MAP). vs.  $\beta$

La Figura No. 6 muestra los experimentos normalizando el denominador y en la Figura No. 7 se muestran los resultados de la formula (2) original tanto con denominador como sin denominador.

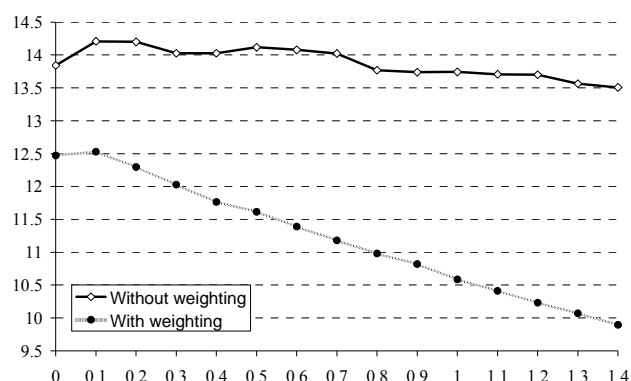


Figura No. 7. Promedio de Precisión Media, sin normalizar los pesos, como una función relativa al peso de las relaciones y los resultados normalizada.

La similitud para el modelo de espacio vectorial corresponde la función (2) con el parámetro  $\beta = 0$ , es decir sin efecto de las relaciones. Si se compara este resultado contra los resultados obtenidos incluyendo a  $\beta > 0$ , se puede observar que los resultados mejoran. Resultado que es mucho más evidente con la formula (2) normalizada. La mejora no es sustantiva pero es claramente observable para  $\beta$ , entre 0 y 0.7.

Con un  $\beta$  infinito, es decir, con  $\alpha = 0$ ,  $\beta = 1$ , es decir incluyendo únicamente las relaciones el resultado fue de 45%. Esto es mucho mejor si se compara contra el modelo utilizado de base, es decir el modelo vectorial que arrojó un 13.5% y el máximo resultado obtenido entre los concursantes de ImageClef2008 para recuperación de información que fue de 29%.

## 4. MÉTODO ESTADÍSTICO NO SUPERVISADO

Uno de los principales problemas del procesamiento de lenguaje natural y la lingüística computacional, es la dependencia de estas disciplinas de corpus especializados para un determinado idioma o dominio. Dicha dependencia implica un alto costo humano y de conocimiento, pues la tarea de obtener estos corpus o compilaciones de conocimiento (léxicos, tesauros, bases de conocimiento, ontologías, entre otros), generalmente depende de personal especializado y muchas horas de trabajo. Los métodos más eficaces, que resuelven muchas de las tareas importantes del procesamiento del lenguaje natural y de la lingüística computacional, están contruidos sobre aprendizaje supervisado basados en estos corpus de entrenamiento.

Lo anterior ha llevado a buscar alternativas que minimicen el costo de conocimiento especializado. Dentro estas alternativas se encuentran los métodos no supervisados basados en estadísticas y en particular probabilidades. Específicamente, en el área de la creación de los Grafos conceptuales no se ha implementado este tipo de métodos. Dentro de los trabajos más cercanos, a este objetivo, se encuentran los analizadores (“*parsing*”) probabilísticos para generar estructuras sintácticas. Por lo anterior a continuación se presentará un resumen de los trabajos desarrollados en el campo de la generación de estructuras sintácticas usando “*parsing*” probabilísticos con métodos no supervisados.

### 4.1 “PARSING” PROBABILÍSTICOS

La tarea de construir estructuras sintácticas utilizando métodos no supervisados ha sido poco explorada.

Específicamente, en el campo de la generación de estructuras sintácticas de dependencia, el trabajo de Klein y Manning en el 2004 [47] exponen una propuesta basada en valencias y hacen un análisis comparativo con los trabajos conocidos en el área. En este trabajo se analizan los tres mejores trabajos desarrollados en esta dirección, Carroll y Charniak [31], Paskin [57] y Yuret [21]. Dado que es importante para nuestra propuesta la discusión que se hace en este trabajo se incluyen algunas de los análisis realizados.

Los experimentos de Kelin y Manning [47] se basan en 4 colecciones dos en inglés (WSJ, porción de Penn Treebank y WSJ10 conjunto de sentencia que contienen 10 palabras o menos, después de renovar la puntuación), una en alemán y la otra en chino. El trabajo presenta resultados con los enlaces dirigidos y no dirigidos. Con el fin de comparar y dado que la propuesta que nosotros presentamos está dirigida al inglés, a grafos no dirigidos y no incluyen un límite de palabras, tomaremos únicamente los resultados que apuntan en esta dirección. El trabajo en mención para la colección WSJ, con grafos no dirigidos, muestra las precisiones obtenidas implementando diferentes modelos. Los modelos implementados corresponden a:

- Caroll y Charniak [31] con tres variaciones: inicializando con aleatoriedad (modelo *Randon*), disminuyendo la aleatoriedad al inicializar (modelo Charniak and Caroll) y enlazando las palabras adyacentes (modelo *Adjacent*).
- Paskin [57] y Yuret [21] (modelo Paskin) y
- Klein y Mannning [47] (modelo DMV)

Según [47] todos los sistemas que pretenden generar estructuras de dependencia a través de métodos no supervisados operan bajo algunos supuestos en común: a) La probabilidad de una estructura es el producto de las calificaciones de las dependencias enlazadas a dicha estructura; b) Las dependencias son pares ordenados (encabezado, dependiente) y varían en las condiciones impuestas sobre las calificaciones; y c) Los símbolos de la gramática no son simétricos.

#### 4.1.1 MODELO DE CAROLL Y CHARNIAK

El Modelo de Caroll y Charniak [31], presenta un aprendizaje probabilístico para gramáticas de dependencia y ejemplos de entrenamiento positivos. El corpus utilizado corresponde a las marcas de partes de la oración (*Pos Tagging*), sin incluir las palabras, más restricciones en las reglas. En el análisis que hacen en [47] aclaran que este trabajo no reporta precisiones que se puedan comparar, sugieren que los resultados obtenidos fueron muy pobres (Precisión para el modelo *Randon* = 41.7%) debido a la inicialización aleatoria de las probabilidades de los enlaces. Ya que, sobre una gramática en que las producciones son inicialmente uniformes, un símbolo  $X$  puede tener una probabilidad no nula que se extiende sobre los enlaces terminales de  $X$ .

En [47] rehacen este experimento y permiten que las iteraciones entre los datos y el modelo estructural rompan la simetría inicial (escogen una estructura de dependencia aleatoria) y obtienen una precisión de 44.7% (modelo Charniak y Caroll) sobre el Penn Treebank. Valor que es bajo si se compara con la opción de simplemente enlazar todas las palabras adyacentes a un encabezado izquierdo dado y bifurcando a la derecha (precisión para el modelo *Adjacent* = 53.2%).

#### 4.1.2 MODELO DE PASKIN Y YURET

El modelo de Paskin [57] y Yuret [21] como método no supervisado se basa en la escogencia aleatoria de un grafo  $G$  y un encabezado fijo que enlaza las diferentes palabras hasta alcanzar la estructura. El enlace de las palabras se realiza con base en la información mutua entre bigramas. En el modelo de Paskin utiliza un corpus etiquetado, con un analizador “*parsing*”, que se utiliza tanto para entrenamiento como para la prueba.

De acuerdo al análisis que se hace en [47], el trabajo de Paskin, presenta un alto costo computacional ya que es entrenado con 30 M de palabras, depende de la escogencia de una estructura arbitraria y la precisión obtenida está por debajo del nivel del azar. Igualmente en [47] analizan que el resultado del modelo se debe a que se enlazan palabras (bigramas) que tienen una alta información mutua, independientemente si están o no sintácticamente relacionadas. Es así que la información mutua entre dos sustantivos similares es mayor que entre una preposición y su objeto. Lo anterior lleva a concluir que los elementos léxicos son demasiado cargados semánticamente para representar las unidades de la estructura léxica. El modelo de Yuret, además se usar un corpus etiquetado y marcado como el de Paskin, usa 100 millones de palabras para encontrar su valor óptimo de 60% de precisión y 50% recall en un tiempo de  $O(n^5)$ . Es de aclarar que este último usa el corpus de Associated Press para sus experimentos.

#### 4.1.3 MODELO DE KLEIN Y MANNING (MODELO DMV)

El modelo propuesto por Klein y Manning [47] incluyen las valencias como nuevo elemento dentro del modelo. Proponen un modelo de tipo encabezado exterior (head-outward) e incluye clases de palabras y valencias. En este modelo se reporta una precisión de 54.4%. Es de aclarar que al igual que los anteriores, este modelo usa corpus etiquetado y marcado.

### 4.2 LAS ESTRUCTURAS DE DEPENDENCIA Y LOS GRAFOS CONCEPTUALES

Dado que la validación del trabajo desarrollado se hace con “*parsing*” no supervisados para estructuras de dependencia, vale la pena hacer un análisis comparativo entre las estructuras de dependencia y los grafos conceptuales. Como ya se había mencionado las estructuras de dependencia se basan en las gramáticas de dependencia propuestas por Tesnière [97] y su teoría es comparable con los átomos que cuentan con valencias que permiten la combinación entre palabras y la función que cada palabra ejerce. Una palabra puede ejercer una de dos funciones: principal o actuante y auxiliar o modificadora. La estructura de dependencia se base en el hecho de que toda palabra tiene un rector con excepción de la raíz, y no todas las palabras tienen un dependiente. Las aristas en una estructura de dependencia son etiquetadas

con roles gramaticales. En la estructura de dependencia de la frase “*From my infancy I was noted for the docility and humanity of my disposition*”<sup>3</sup>, la raíz corresponde al verbo “noted” (ver Figura No. 8), todas las palabras están incluidas dentro de la estructura, las etiquetas de los arcos corresponden a categorías gramaticales (*actor, theme*).

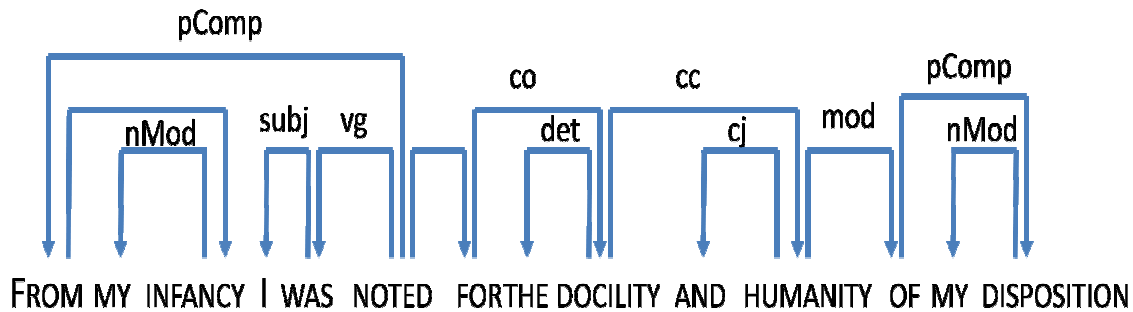


Figura No. 8. Estructura de dependencia para “From my infancy I was noted for the docility and humanity of my disposition”

En las estructuras de Grafos conceptuales, los nodos concepto están definidos por las palabras contenido (sustantivos, verbos, adjetivos y adverbios) y los nodos relación están definidos por algunas palabras funcionales (como preposiciones) y roles semánticos como (“*attr*”, “*agent*”, “*patient*”, “*actor*”, “*theme*”, entre otros). En la Figura No. 9 se puede apreciar el GC correspondiente a la frase “*From my infancy I was noted for the docility and humanity of my disposition*”. Es de observar que el tiempo verbal es un elemento externo, que palabras como “*my*” complementa al concepto “*disposition*”, “*the*” complementa a “*docility*” y aparecen relaciones como “*theme*”, “*actor*” y “*attr*” que son roles semánticos a diferencia de la estructura de dependencia donde aparecen roles gramaticales. Por último es de resaltar la relación entre “*noted*” y “*disposition*” es mucho más directa y semántica que en la estructura de dependencia que es indirecta y claramente sintáctica.

<sup>3</sup> Frase tomada del cuento de Edgar Allan Poe “The Black Cat”



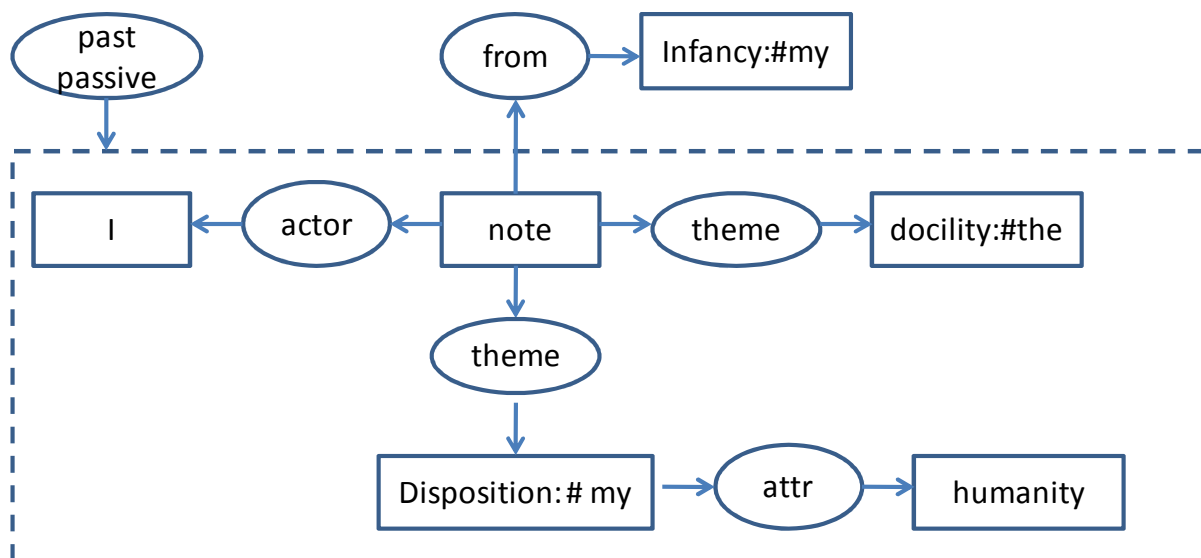


Figura No. 9. Estructura de GC para “From my infancy I was noted for the docility and humanity of my disposition”

Como ya se ha expuesto previamente, existen claras diferencias entre las estructuras de dependencia y los grafos conceptuales. Sin embargo, dado que no existen trabajos previos en este campo en los grafos conceptuales, es decir que se creen automáticamente los grafos conceptuales a través de métodos no supervisados, se recurre a comparar el trabajo propuesto con los desarrollados en el campo de las estructuras de dependencia, por contar estas con trabajos en esta dirección y tener algunas similitudes con los grafos conceptuales.

### 4.3 ASOCIACIONES POR CO-OCURRENCIAS

En el estudio presentado por Rapp [75], se precisan las relaciones sintagmáticas y paradigmáticas. Las frecuencias de co-ocurrencia de las palabras en un texto pueden ser aprovechadas para estimar asociaciones entre palabras. Aplicando estadísticas de primer orden sobre dichas frecuencias se pueden estimar relaciones semánticas y aplicando estadísticas de segundo orden se pueden encontrar relaciones de sinonimia y de colocación.

**Relaciones sintagmáticas.** Existe una relación sintagmática entre dos palabras si se presenta una co-ocurrencia en el lenguaje hablado o escrito más frecuentemente que lo esperado y si tienen diferentes relaciones gramaticales en el texto en que se produce. Las relaciones sintagmáticas son las que se establecen entre unidades o palabras coexistentes en un mismo enunciado, como la relación que existe entre sujeto y verbo.

**Relaciones paradigmáticas.** Dos palabras están relacionadas paradigmáticamente, cuando tienen una alta similitud semántica y se puede determinar de acuerdo a la proximidad. Por ejemplo la similitud semántica entre dos palabras como *red* y *blue* puede establecerse por la frecuencia de aparición de las palabras cercanas como *color*, *flowers*, *car*, etc. Se dice que

existe una relación entre dos palabras, si una de las dos palabras puede sustituirse por la otra en la sentencia sin afectar la gramática ya que generalmente las palabras ocupan el mismo lugar de la gramática y corresponden al mismo tipo gramatical. Este tipo de relación es muy usado para establecer la sinonimia, la antonimia y en general análisis de colocación.

#### 4.4 ESTRUCTURA SIMPLIFICADA DE GRAFOS CONCEPTUALES

Con base en las asociaciones basadas en las relaciones sintagmáticas, en este trabajo se establece una estructura simplificada de los grafos conceptuales. Dicha estructura aun cuando no incluye todos los elementos que hacen parte del estándar de los grafos conceptuales sí presenta las relaciones semánticas, los nodos concepto y la estructura base que permite que los nodos relación sean estimados a través de trabajos desarrollados para la búsqueda de roles semánticos como los presentados por Hensman [36,37,38]. La estructura simplificada es presentada a través del corpus marcado para el cuento de Edgar Allan Poe “*The black cat*” que se incluye en el anexo A de este trabajo.

La estructura propuesta tiene como característica fundamental incluir y relacionar los elementos semánticos de los grafos conceptuales, es decir los nodos-concepto. Por lo anterior la estructura simplificada incluye únicamente las palabras contenido o que pueden ejercer la función de palabra contenido (sustantivos, adjetivos, verbos y adverbios, pronombres), teniendo en cuenta que además de contener significado, forman la estructura principal de los Grafos conceptuales. Las palabras funcionales tales como preposiciones, artículos, conjunciones, determinantes, interrogantes, verbos modales y verbos auxiliares no tienen un significado sino que cumplen una función en el lenguaje. Existen para explicar o crear relaciones gramaticales o estructurales entre las palabras contenido. Por lo anterior, no se tienen en cuenta en la estructura simplificada de grafos conceptuales.

Por ejemplo en la frase “*From my infancy I was noted for the docility and humanity of my disposition*” extractada del cuento de Poe, la estructura simplificada se puede apreciar en la Figura No. 10. En dicha estructura a pesar de que no aparecen palabras *from, my, I, was, for, the, and* y *of*, la frase sigue transmitiendo un significado. La idea de su significado se mantiene.

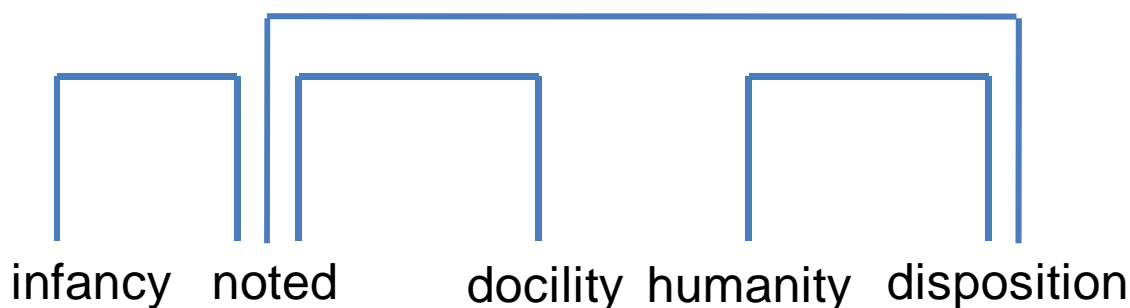


Figura No. 10. Estructura simplificada de GC para “*From my infancy I was noted for the docility and humanity of my disposition*”

Se observa que en la nueva estructura, GC simplificada, se establecen las relaciones entre los nodos concepto *noted* - *docility*, *noted* - *disposition* y *humanity* - *disposition* que permiten que con procesos ya establecidos para la determinación de roles semánticos como los presentados por Hensman [36,37,38], se puedan estimar los nodos relación que deben existir entre los nodos concepto. Esto es, que para aproximar la estructura de GC simplificado a la de GC completa solo queda por incluir los roles semántico.



Figura No. 11. Estructura simplificada de GC para “noted” y “docility” de la frase “From my infancy I was noted for the docility and humanity of my disposition”

Por ejemplo, si se le incluye el rol semántico (*theme*) entre la relación definida en *noted* y *docility* en la estructura de GC simplificada de la Figura No. 11, se obtendrá una aproximación más certera a la estructura de GC completa (ver Figura No. 12).

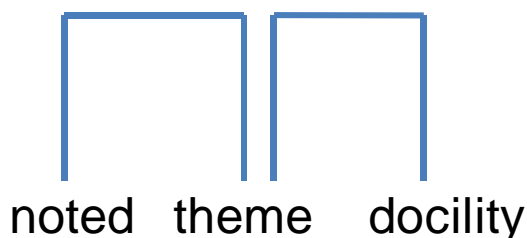


Figura No. 12. Estructura de GC para “noted” y “docility” de la frase “From my infancy I was noted for the docility and humanity of my disposition”. Donde theme se assume como un rol semántico

En la estructura simplificada de GC (Figura No. 13) de la frase “*When reason returned with the morning - when I had slept off the fumes of the night's debauch - I experienced a sentiment half of horror, half of remorse, for the crime of which I had been guilty; but it was, at best, a feeble and equivocal feeling, and the soul remained untouched*”, un poco más compleja que la anterior y también extractada del cuento de Poe, se utilizan 27 de 52 palabras y sin embargo es claro visualizar que dicha estructura si bien no contiene todos los elementos gramaticales y por ende no es una fiel copia del lenguaje natural, si transmite en gran parte la semántica de dicho lenguaje. En las estructuras como el modelo vectorial se eliminan las palabras que estadísticamente no aportan (“stop-words”), en este caso se eliminan aquellas que solo cumplen un papel funcional en el texto. Sin embargo en el modelo vectorial se cuenta con una bolsa de palabras, en este caso se obtiene más que una bolsa de palabras, una estructura con relaciones.

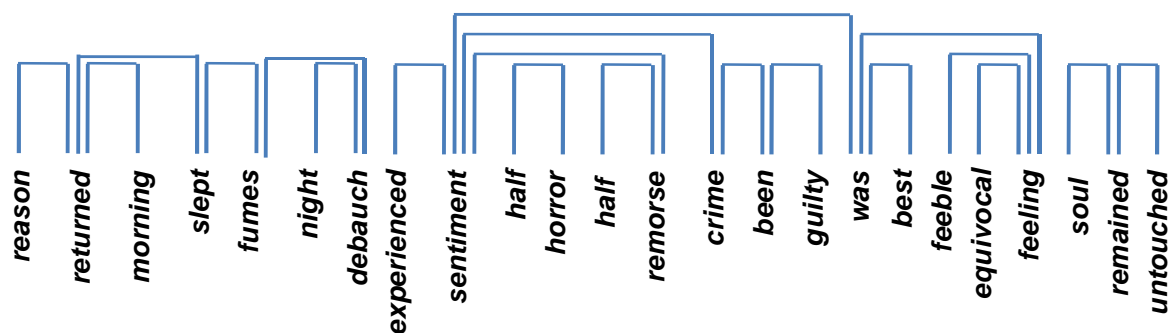


Figura No. 13. Estructura simplificada de GC para “When reason returned with the morning - when I had slept off the fumes of the night's debauch - I experienced a sentiment half of horror, half of remorse, for the crime of which I had been guilty; but it was, at best, a feeble and equivocal feeling, and the soul remained untouched”

Si se compara la estructura de dependencia (Figura No. 8) y la estructura simplificada de GC (Figura No. 10), es claro que a pesar de no incluir todas las palabras y definir nuevas relaciones, sigue conservando el sentido de la oración. La disminución del número de variables (14 a 5 palabras, es decir que se utilizan únicamente 36% de las palabras), por un lado no afecta en mayor grado la semántica de la oración y por el otro disminuye la complejidad si se quiere automáticamente conseguir dicha estructura. Aparecen relaciones más directas como por ejemplo entre *noted* y *docility*.

En la estructura propuesta las relaciones se hacen más directas, es así que no se hace necesario relacionar los verbos pertenecientes a dos cláusulas de una oración. Por ejemplo en la frase “*My tenderness of heart was even so conspicuous as to make me the jest of my companions*” en la estructura de dependencia (ver Figura No. 14) existe una relación directa entre *was* y *make*, mientras que no existe relación entre *conspicuous* y *make*.

En la estructura de GC simplificada (ver Figura No. 15 ) se presenta una relación entre *was* y *conspicuous* y entre *conspicuous* y *make*. Y no hay relación directa entre *was* y *make*.

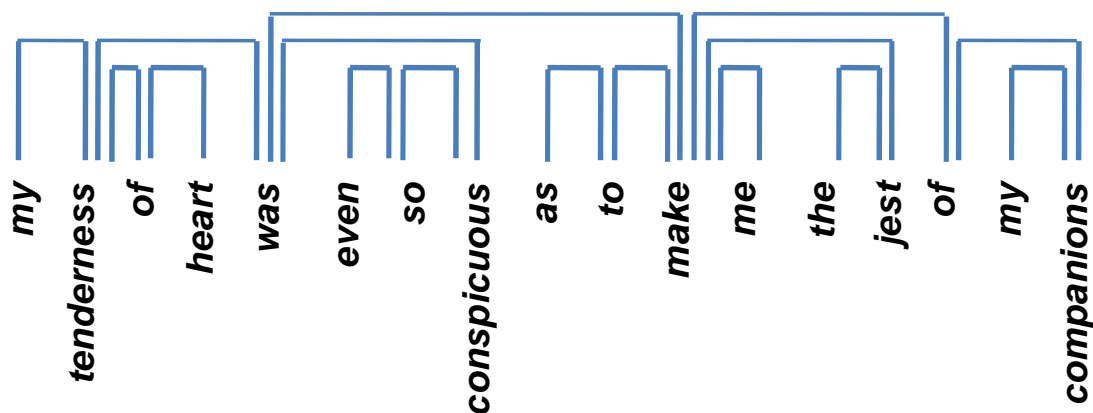


Figura No. 14. Estructura de dependencia para “My tenderness of heart was even so conspicuous as to make me the jest of my companions”

Otra característica que diferencia a la estructuras de GC (simplificada y completa) con la de dependencia, en frases donde una palabra *a* afecta a varias palabras (*b,c,d*) divididas por un operador, la relación en los GC siempre se da entre *ab*, *ac* y *ad*. Mientras que en la estructura de dependencia la relación no siempre se presenta en estos términos. En la frase “*succession of causes and effects*”. Se cuenta con sucesión de causas y una sucesión de efectos. En los GC (simplificados o completos) la palabra *succession* se relaciona con *causes* y *succession* con *effects*. En la relación de dependencia *succession* se relaciona con *causes* y *causes* con *effects* (ver Figura No. 16).

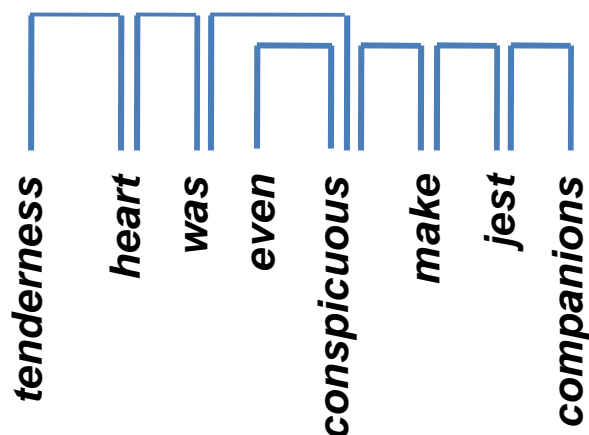


Figura No. 15. Estructura de GC simplificada para: “My tenderness of heart was even so conspicuous as to make me the jest of my companions”

En conclusión la nueva estructura de GC simplificada que se ha planteado, busca conservar la estructura de los GC y para ello tiene en cuenta que la estructura de los GC, no incluye todas las palabras funcionales y se opta por eliminar en su totalidad dichas palabras de la estructura sin perder totalmente la semántica de la oración.

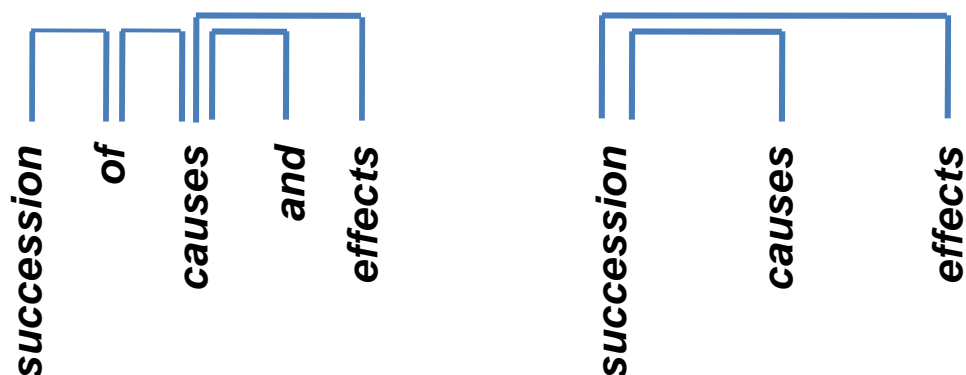


Figura No. 16. Estructura de dependencia (izquierda) vs GC simplificada (derecha) para “succession of causes and effects”

## 4.5 PARSING PARA GENERAR UNA ESTRUCTURA SIMPLIFICADA DE GC

Con base en la estructura simplificada de GC se propone un *parsing* basado en un método no supervisado y pobre en conocimiento. El *parsing* propuesto está basado en dos medidas básicas. La información mutua entre pares de palabra de tipo contenido (medida en un corpus grande y diferente a la colección de trabajo) y la distancia entre las palabras medida en la colección de trabajo.

El método utilizado por el *parsing* desarrollado incluye las ventajas que tienen las medidas sintagmáticas entre las palabras. Se aprovecha de esta medida estimando la asociación que puede existir entre dos palabras que co-ocurren en un texto y que no siempre tienen la misma relación gramatical.

Con el *parsing* desarrollado se puede construir una aproximación a los GCs sin necesidad de incluir un árbol sintáctico. El *parsing* a partir de estadísticas de primer orden logra obtener una estructura de los GCs simplificados. Con este *parsing* se presenta una solución no existente hasta el momento, se construye una aproximación a los GCs con un método no supervisado.

Debido a la inexistencia de este tipo de soluciones, el *parsing* presentado, se compara con los *parsing* desarrollados para árboles de dependencia, dado que estas estructuras cuentan con algunos elementos similares a los Grafos conceptuales como previamente se discutió en capítulos anteriores. Los *parsing* no supervisados para estructuras de dependencia utilizan grandes corpus de entrenamiento y por ende elevados costos computacionales. El *parsing* que en este trabajo se presenta, por el contrario no requiere de grandes corpus y horas de entrenamiento, con un corpus externo y con complejidad  $O(Nn^2)$  ( $N$  número de oraciones y  $n$  número de palabras contenido por oración) logra obtener una aproximación a la estructura de los grafos conceptuales, con una precisión mejor a la presentada por los *parsing* supervisados desarrollados para estructuras de dependencia y entrenados con grandes corpus y horas de entrenamiento.

El método no supervisado utilizado no requiere que de un texto controlado o especializado, sirve para cualquier tipo de texto y de vocabulario.

### 4.5.1 MÉTODO

El *parsing* para transformar de manera automática un texto  $T$  escrito en lenguaje natural a la estructura simplificada de GC se basa en un método de análisis no supervisado y pobre en conocimiento. Este método incluye una medida que utiliza por un lado, estadísticas de un gran corpus  $C$  cuyo único requisito es el que este en el mismo idioma que  $T$ , en este caso el inglés. Por el otro, utiliza la distancia entre las palabras del texto  $T$ . El método se prueba calculando la información mutua a partir de las frecuencias encontradas en el corpus de n-gramas de Google o consultando el motor de búsqueda Yandex. Es así que los resultados de

dichas frecuencias permitieron probar el método independientemente tanto con Google como con Yandex.

Para la validación del método se utilizó el texto correspondiente a una porción del cuento “El Gato Negro” de Edgar Allan Poe marcado de manera manual como estructuras simplificadas de GC e incluido en el anexo A de esta tesis.

El método se presenta como un algoritmo de optimización que incluye:

Sea  $w_i$  y  $w_j$  dos palabras que co-ocurren en un texto  $T$

Sea  $C$  un gran conjunto de datos (gran corpus) tal que  $T$  no pertenece  $C$

Sea  $f_{ij}$  la frecuencia de co-ocurrencia de la palabra  $w_i$  y la palabra  $w_j$  extractada de  $C$

Sea  $f_i$  y  $f_j$  la frecuencia de aparición de la palabra  $w_i$  y  $w_j$  respectivamente extractadas de  $C$

Sea  $M_{ij}$  la información mutua para la palabra  $w_i$  y  $w_j$ , donde  $\alpha$  corresponde a un parámetro que puede ser 0.

$$M_{ij} = \alpha + \log\left(\frac{f_{ij}}{f_i f_j}\right) \quad (5)$$

Sea  $d_{ij}$  la distancia entre la palabra  $w_i$  y la palabra  $w_j$  medida en el texto  $T$

Sea  $f(s)$  la función sobre la estructura que se desea maximizar. Y  $\beta$  un parámetro que puede ser 1 y que permite encontrar la mejor función.

$$Maxf(s) = \sum_i \sum_j \frac{M_{ij}}{d_{ij}^\beta} \quad (6)$$

Remplazando (5) en (6) resulta:

$$Maxf(s) = \sum_i \sum_j \frac{\alpha + \log\left(\frac{f_{ij}}{f_i f_j}\right)}{d_{ij}^\beta} \quad (7)$$

Bajo la restricción

- n-1 conexiones bajo un árbol proyectivo

Las restricciones definidas caracterizan un árbol proyectivo de n-1 conexiones, por lo que el algoritmo desarrollado, consiste en extraer por cada texto  $T$  el subconjunto de árboles proyectivos y sobre este elegir aquel que maximice la función.

### 4.5.2 ALGORITMO VORAZ

La optimización de la función planteada en (6) se logró a través de la implementación de un algoritmo voraz. Dicho algoritmo permite resolver el problema de optimización eligiendo en cada paso una solución óptima local esperando que esta sea la solución óptima general.

Dado  $M$  el conjunto de todos los pares de palabras del texto  $T$  y  $S$  el conjunto solución. En cada paso el algoritmo:

- Escoge la pareja de mayor valor  $x = (w_i w_j)$  (elemento más prometedor)
- Incluye  $x$  en  $S$  y valida que sea un árbol proyectivo
- Elimina  $x$  de  $M$
- Si el nuevo  $S$  es una solución factible pasa al siguiente candidato
- Si el nuevo  $S$  no es una solución factible elimina  $x$  de  $S$

### 4.5.3 RESULTADOS EXPERIMENTALES DEL PARSING

Después de realizar varios experimentos se encontraron el alfa y el beta que maximizan la Precisión. Como se puede apreciar en el Cuadro No. 9, para el motor Yandex, el valor de  $\alpha = 40$  y el de  $\beta = 0,2$  conduce a una *precision* = 64,3% y a un *recall* = 71,5%. Para el motor de Google los resultados correspondieron a un  $\alpha = 60$  y el de  $\beta = 0,8$  con una *precision* = 64,1% y a un *recall* = 71,3%.

Motor	Alfa	Beta	Precission	Recall
Yandex	40	0,2	0,6434635	0,71509434
Google	60	0,8	0,6417657	0,71320755

Cuadro No. 9. Valores de Alfa y beta para máxima precisión

Es de notar que la *precision* se mantiene muy cercana 0,64 frente a diferentes valores de alfa y beta, tal como se puede apreciar en la Figura No. 17.

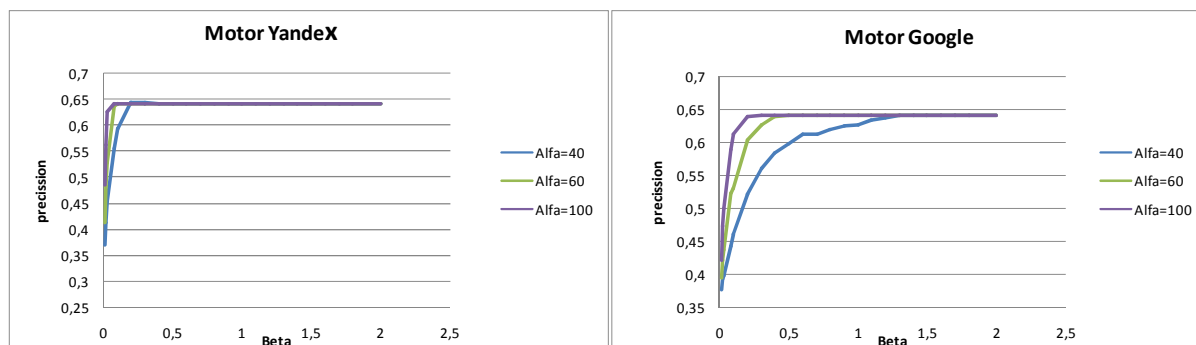


Figura No. 17. Comportamiento de la precisión para diferentes valores de alfa y beta par



El *parsing*, no supervisado pobre en conocimiento, resultado de esta investigación con un corpus de co-ocurrencias externo y con complejidad  $O(Nn^2)$  ( $N$  número de oraciones y  $n$  número de palabras de contenido por oración) logra obtener una aproximación a la estructura de los grafos conceptuales, con una precisión mejor que los *parsing* no supervisados desarrollados para estructuras de dependencias. Como ampliamente se explicó en 4.1, se toma como punto de comparación los *parsing* para estructuras de dependencia dadas algunas similitudes con los grafos conceptuales y debido a la falta de este tipo de trabajos para GCs.

Structure	Modelo	% precision
Dependency	Paskin 01	39,7
Dependency	Random	41,7
Dependency	Charmiak an Carroll	44,7
Dependency	Adjacent	53,2
Dependency	Yuret	60,0
Dependency	DMV	54,4
GC	Parsing CG	64,3

Cuadro No. 10. Resultados para diferentes Modelos de Parsing no supervisado.

En el Cuadro No. 10, se puede observar que el parsing co-ocurrence desarrollado por esta investigación, presenta una mejora con relación a los previamente desarrollados para las estructuras de dependencias que previamente fueron explicadas en la sección 4.1. El resultado es satisfactorio si se tiene en cuenta que si bien todos se basan en métodos no supervisados, el parsing co-ocurrence desarrollado, a diferencia de los demás, está basado en un método pobre en conocimiento que no requiere de grandes corpus ni tiempos de entrenamiento y que conduce tan a solo una complejidad de  $O(Nn^2)$ .

## 5. RESULTADOS DE LA INVESTIGACIÓN

### 5.1 CONCLUSIONES Y TRABAJO FUTURO

Esta investigación presenta dos grandes conclusiones alrededor de búsqueda de alternativas para la construcción automática de los GCs a partir de texto:

- ✓ Se ha demostrado que se puede transformar el texto en GC a partir de una gramática. Con la ventaja de que dicha gramática se construye de manera automática utilizando un recurso léxico desarrollado para otros fines distintos a la construcción de la gramática.
- ✓ Se ha demostrado que se puede transformar el texto en una estructura de GCs simplificados a partir de un método no supervisado, pobre en conocimiento, sin grandes corpus de entrenamiento, con complejidad de  $O(Nn^2)$  y que presenta mejores resultados en comparación a los similares.

### 5.2 APORTES

- ✓ El trabajo aporta una metodología para obtener una gramática a partir de un recurso léxico construido con fines distintos a los que compete en este trabajo. La metodología incluye los aspectos más importantes que se deben tener en cuenta a la hora de construir una gramática para los grafos conceptuales. Muestra como pasar cada marco definido en el recurso léxico VerbNet a una o varias reglas gramaticales. Es de resaltar que dado que la construcción de la gramática se hizo de manera automática y a partir de un recurso léxico ajeno al objeto del trabajo se diferencia de las demás construcciones por su costo en conocimiento muy bajo ya que no se requirió la ayuda del experto.
- ✓ Como resultado del conjunto de reglas gramaticales se construye una gramática para generar los Grafos conceptuales a partir de texto expresado en idioma inglés. La característica principal de la gramática es que incluye no solo los elementos estructurales y sintácticos de los grafos sino los semánticos. Hasta el momento no existía una gramática especializada para tal fin.
- ✓ Los Grafos conceptuales generados con la anterior gramática son lo suficiente buenos ya que mejoran la tarea de recuperación de información. Dado que no existen colecciones de

prueba que permitan validar la estructura encontrada, se validó la estructura resolviendo la tarea de recuperación de información y obteniendo resultados mejores que los obtenidos a través del modelo estándar de espacio vectorial.

- ✓ Se logró construir una aproximación a los Grafos conceptuales sin necesidad de pasar por una estructura sintáctica. Hasta el momento la construcción de los Grafos conceptuales han partido de un árbol sintáctico que es posteriormente reformado y modificado. En este trabajo se logró construir los Grafos conceptuales sin incluir este árbol y sin necesidad de incluir procedimientos que modificaran el árbol. Un analizador gramatical y una gramática construida para este fin, transforma el texto en idioma inglés a Grafos Conceptuales.
- ✓ A través de un método no supervisado y pobre en conocimiento se creó un analizador “parsing” que permite obtener una estructura simplificada de los grafos conceptuales. Hasta el momento no existía un analizador “parsing” especializado para obtener la estructura de Grafos.
- ✓ A partir de un corpus de estadísticas de co-ocurrencias, en este trabajo se logró construir un “parsing”. Hasta el momento los parsing logrados con métodos no supervisados y para estructuras diferentes a los Grafos conceptuales, deben utilizar gigantescos corpus de entrenamiento y muchas horas de trabajo. El “parsing” resultado de esta investigación con un corpus de co-ocurrencias externo y con complejidad  $O(Nn^2)$  ( $N$  número de oraciones y  $n$  número de palabras de contenido por oración) logra obtener una aproximación a la estructura de los grafos conceptuales, con una precisión mejorada de los parsing similares. Por último, el método no supervisado que se utilizó para crear el “parsing” sirve para cualquier texto, no requiere ser controlado.
- ✓ En este trabajo se ha obtenido una estructura simplificada de los Grafos conceptuales, que aun cuando no incluye todos los detalles de los Grafos, sí captura aspectos semánticos de los grafos.
- ✓ Por último el trabajo aporta un corpus marcado (“*The Black Cat*” - Edgar Allan Poe) con la estructura anteriormente definida.

### 5.3 PRODUCTOS ENTREGABLES

- ✓ Una gramática para la creación de Grafos conceptuales a partir de texto en inglés.
- ✓ Una estructura simplificada de Grafos Conceptuales
- ✓ Un “parsing” no supervisado y pobre en conocimiento basado en estadísticas de Co-ocurrencia.

- ✓ Un corpus marcado de “*The Black Cat*” (Edgar Allan Poe) con una simplificada estructura.

## 5.4 PUBLICACIONES

En proceso de publicación

- ✓ S. Ordoñez, A. Gelbukh. Estructuras computacionales en el procesamiento del lenguaje natural. Revista de Ingeniería de la Universidad Distrital F.J.C

Capítulo de libro

- ✓ S. Ordoñez, A. Gelbukh. *Generación de grafos conceptuales*. Avances en sistemas inteligentes en México. Edit. Miguel González Mendoza – Oscar Herrera Alcantarra. Sociedad Mexica de Inteligencia Artificial. ISBN: 978-607-95367-2-5 pp. 139-150, 2010.

Artículo en revista

- ✓ S. Ordoñez, A. Gelbukh. *Automatic Semantic Analysis of Short Natural Language Sentences via Supervised Learning*. Conference on Computing CORE-2010, Mexico City, Mexico, May 2010. Advances in Computer Science and Engineering. J. Research in Computing Science (latindex).
- ✓ S. Ordoñez. *Aprendizaje de Reglas Encadenas para la Creación de Grafos Conceptuales*. Research journal on Computer science and computer engineering with applications *POLIBITS*. Special Section. Information Retrieval and Natural Language Processing. PP. 39-45 ISSN 1870-9044 (latindex)

Lectures Notes

- ✓ S. Ordoñez, A. Gelbukh. Information Retrieval with a Simplified Conceptual Graph-Like Representation. In Proc. 9th Mexican International Conference on Artificial Intelligence Nov 2010. Lecture Notes in Artificial Intelligence. pp. 92 – 104, 2010

## *Referencias*

- 1 Abdulrub S., Polovina S. y Hill, R. Implementing Interoperability through an Ontology Importer for Amine. Conceptual Structures Tools an the Web - Third Conceptual Structures Tool Interoperability Workshop, 2008.
- 2 Amghar, T., Battistelli, D., Charnois, T. Reasoning on aspectual-temporal information in French within conceptual graphs. Tools with Artificial Intelligence, 2002. (ICTAI 2002). Proceedings. 14th IEEE International Conference on, 2002, pp. 315-322.
- 3 Aravind J.; Levy L., and Takahashi M. «Tree adjunct grammars». Journal of Computer and System Sciences, 10(1). pp. 136-162. 1975.
- 4 Ausube I, D., Novak, J. Hanesian, H. Psicología Educacional: Una visión cognitiva; Halt, Reinhart and Winston, New York, 1978
- 5 Baader F., Nutt W. Basic Description Logics. The Description Logic Handbook, Cambridge University Press, pp 47-100, 2002.
- 6 Badia, A. y Kantardzic, M. Graph building as a mining activity: finding links in the small. LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery, PP. 17-24. ACM, 2005.
- 7 Barceló, G., Cendejas, E., Bolshakov, I. y Sidorov G. Ambigüedad en nombres hispanos. Revista Signos. Estudios de Lingüística 42 (70). PP. 153-169. 2009.
- 8 Barrière, C., Barrière, N. C. From a Children's First Dictionary to a Lexical Knowledge Base of Conceptual Graphs. Tesis presentada en Ecole Polytechnique de Montreal, para optar el título de Doctor en Philosophy in the School of Computing Science, 1997.
- 9 Barski C. The enigmatic art of knowledge representation. [www.lisperati.com/tellstuff/index.html](http://www.lisperati.com/tellstuff/index.html), visitada en marzo 2009.
- 10 Brachman R., Schmolze J., An overview of the KL-ONE Knowledge Representation System, Cognitive Science, Vol. 9, No. 2, pp. 171-216. 1985.
- 11 Campbell, K. E., Musen M. A. Representation of clinical data using SNOMED III and conceptual graphs. Proceedings of the Annual Symposium on Computer Application in Medical Care. 1992: pp. 354-358
- 12 Castro-Sánchez, N. A., y Sidorov, G. Analysis of Definitions of Verbs in an Explanatory Dictionary for Automatic Extraction of Actants based on Detection of Patterns. Lecture Notes in Computer Science, N 6177. PP. 233-239. 2010.

- 13 Chang C.S., Chen A.L.P., Supporting Conceptual and Neighborhood Queries on WWW, IEEE Transactions on Systems, Man, and Cybernetics, Part C: application and reviews (EI,SCI), Vol.28, No.2, 1998, pp.300-308
- 14 Chomsky, N. Syntactic structures. La Haya, Mouton. 1957
- 15 Corby O., Dieng-Kuntz R., Faron-Zucker C. Querying the Semantic Web with the CORESE search engine. Lopez de Mantaras R., L. Saitta L., (ed.). Proc. of the 16th European Conference on Artificial Intelligence (ECAI'2004), subconference PAIS'2004, Valencia, 2004, pp. 705-709.
- 16 Cruz, M. H. Generador de los grafos conceptuales a partir del texto en español. Instituto Politécnico Nacional . Centro de Investigación en computación, 2007.
- 17 Deerwester S., Dumais S. T., Furnas G. W., Landauer T. K. y Harshman R. Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science. Vol. 41. No. 6. PP. 391-407. 1990.
- 18 Delugach H. S. Towards. Conceptual Structures Interoperability Using Common Logic Computer. Science Department Univ. of Alabama in Huntsville. Third Conceptual Structures Tool Interoperability Workshop. 2008.
- 19 Delugach H., A Conceptual Graph Editor, CharGer Manual v3.5b1 User's Guide, 2005.
- 20 Denecke K., I., Kohlhof, J. B. Use Of Mul-tiaxial Indexing for Information Extraction From Medical Texts Integrating Biomedical Information: From E-cell to E-patient. Proceedings of the European Federation for Medical Informatics Special Topic Conference 2006. , IOS Press, April, 2006.
- 21 Deniz Yuret. 1998. Discovery of Linguistic Relations Using Lexical Attraction. Ph.D. thesis, MIT.
- 22 Farkas, J., Improving the classification accuracy of automatic text processing systems using context vectors and backpropagation algorithms., Electrical and Computer Engineering, 1996. Canadian Conference. Vol.2, May 1996., pp.696-699
- 23 Fernández, M.; de la Clergerie, E. & Vilares, M. Mining conceptual graphs for knowledge acquisition. iNEWS '08: Proceeding of the 2nd ACM workshop on Improving non english web searching, ACM, 25-32, 2008.
- 24 Fikes R.,Farquhar A., Rice J., Tools for Assembling Modular Ontologies in Ontolingua. Knowledge Systems, AI Laboratory, 1997.
- 25 Fillmore, c. J."The Case for Case" (1968). In Bach and Harms (Ed.): *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, 1-88.
- 26 Gao J., Nie J.-Y, Wu G., Cao G. Dependence Language Model for Information Retrieval, Microsoft Research, Asia, Brooks Cole Publishing Co., Pacific Grove, 1999.
- 27 Gelbukh A., Sidorov G. Procesamiento automático del español con enfoque en recursos léxicos grandes. Centro de Investigación en Computación, Instituto Politécnico Nacional, México, 2006.
- 28 Gelbukh A., Sidorov, G., Galicia, S., Bolshakov, I. Environment for Development of a Natural Language Syntactic Analyzer. Acta Academia, Moldova, pp. 206-213, 2002.

- 29 Genest, D., Salvat, E. A platform allowing typed nested graphs: How CoGITO became CoGITaNT in Conceptual Structures: Theory, Tools and Applications. Lecture Notes in Computer Science. Editor Mugnier, M., hein, M., Publisher Springer Berlin – Heidelberg. 1.998 pp. 154-161.
- 30 Ginsberg M., Knowledge Interchange Format: The KIF of Death, journal AI Magazine. Vol. 12, pp 57-63, 1991.
- 31 Glenn Carroll and Eugene Charniak. 1992. Two experiments on learning probabilistic dependency grammars from corpora. In Carl Weir, Stephen Abney, Ralph Grishman, and Ralph Weischedel, editors, Working Notes of the Workshop Statistically-Based NLP Techniques, pages 1–13. AAAI Press, Menlo Park, CA.
- 32 Guarino, N.; Masolo, C., Vetere, G. Onto-Seek: Content-Based Access to the Web. IEEE Intelligent Systems, IEEE Educational Activities Department, 1999.
- 33 Harabagiu S., Maiorano S. Acquisition of linguistic patterns for knowledge-based information extraction. Proceedings of LREC 2000, Athens, 2000.
- 34 Helbig H.. Knowledge Representation and the Semantics of Natural Language. Lecture Notes in Computer Science Publisher Springer-Verlag Berlin Heidelberg 2006.
- 35 Henderson, J., Merlo, P., Petroff, I. y Schneider, G. Using NLP to efficiently visualize text collections with SOMs. Proceedings. 13th International Workshop on Database and Expert Systems Applications, 2002. Proceedings. 13th International Workshop, 2002.
- 36 Hensman S.. Construction of Conceptual Graph representation of texts. Department of Computer Science, University College Dublin. Belfield, Dublin 4. Proceedings of Student Research Workshop at HLT-NAACL, 2004.
- 37 Hensman, S., Dunnion, J. Automatically building conceptual graphs using VerbNet and WordNet. Proceedings of the 2004 international Symposium on information and Communication Technologies. ACM International Conference Proceeding Series, vol. 90. Trinity College Dublin. Las Vegas, June 2004, pp. 115-120.
- 38 Hensman, S., Dunnion, J., Constructing conceptual graphs using linguistic resources. In Proceedings of the 4th WSEAS international Conference on Telecommunications and informatics. Husak M. and Mastorakis N., (Ed). World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, Mar. 2005, pp. 1-6.
- 39 Hernández Cruz, M. Generador de los grafos conceptuales a partir del texto en español. Tesis de Maestría. Instituto Politécnico Nacional. Centro de Investigación en computación. 2007.
- 40 ISO/IEC 14481: Information Tecnology (IT). Conceptual Schema Modeling Facilities CSMF., 1998
- 41 Kabbaj, A., Janta-Polczynski, M., From PROLOG++ to PROLOG+CG: A CG Object-Oriented Logic Programming Language in Conceptual Structures: Logical, Linguistic, and Computational Issues, Lecture Notes in Computer Science. Editor Bernhard G. and Guy M., publisher Springer Berlin Heidelberg., Vol. 1867 2000. pp. 540-554.
- 42 Kamaruddin, S.S.; Bakar, A.A.; Hamdan, A.R.; Nor, F.M., Conceptual graph formalism for financial text representation, Information Technology, 2008. ITSIm 2008. International Symposium , Vol.3, Aug. 2008., pp.1-6.

- 43 Kifer M, Lausen G, Wu J. Logical Foundations of Object-Oriented and Frame-Based Languages. Journal of ACM, May 1995.
- 44 Kimura, M., Saito, K., Ueda, N. Multinomial PCA for extracting major latent topics from document streams. Neural Networks 2005, IJCNN '05. Proceedings. 2005 IEEE International Joint Conference, 2005.
- 45 Kipper K., Korhonen A., Ryant N., Palmer M.. Extending VerbNet with Novel Verb Classes. Fifth International Conference on Language Resources and Evaluation (LREC 2006). Genoa, Italy. June, 2006
- 46 Klabbankoh B. y Pinngern Q. Applied Genetic Algorithms in Information Retrieval. Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, 2000.
- 47 Klein D. and Manning C.D. Corpus-Based Induction of Syntactic Structure: Models of Dependency and Constituency. In Proceedings of the 42nd Annual Meeting of the ACL, pp 479-486, 2004.
- 48 Kovacs, L., Baksa-Varga, E., Dependency-based mapping between symbolic language and Extended Conceptual Graph, Intelligent Systems and Informatics, 2008. SISY 2008. 6th International Symposium, Sept. 2008, pp.1-6.
- 49 Krauskopf T., Miller. J, Resnick P., Treese W. PICS Label Distribution Label Syntax and Communication Protocols, Version 1.1, W3C Recommendation, 1996.
- 50 Lafferty J., Sleator D., Temperley D., Grammatical Trigrams: A Probabilistic Model of Link Grammar. Proceedings of the AAAI Conference on Probabilistic Approaches to Natural Language, October, 1992
- 51 Lassila O., Swick R. y World Wide and Web Consortium. Resource Description Frame-work (RDF) Model and Syntax Specification. Recommendation, World Wide Web W3C. Consortium, Cambridge (MA), 1999.
- 52 Last M., Maimon O. A compact and Accurate Model for Classification, IEEE Transactions on Knowledge and Data Engineering. Vol. 16. No. 2. PP. 203-215. 2004.
- 53 Losada, D. y Barreiro A. Rating the impact of logical representations on retrieval performance. Database and Expert Systems Applications Proceedings. 12th International Workshop, PP. 247-253. 2001.
- 54 Maisonnasse L., Gaussier E., J. C. Multiplying Concept Sources for Graph Modeling. This paper presents the LIG contribution to the CLEF 2007 medical retrieval task (i.e. ImageCLEFmed). 2007.
- 55 Manning C., Schütze H., Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999.
- 56 Manning, C. D., Raghavan, P., and Schütze, H. Introduction to Information Retrieval, Cambridge University Press. 2008.
- 57 Mark A. Paskin. 2002. Grammatical bigrams. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, Advances in Neural Information Processing Systems 14, Cambridge, MA. MIT Press.
- 58 Maron, M.E. y Kuhns, J.L. On relevance, probabilistic indexing and information retrieval. Journal of the ACM, 1960.



- 59 Martin P. A., Eklund P. Large-scale cooperatively-built heterogeneous KBs. Proceedings of ICCS, 9th International Conference on Conceptual Structures, Stanford University, California, LNAI 2120 Springer Verlag USA, Jul. 2001 pp. 231-244.
- 60 Martin, P. Knowledge Representation in CGLF, CGIF, KIF, Frame-CG and Formalized-English, Conceptual Structures: Integration and Interfaces. Lecture Notes in Computer Science. Springer, Vol. 2393, 2002., pp. 77-91.
- 61 Martin, P.A. and Eklund, P. WebKB. Proceedings of the 7th International Conference on Conceptual Structures, 1999.
- 62 McGuinness D. L., Fikes R., Hendler J., Stein L. A., "DAML+OIL: An Ontology Language for the Semantic Web," IEEE Intelligent Systems, vol. 17, no. 5, pp. 72-80, Sep./Oct. 2002.
- 63 Medical Image Retrieval Challenge Evaluation P., <http://ir.ohsu.edu/image>.
- 64 Miller G., A., (1995). WordNet: A Lexical Database for English Communications of the ACM Vol. 38, No. 11, 1995. pp. 39-41.
- 65 Mineau G. W., Moulin B., Conceptual Grpahs for Knowledge Representation, First International Conference on Conceptual Structures, ICCS'93., Lectures Notes in Artificial Intelligence. Sowa J.F. (Ed.), Springer-Verland., 1993.
- 66 Mineau, G. W., Stumme, G, Wille, R. Conceptual Structures Represented by Conceptual Graphs and Formal Concept Analysis, International Conference on Conceptual Structures, 1999.
- 67 Montes y Gómez M., Minería de Texto Empleando la Semejanza entre Estructuras semánticas. Tesis Doctoral. Instituto Politécnico Nacional Centro de Investigación en Computación. Laboratorio de Lenguaje Natural y Procesamiento de Texto, México 2002.
- 68 Montes-y-Gómez M. Minería de texto: Un nuevo reto computacional. 3er Taller Internacional de Minería de Datos MINDAT-2001, Universidad Panamericana, Ciudad de México, Octubre 2001.
- 69 Noy N., Grosso W, Musen A. Knowledge acquisition Interfaces for Domain Experts: An Empirical Evaluation of Protege-2000. Twelfth International Conference on Software Engineering and Knowledge Engineering (SEKE 2000), Chicago, IL. 2000.
- 70 Peltonen, J., Sinkkonen J. y Kaski, S. Discriminative clustering of text documents. Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference. Vol.4. PP.1956-1960. 2002.
- 71 Pfeiffer D, Chavez N., Heather D., Pfeiffer J. CPE Design Considering Interoperability, New Mexico State University, 2004.
- 72 Polovina, S. CARE: A Practical Conceptual Graphs Software Tool.,2007.
- 73 Porter M. An algorithm for suffix stripping, Program, Vol. 14, no. 3, pp 130–137, 1980, Available on <http://tartaus.org/~martin/PorterStemmer/>.
- 74 Quijano B. H., Espinosa M. A. Grafitos: un editor didáctico para grafos conceptuales. programa Ingeniería de Sistemas y Computación, de la Universidad Tecnológica de Pereira., 2004.

- 75 Rapp R.. The Computation of Word Associations: Comparing Syntagmatic and Paradigmatic Approaches
- 76 Rassinoux, A. M., Baud, R. H., Lovis, C., Wagner, J. C., Scherrer, J. R. Tuning Up Conceptual Graph Representation for Multilingual Natural Language Processing in Medicine. 6th International Conference on Conceptual Structures, ICCS'98. Conceptual Structures: Theory, Tools, and Applications. Lecture Notes in Computer Science Montpellier. Springer. Vol. 1453, Aug. 1998.
- 77 Rassinoux, A. M., Baud, R. H., Scherrer, J. R. A Multilingual Analyser of Medical Texts. Second International Conference ICCS'94. Conceptual Structures: Theory, Tools, and Applications. Lecture Notes in Computer Science, Springer, USA, Vol. 835, Aug. 1994 pp. 84-96.
- 78 Rijsbergen C.J. van, Robertson S.E. y Porter M.F. New models in probabilistic information retrieval. London: British Library. (British Library Research and Development Report, no. 5587). 1980.
- 79 Rijsbergen Van C.J. Information Retrieval. Department of Computing Science, University of Glasgow Second edition. 1.979.
- 80 Roberts R, Goldstein I.. The FRL manual. Memo 409, Massachusetts Institute of Technology Artificial Intelligence Laboratory, 1977.
- 81 Salton, G. y Lesk, M. E. The SMART automatic document retrieval systems and illustration Common. ACM, 1965.
- 82 Schenker A., Bunke H., M. L. y Kandel, A. A Graph-Based framework for Web document mining. Lecture Notes in Computer Science Publisher Springer Berlin Heidelberg. Vol. 3163, PP. 401-412. 2004.
- 83 Schenker A., Bunke Horst, M. L. A. K. Graph-theoretic techniques for Web content mining. World Scientific Publishing, 2005.
- 84 Schröder, M. Knowledge based analysis of radiology reports using conceptual graphs Conceptual Structures: Theory and Implementation, Lecture Notes in Computer Science Springer.. Vol. 754. 1993. pp. 293-302.
- 85 Shannon, C. A mathematical theory of communication. The Bell System Technical Journal, vol. 27, pp. 379-423, October 1948.
- 86 Shehata, S., Karray, F. y Kamel, M. Enhancing Text Retrieval Performance using Conceptual Ontological Graph. Data Mining Workshops, International Conference, pp. 39-44, Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), 2006.
- 87 Shin S-J The Logical Status of Diagrams Cambridge University Press 1994.
- 88 Sirin E., Hendler J., Parsia B. Semi-automatic Composition of Web Services using Semantic Descriptions. Web Services: Modeling, Architecture and Infrastructure workshop in ICEIS 2003. PP. 17-24. 2002.
- 89 Sleator, D. D., Temperley, D. Parsing English with a link grammar. Third International Workshop on Parsing Technologies, 1993.
- 90 Southey, F., Linders, J. G., Notio. A - Java API for developing CG tools in Conceptual Structures: Standards and Practices, Springer, 1999, pp. 262-71

- 91 Sowa J., F Conceptual Graph Standard. Committee on Information Interchange and Interpretation, 2000.
- 92 Sowa J., F, Common Logic. A Framework for a Family Of Logic-Based Languages, 2008
- 93 Sowa J., F. Conceptual Structures, Information Processing in Mind and Machine, The Systems Programming Series Addison-Wesley Publishing Company, 1984.
- 94 Sowa J., F., Way E. C. Implementing a semantic interpreter using conceptual graphs. IBM Journal of Research and Development, 30(1):57–69, 1986.
- 95 Sowa J.F., Way, E. C. Implementing a semantic interpreter using conceptual graphs. IBM J. Res. Dev. 30, 1 Jan. 1986, pp. 57-69
- 96 Sowa, J. F., Conceptual Graphs., Handbook of Knowledge Representation., Foundations of Artificial Intelligence., Vol. 3, Harmelen F.,V., Lifschitz V., Porter B., (Ed.), Elsevier, 2008, pp 213-237.
- 97 Tesnière, A. L. Elements de syntaxe structurale. Libraire C. Klincksieck, 1959.
- 98 Varlamis I., Vazirgiannis M., Halkidi M., Nguyen B. Thesus, A Closer View on Web Content Management Enhanced with Link Semantics. IEEE Transactions on Knowledge and Data Engineering. Vol. 16. No. 6. PP. 685-700. 2004.
- 99 W3. RDF/XML Syntax Specification. W3C Recommendation., Dave Beckett, (ed). February 10, 2004.
- 100 Wermelinger, M. GET: An Implementation of Conceptual Graphs. Centro de Inteligencia Artificial / UNINOVA Portugal., 1991
- 101 Williams, R. A., Computational Effective Document Semantic Representation. Digital EcoSystems and Technologies Conference, 2007. DEST '07. Inaugural IEEE-IES., IEEE IES.,Feb. 2007., pp. 410-415.
- 102 Zhang, L. and Yu, Y. Learning to Generate CGs from Domain Specific Sentences Seidel Mean Value Coordinates for Arbitrary Spherical Polygons and Polyhedra in R 3, Springer Verlag, PP. 44-57, 2001.



# *Anexo A*



1. most wild most homely  
 narrative i pen i neither  
 expect solicit belief  
 most|wild 1  
 most|most  
 most|homely  
 most|narrative  
 most|i  
 most|pen  
 most|i  
 most|neither  
 most|expect  
 most|solicit  
 most|belief  
 wild|most  
 wild|homely  
 wild|narrative  
 wild|i  
 wild|pen  
 wild|i  
 wild|neither  
 wild|expect  
 wild|solicit  
 wild|belief  
 most|homely 1  
 most|narrative  
 most|i  
 most|pen  
 most|i  
 most|neither  
 most|expect  
 most|solicit  
 most|belief  
 homely|narrative 1  
 homely|i  
 homely|pen  
 homely|i  
 homely|neither  
 homely|expect  
 homely|solicit  
 homely|belief  
 narrative|i 1  
 narrative|pen  
 narrative|i  
 narrative|neither  
 narrative|expect

narrative|solicit  
 narrative|belief  
 i|pen 1  
 i|i  
 i|neither 1  
 i|expect  
 i|solicit  
 i|belief  
 pen|i  
 pen|neither  
 pen|expect  
 pen|solicit  
 pen|belief  
 i|neither  
 i|expect  
 i|solicit  
 i|belief  
 neither|expect 1  
 neither|solicit  
 neither|belief  
 expect|solicit 1  
 expect|belief 1  
 solicit|belief  
 2. mad indeed i expect it case  
 my very senses reject their  
 own evidence  
 mad|indeed 1  
 mad|i  
 mad|expect  
 mad|it  
 mad|case  
 mad|my  
 mad|very  
 mad|senses  
 mad|reject  
 mad|their  
 mad|own  
 mad|evidence  
 indeed|i 1  
 indeed|expect  
 indeed|it  
 indeed|case  
 indeed|my  
 indeed|very  
 indeed|senses  
 indeed|reject

indeed|their  
indeed|own  
indeed|evidence  
i|expect 1  
i|it  
i|case  
i|my  
i|very  
i|senses  
i|reject  
i|their  
i|own  
i|evidence  
expect|it 1  
expect|case  
expect|my  
expect|very  
expect|senses  
expect|reject  
expect|their  
expect|own  
expect|evidence  
it|case 1  
it|my  
it|very  
it|senses  
it|reject  
it|their  
it|own  
it|evidence  
case|my  
case|very  
case|senses 1  
case|reject  
case|their  
case|own  
case|evidence  
my|very  
my|senses  
my|reject  
my|their  
my|own  
my|evidence  
very|senses 1  
very|reject  
very|their

very|own  
very|evidence  
senses|reject 1  
senses|their  
senses|own  
senses|evidence  
reject|their  
reject|own  
reject|evidence 1  
their|own  
their|evidence  
own|evidence 1  
3. mad i very surely do i  
dream  
mad|i 1  
mad|very  
mad|surely  
mad|do  
mad|i  
mad|dream  
i|very 1  
i|surely  
i|do  
i|i  
i|dream  
very|surely 1  
very|do  
very|i  
very|dream  
surely|do  
surely|i 1  
surely|dream  
do|i  
do|dream  
i|dream 1  
4. morrow i die day i  
unburthen my soul  
morrow|i 1  
morrow|die  
morrow|day  
morrow|i  
morrow|unburthen  
morrow|my  
morrow|soul  
i|die  
i|day



i|i  
i|unburthen  
i|my  
i|soul  
die|day 1  
die|i  
die|unburthen  
die|my  
die|soul  
day|i 1  
day|unburthen  
day|my  
day|soul  
i|unburthen 1  
i|my  
i|soul  
unburthen|my  
unburthen|soul 1  
my|soul  
5. my immediate purpose  
place world plainly  
succinctly comment series  
mere household events  
my|immediate 1  
my|purpose  
my|place  
my|world  
my|plainly  
my|succinctly  
my|comment  
my|series  
my|mere  
my|household  
my|events  
immediate|purpose 1  
immediate|place  
immediate|world  
immediate|plainly  
immediate|succinctly  
immediate|comment  
immediate|series  
immediate|mere  
immediate|household  
immediate|events  
purpose|place 1  
purpose|world

purpose|plainly  
purpose|succinctly  
purpose|comment  
purpose|series  
purpose|mere  
purpose|household  
purpose|events  
place|world 1  
place|plainly 1  
place|succinctly 1  
place|comment 1  
place|series 1  
place|mere  
place|household  
place|events  
world|plainly  
world|succinctly  
world|comment  
world|series  
world|mere  
world|household  
world|events  
plainly|succinctly  
plainly|comment  
plainly|series  
plainly|mere  
plainly|household  
plainly|events  
succinctly|comment  
succinctly|series  
succinctly|mere  
succinctly|household  
succinctly|events  
comment|series  
comment|mere  
comment|household  
comment|events  
series|mere  
series|household  
series|events 1  
mere|household 1  
mere|events  
household|events 1  
6. their consequences these  
events terrified tortured  
destroyed me

their|consequences  
 their|these  
 their|events  
 their|terrified  
 their|tortured  
 their|destroyed  
 their|me  
 consequences|these 1  
 consequences|events  
 consequences|terrified  
 consequences|tortured  
 consequences|destroyed  
 consequences|me  
 these|events 1  
 these|terrified  
 these|tortured  
 these|destroyed  
 these|me  
 events|terrified 1  
 events|tortured 1  
 events|destroyed 1  
 events|me  
 terrified|tortured  
 terrified|destroyed  
 terrified|me 1  
 tortured|destroyed  
 tortured|me 1  
 destroyed|me 1  
 7. i attempt expound them  
 i|attempt 1  
 i|expound  
 i|them  
 attempt|expound 1  
 attempt|them  
 expound|them 1  
 8. me they presented little  
 horror many they seem less  
 terrible barroques  
 me|they  
 me|presented  
 me|little  
 me|horror  
 me|many  
 me|they  
 me|seem  
 me|less

me|terrible  
 me|barroques  
 they|presented 1  
 they|little  
 they|horror  
 they|many  
 they|they  
 they|seem  
 they|less  
 they|terrible  
 they|barroques  
 presented|little 1  
 presented|horror  
 presented|many  
 presented|they  
 presented|seem  
 presented|less  
 presented|terrible  
 presented|barroques  
 little|horror 1  
 little|many  
 little|they  
 little|seem  
 little|less  
 little|terrible  
 little|barroques  
 horror|many  
 horror|they  
 horror|seem  
 horror|less  
 horror|terrible  
 horror|barroques  
 many|they 1  
 many|seem  
 many|less  
 many|terrible  
 many|barroques  
 they|seem 1  
 they|less  
 they|terrible  
 they|barroques  
 seem|less 1  
 seem|terrible  
 seem|barroques  
 less|terrible 1  
 less|barroques

terrible|barroques 1  
9. hereafter perhaps some  
intellect may found reduce  
my phantasm common place  
hereafter|perhaps 1  
hereafter|some  
hereafter|intellect  
hereafter|may  
hereafter|found  
hereafter|reduce  
hereafter|my  
hereafter|phantasm  
hereafter|common  
hereafter|place  
perhaps|some  
perhaps|intellect 1  
perhaps|may  
perhaps|found  
perhaps|reduce  
perhaps|my  
perhaps|phantasm  
perhaps|common  
perhaps|place  
some|intellect 1  
some|may  
some|found  
some|reduce  
some|my  
some|phantasm  
some|common  
some|place  
intellect|may  
intellect|found 1  
intellect|reduce  
intellect|my  
intellect|phantasm  
intellect|common  
intellect|place  
may|found  
may|reduce  
may|my  
may|phantasm  
may|common  
may|place  
found|reduce 1  
found|my

found|phantasm  
found|common  
found|place  
reduce|my  
reduce|phantasm 1  
reduce|common  
reduce|place  
my|phantasm  
my|common  
my|place  
phantasm|common 1  
phantasm|place  
common|place 1  
10. some intellect more calm  
more logical far less  
excitable my own perceive  
circumstances i detail awe  
nothing more ordinary  
succession very natural  
causes effects  
some|intellect 1  
some|more  
some|calm  
some|more  
some|logical  
some|far  
some|less  
some|excitable  
some|my  
some|own  
some|perceive  
some|circumstances  
some|i  
some|detail  
some|awe  
some|nothing  
some|more  
some|ordinary  
some|succession  
some|very  
some|natural  
some|causes  
some|effects  
intellect|more 1  
intellect|calm  
intellect|more 1

intellect|logical  
 intellect|far 1  
 intellect|less  
 intellect|excitable  
 intellect|my  
 intellect|own  
 intellect|perceive 1  
 intellect|circumstances  
 intellect|i  
 intellect|detail  
 intellect|awe  
 intellect|nothing  
 intellect|more  
 intellect|ordinary  
 intellect|succession  
 intellect|very  
 intellect|natural  
 intellect|causes  
 intellect|effects  
 more|calm 1  
 more|more  
 more|logical  
 more|far  
 more|less  
 more|excitable  
 more|my  
 more|own  
 more|perceive  
 more|circumstances  
 more|i  
 more|detail  
 more|awe  
 more|nothing  
 more|more  
 more|ordinary  
 more|succession  
 more|very  
 more|natural  
 more|causes  
 more|effects  
 calm|more  
 calm|logical  
 calm|far  
 calm|less  
 calm|excitable  
 calm|my

calm|own  
 calm|perceive  
 calm|circumstances  
 calm|i  
 calm|detail  
 calm|awe  
 calm|nothing  
 calm|more  
 calm|ordinary  
 calm|succession  
 calm|very  
 calm|natural  
 calm|causes  
 calm|effects  
 more|logical 1  
 more|far  
 more|less  
 more|excitable  
 more|my  
 more|own  
 more|perceive  
 more|circumstances  
 more|i  
 more|detail  
 more|awe  
 more|nothing  
 more|more  
 more|ordinary  
 more|succession  
 more|very  
 more|natural  
 more|causes  
 more|effects  
 logical|far  
 logical|less  
 logical|excitable  
 logical|my  
 logical|own  
 logical|perceive  
 logical|circumstances  
 logical|i  
 logical|detail  
 logical|awe  
 logical|nothing  
 logical|more  
 logical|ordinary

logical|succession

logical|very

logical|natural

logical|causes

logical|effects

far|less 1

far|excitable

far|my

far|own

far|perceive

far|circumstances

far|i

far|detail

far|awe

far|nothing

far|more

far|ordinary

far|succession

far|very

far|natural

far|causes

far|effects

less|excitable 1

less|my

less|own

less|perceive

less|circumstances

less|i

less|detail

less|awe

less|nothing

less|more

less|ordinary

less|succession

less|very

less|natural

less|causes

less|effects

excitable|my 1

excitable|own

excitable|perceive

excitable|circumstances

excitable|i

excitable|detail

excitable|awe

excitable|nothing

excitable|more

excitable|ordinary

excitable|succession

excitable|very

excitable|natural

excitable|causes

excitable|effects

my|own 1

my|perceive

my|circumstances

my|i

my|detail

my|awe

my|nothing

my|more

my|ordinary

my|succession

my|very

my|natural

my|causes

my|effects

own|perceive

own|circumstances

own|i

own|detail

own|awe

own|nothing

own|more

own|ordinary

own|succession

own|very

own|natural

own|causes

own|effects

perceive|circumstances 1

perceive|i

perceive|detail

perceive|awe

perceive|nothing 1

perceive|more

perceive|ordinary

perceive|succession

perceive|very

perceive|natural

perceive|causes

perceive|effects

circumstances|i 1  
 circumstances|detail  
 circumstances|awe  
 circumstances|nothing  
 circumstances|more  
 circumstances|ordinary  
 circumstances|succession  
 circumstances|very  
 circumstances|natural  
 circumstances|causes  
 circumstances|effects  
 i|detail 1  
 i|awe  
 i|nothing  
 i|more  
 i|ordinary  
 i|succession  
 i|very  
 i|natural  
 i|causes  
 i|effects  
 detail|awe 1  
 detail|nothing  
 detail|more  
 detail|ordinary  
 detail|succession  
 detail|very  
 detail|natural  
 detail|causes  
 detail|effects  
 awe|nothing  
 awe|more  
 awe|ordinary  
 awe|succession  
 awe|very  
 awe|natural  
 awe|causes  
 awe|effects  
 nothing|more 1  
 nothing|ordinary  
 nothing|succession  
 nothing|very  
 nothing|natural  
 nothing|causes  
 nothing|effects  
 more|ordinary 1

more|succession  
 more|very  
 more|natural  
 more|causes  
 more|effects  
 ordinary|succession 1  
 ordinary|very  
 ordinary|natural  
 ordinary|causes  
 ordinary|effects  
 succession|very 1  
 succession|natural  
 succession|causes  
 succession|effects  
 very|natural 1  
 very|causes  
 very|effects  
 natural|causes 1  
 natural|effects  
 causes|effects 1  
 11. my infancy i was noted  
 docility humanity my  
 disposition  
 my|infancy 1  
 my|i  
 my|was  
 my|noted  
 my|docility  
 my|humanity  
 my|my  
 my|disposition  
 infancy|i 1  
 infancy|was  
 infancy|noted  
 infancy|docility  
 infancy|humanity  
 infancy|my  
 infancy|disposition  
 i|was 1  
 i|noted  
 i|docility  
 i|humanity  
 i|my  
 i|disposition  
 was|noted 1  
 was|docility

was|humanity  
was|my  
was|disposition  
noted|docility 1  
noted|humanity 1  
noted|my  
noted|disposition  
docility|humanity 1  
docility|my  
docility|disposition  
humanity|my  
humanity|disposition 1  
my|disposition  
12. my tenderness heart was  
even conspicuous make me  
jest my companions  
my|tenderness 1  
my|heart  
my|was  
my|even  
my|conspicuous  
my|make  
my|me  
my|jest  
my|my  
my|companions  
tenderness|heart 1  
tenderness|was  
tenderness|even  
tenderness|conspicuous  
tenderness|make  
tenderness|me  
tenderness|jest  
tenderness|my  
tenderness|companions  
heart|was 1  
heart|even  
heart|conspicuous  
heart|make  
heart|me  
heart|jest  
heart|my  
heart|companions  
was|even 1  
was|conspicuous  
was|make

was|me  
was|jest  
was|my  
was|companions  
even|conspicuous 1  
even|make  
even|me  
even|jest  
even|my  
even|companions  
conspicuous|make 1  
conspicuous|me  
conspicuous|jest  
conspicuous|my  
conspicuous|companions  
make|me  
make|jest 1  
make|my  
make|companions  
me|jest  
me|my  
me|companions  
jest|my  
jest|companions 1  
my|companions  
13. i was especially fond  
animals was indulged my  
parents great variety pets  
i|was 1  
i|especially  
i|fond  
i|animals  
i|was  
i|indulged  
i|my  
i|parents  
i|great  
i|variety  
i|pets  
was|especially 1  
was|fond  
was|animals  
was|was  
was|indulged  
was|my  
was|parents

was|great  
was|variety  
was|pets  
especially|fond 1  
especially|animals  
especially|was  
especially|indulged  
especially|my  
especially|parents  
especially|great  
especially|variety  
especially|pets  
fond|animals 1  
fond|was  
fond|indulged  
fond|my  
fond|parents  
fond|great  
fond|variety  
fond|pets  
animals|was  
animals|indulged  
animals|my  
animals|parents  
animals|great  
animals|variety  
animals|pets  
was|indulged 1  
was|my  
was|parents  
was|great  
was|variety  
was|pets  
indulged|my 1  
indulged|parents  
indulged|great  
indulged|variety  
indulged|pets 1  
my|parents 1  
my|great  
my|variety  
my|pets  
parents|great  
parents|variety  
parents|pets  
great|variety 1

great|pets  
variety|pets 1  
14. these i spent most my  
time never was happy  
feeding caressing them  
these|i  
these|spent  
these|most  
these|my  
these|time  
these|never  
these|was  
these|happy  
these|feeding  
these|caressing  
these|them  
i|spent 1  
i|most  
i|my  
i|time  
i|never  
i|was  
i|happy  
i|feeding  
i|caressing  
i|them  
spent|most 1  
spent|my  
spent|time  
spent|never  
spent|was  
spent|happy  
spent|feeding  
spent|caressing  
spent|them  
most|my 1  
most|time  
most|never  
most|was  
most|happy  
most|feeding  
most|caressing  
most|them  
my|time 1  
my|never  
my|was



my|happy  
 my|feeding  
 my|caressing  
 my|them  
 time|never  
 time|was  
 time|happy  
 time|feeding  
 time|caressing  
 time|them  
 never|was 1  
 never|happy  
 never|feeding  
 never|caressing  
 never|them  
 was|happy 1  
 was|feeding  
 was|caressing  
 was|them  
 happy|feeding 1  
 happy|caressing 1  
 happy|them  
 feeding|caressing 1  
 feeding|them  
 caressing|them  
 15. this peculiarity character  
 grew my growth my  
 manhood i derived it one my  
 principal sources pleasure  
 this|peculiarity 1  
 this|character  
 this|grew  
 this|my  
 this|growth  
 this|my  
 this|manhood  
 this|i  
 this|derived  
 this|it  
 this|one  
 this|my  
 this|principal  
 this|sources  
 this|pleasure  
 peculiarity|character 1  
 peculiarity|grew

peculiarity|my  
 peculiarity|growth  
 peculiarity|my  
 peculiarity|manhood  
 peculiarity|i  
 peculiarity|derived  
 peculiarity|it  
 peculiarity|one  
 peculiarity|my  
 peculiarity|principal  
 peculiarity|sources  
 peculiarity|pleasure  
 character|grew 1  
 character|my  
 character|growth  
 character|my  
 character|manhood  
 character|i  
 character|derived  
 character|it  
 character|one  
 character|my  
 character|principal  
 character|sources  
 character|pleasure  
 grew|my 1  
 grew|growth 1  
 grew|my  
 grew|manhood  
 grew|i  
 grew|derived  
 grew|it  
 grew|one  
 grew|my  
 grew|principal  
 grew|sources  
 grew|pleasure  
 my|growth 1  
 my|my  
 my|manhood  
 my|i  
 my|derived  
 my|it  
 my|one  
 my|my  
 my|principal

my|sources  
my|pleasure  
growth|my  
growth|manhood 1  
growth|i  
growth|derived  
growth|it  
growth|one  
growth|my  
growth|principal  
growth|sources  
growth|pleasure  
my|manhood 1  
my|i  
my|derived  
my|it  
my|one  
my|my  
my|principal  
my|sources  
my|pleasure  
manhood|i  
manhood|derived  
manhood|it  
manhood|one  
manhood|my  
manhood|principal  
manhood|sources  
manhood|pleasure  
i|derived 1  
i|it  
i|one  
i|my  
i|principal  
i|sources  
i|pleasure  
derived|it  
derived|one  
derived|my  
derived|principal  
derived|sources 1  
derived|pleasure  
it|one  
it|my  
it|principal  
it|sources

it|pleasure  
one|my  
one|principal  
one|sources  
one|pleasure  
my|principal 1  
my|sources  
my|pleasure  
principal|sources 1  
principal|pleasure  
sources|pleasure 1  
16. those cherished affection  
faithful sagacious dog i need  
hardly trouble explaining  
nature intensity gratification  
derivable  
those|cherished 1  
those|affection  
those|faithful  
those|sagacious  
those|dog  
those|i  
those|need  
those|hardly  
those|trouble  
those|explaining  
those|nature  
those|intensity  
those|gratification  
those|derivable  
cherished|affection 1  
cherished|faithful  
cherished|sagacious  
cherished|dog  
cherished|i  
cherished|need  
cherished|hardly  
cherished|trouble  
cherished|explaining  
cherished|nature  
cherished|intensity  
cherished|gratification  
cherished|derivable  
affection|faithful  
affection|sagacious  
affection|dog 1

affection|i  
affection|need  
affection|hardly  
affection|trouble  
affection|explaining  
affection|nature  
affection|intensity  
affection|gratification  
affection|derivable  
faithful|sagacious  
faithful|dog 1  
faithful|i  
faithful|need  
faithful|hardly  
faithful|trouble  
faithful|explaining  
faithful|nature  
faithful|intensity  
faithful|gratification  
faithful|derivable  
sagacious|dog 1  
sagacious|i  
sagacious|need  
sagacious|hardly  
sagacious|trouble  
sagacious|explaining  
sagacious|nature  
sagacious|intensity  
sagacious|gratification  
sagacious|derivable  
dog|i  
dog|need  
dog|hardly  
dog|trouble  
dog|explaining  
dog|nature  
dog|intensity  
dog|gratification  
dog|derivable  
i|need 1  
i|hardly  
i|trouble  
i|explaining  
i|nature  
i|intensity  
i|gratification

i|derivable  
need|hardly 1  
need|trouble  
need|explaining  
need|nature  
need|intensity  
need|gratification  
need|derivable  
hardly|trouble 1  
hardly|explaining  
hardly|nature  
hardly|intensity  
hardly|gratification  
hardly|derivable  
trouble|explaining 1  
trouble|nature  
trouble|intensity  
trouble|gratification  
trouble|derivable  
explaining|nature 1  
explaining|intensity 1  
explaining|gratification  
explaining|derivable  
nature|intensity  
nature|gratification  
nature|derivable  
intensity|gratification 1  
intensity|derivable  
gratification|derivable 1  
17. there something unselfish  
self sacrificing love brute  
goes directly heart him has  
had frequent occasion test  
paltry friendship gossamer  
fidelity mere man  
there|something 1  
there|unselfish  
there|self  
there|sacrificing  
there|love  
there|brute  
there|goes  
there|directly  
there|heart  
there|him  
there|has

there|had  
 there|frequent  
 there|occasion  
 there|test  
 there|paltry  
 there|friendship  
 there|gossamer  
 there|fidelity  
 there|mere  
 there|man  
 something|unselfish 1  
 something|self  
 something|sacrificing  
 something|love  
 something|brute  
 something|goes  
 something|directly  
 something|heart  
 something|him  
 something|has  
 something|had  
 something|frequent  
 something|occasion  
 something|test  
 something|paltry  
 something|friendship  
 something|gossamer  
 something|fidelity  
 something|mere  
 something|man  
 unselfish|self  
 unselfish|sacrificing  
 unselfish|love 1  
 unselfish|brute  
 unselfish|goes  
 unselfish|directly  
 unselfish|heart  
 unselfish|him  
 unselfish|has  
 unselfish|had  
 unselfish|frequent  
 unselfish|occasion  
 unselfish|test  
 unselfish|paltry  
 unselfish|friendship  
 unselfish|gossamer

unselfish|fidelity  
 unselfish|mere  
 unselfish|man  
 self|sacrificing 1  
 self|love  
 self|brute  
 self|goes  
 self|directly  
 self|heart  
 self|him  
 self|has  
 self|had  
 self|frequent  
 self|occasion  
 self|test  
 self|paltry  
 self|friendship  
 self|gossamer  
 self|fidelity  
 self|mere  
 self|man  
 sacrificing|love 1  
 sacrificing|brute  
 sacrificing|goes  
 sacrificing|directly  
 sacrificing|heart  
 sacrificing|him  
 sacrificing|has  
 sacrificing|had  
 sacrificing|frequent  
 sacrificing|occasion  
 sacrificing|test  
 sacrificing|paltry  
 sacrificing|friendship  
 sacrificing|gossamer  
 sacrificing|fidelity  
 sacrificing|mere  
 sacrificing|man  
 love|brute 1  
 love|goes  
 love|directly  
 love|heart  
 love|him  
 love|has  
 love|had  
 love|frequent

love|occasion  
love|test  
love|paltry  
love|friendship  
love|gossamer  
love|fidelity  
love|mere  
love|man  
brute|goes 1  
brute|directly  
brute|heart  
brute|him  
brute|has  
brute|had  
brute|frequent  
brute|occasion  
brute|test  
brute|paltry  
brute|friendship  
brute|gossamer  
brute|fidelity  
brute|mere  
brute|man  
goes|directly 1  
goes|heart  
goes|him  
goes|has  
goes|had  
goes|frequent  
goes|occasion  
goes|test  
goes|paltry  
goes|friendship  
goes|gossamer  
goes|fidelity  
goes|mere  
goes|man  
directly|heart 1  
directly|him  
directly|has  
directly|had  
directly|frequent  
directly|occasion  
directly|test  
directly|paltry  
directly|friendship

directly|gossamer  
directly|fidelity  
directly|mere  
directly|man  
heart|him 1  
heart|has  
heart|had  
heart|frequent  
heart|occasion  
heart|test  
heart|paltry  
heart|friendship  
heart|gossamer  
heart|fidelity  
heart|mere  
heart|man 1  
him|has 1  
him|had  
him|frequent  
him|occasion  
him|test  
him|paltry  
him|friendship  
him|gossamer  
him|fidelity  
him|mere  
him|man  
has|had 1  
has|frequent  
has|occasion  
has|test  
has|paltry  
has|friendship  
has|gossamer  
has|fidelity  
has|mere  
has|man  
had|frequent 1  
had|occasion  
had|test  
had|paltry  
had|friendship  
had|gossamer  
had|fidelity  
had|mere  
had|man

frequent|occasion 1  
frequent|test  
frequent|paltry  
frequent|friendship  
frequent|gossamer  
frequent|fidelity  
frequent|mere  
frequent|man  
occasion|test 1  
occasion|paltry  
occasion|friendship  
occasion|gossamer  
occasion|fidelity  
occasion|mere  
occasion|man  
test|paltry  
test|friendship 1  
test|gossamer  
test|fidelity 1  
test|mere  
test|man  
paltry|friendship 1  
paltry|gossamer  
paltry|fidelity  
paltry|mere  
paltry|man  
friendship|gossamer  
friendship|fidelity  
friendship|mere  
friendship|man  
gossamer|fidelity 1  
gossamer|mere  
gossamer|man  
fidelity|mere 1  
fidelity|man  
mere|man 1  
18. i married early was  
happy find my wife  
disposition uncongenial my  
own  
i|married 1  
i|early  
i|was  
i|happy  
i|find  
i|my

i|wife  
i|disposition  
i|uncongenial  
i|my  
i|own  
married|early 1  
married|was 1  
married|happy  
married|find  
married|my  
married|wife  
married|disposition  
married|uncongenial  
married|my  
married|own  
early|was  
early|happy  
early|find  
early|my  
early|wife  
early|disposition  
early|uncongenial  
early|my  
early|own  
was|happy 1  
was|find  
was|my  
was|wife  
was|disposition  
was|uncongenial  
was|my  
was|own  
happy|find 1  
happy|my  
happy|wife  
happy|disposition  
happy|uncongenial  
happy|my  
happy|own  
find|my  
find|wife 1  
find|disposition 1  
find|uncongenial  
find|my  
find|own  
my|wife 1

my|disposition  
my|uncongenial  
my|my  
my|own  
wife|disposition  
wife|uncongenial  
wife|my  
wife|own  
disposition|uncongenial 1  
disposition|my  
disposition|own  
uncongenial|my  
uncongenial|own 1  
my|own 1  
19. observing my partiality  
domestic pets she lost  
opportunity procuring those  
most agreeable kind  
observing|my 1  
observing|partiality  
observing|domestic  
observing|pets  
observing|she  
observing|lost  
observing|opportunity  
observing|procuring  
observing|those  
observing|most  
observing|agreeable  
observing|kind  
my|partiality 1  
my|domestic  
my|pets  
my|she  
my|lost  
my|opportunity  
my|procuring  
my|those  
my|most  
my|agreeable  
my|kind  
partiality|domestic  
partiality|pets 1  
partiality|she  
partiality|lost  
partiality|opportunity

partiality|procuring  
partiality|those  
partiality|most  
partiality|agreeable  
partiality|kind  
domestic|pets 1  
domestic|she  
domestic|lost  
domestic|opportunity  
domestic|procuring  
domestic|those  
domestic|most  
domestic|agreeable  
domestic|kind  
pets|she  
pets|lost  
pets|opportunity  
pets|procuring  
pets|those  
pets|most  
pets|agreeable  
pets|kind  
she|lost 1  
she|opportunity  
she|procuring  
she|those  
she|most  
she|agreeable  
she|kind  
lost|opportunity 1  
lost|procuring  
lost|those  
lost|most  
lost|agreeable  
lost|kind  
opportunity|procuring 1  
opportunity|those  
opportunity|most  
opportunity|agreeable  
opportunity|kind  
procuring|those  
procuring|most  
procuring|agreeable  
procuring|kind  
those|most  
those|agreeable 1

those|kind  
most|agreeable 1  
most|kind  
agreeable|kind 1  
20. we had birds gold fish  
fine dog rabbits small  
monkey cat  
we|had 1  
we|birds  
we|gold  
we|fish  
we|fine  
we|dog  
we|rabbits  
we|small  
we|monkey  
we|cat  
had|birds 1  
had|gold  
had|fish 1  
had|fine  
had|dog 1  
had|rabbits 1  
had|small  
had|monkey 1  
had|cat 1  
birds|gold  
birds|fish  
birds|fine  
birds|dog  
birds|rabbits  
birds|small  
birds|monkey  
birds|cat  
gold|fish 1  
gold|fine  
gold|dog  
gold|rabbits  
gold|small  
gold|monkey  
gold|cat  
fish|fine  
fish|dog  
fish|rabbits  
fish|small  
fish|monkey

fish|cat  
fine|dog 1  
fine|rabbits  
fine|small  
fine|monkey  
fine|cat  
dog|rabbits  
dog|small  
dog|monkey  
dog|cat  
rabbits|small  
rabbits|monkey  
rabbits|cat  
small|monkey 1  
small|cat  
monkey|cat  
21. this latter was  
remarkably large beautiful  
animal entirely black  
sagacious astonishing degree  
this|latter 1  
this|was  
this|remarkably  
this|large  
this|beautiful  
this|animal  
this|entirely  
this|black  
this|sagacious  
this|astonishing  
this|degree  
latter|was 1  
latter|remarkably  
latter|large  
latter|beautiful  
latter|animal  
latter|entirely  
latter|black  
latter|sagacious  
latter|astonishing  
latter|degree  
was|remarkably  
was|large  
was|beautiful  
was|animal 1  
was|entirely



was|black  
 was|sagacious  
 was|astonishing  
 was|degree  
 remarkably|large 1  
 remarkably|beautiful  
 remarkably|animal  
 remarkably|entirely  
 remarkably|black  
 remarkably|sagacious  
 remarkably|astonishing  
 remarkably|degree  
 large|beautiful  
 large|animal 1  
 large|entirely  
 large|black  
 large|sagacious  
 large|astonishing  
 large|degree  
 beautiful|animal 1  
 beautiful|entirely  
 beautiful|black  
 beautiful|sagacious  
 beautiful|astonishing  
 beautiful|degree  
 animal|entirely  
 animal|black 1  
 animal|sagacious 1  
 animal|astonishing  
 animal|degree  
 entirely|black 1  
 entirely|sagacious  
 entirely|astonishing  
 entirely|degree  
 black|sagacious  
 black|astonishing  
 black|degree  
 sagacious|astonishing  
 sagacious|degree 1  
 astonishing|degree 1  
 22. speaking his intelligence  
 my wife heart was little  
 tintured superstition made  
 frequent allusion ancient  
 popular notion regarded all  
 black cats witches disguise

speaking|his  
 speaking|intelligence 1  
 speaking|my  
 speaking|wife 1  
 speaking|heart  
 speaking|was  
 speaking|little  
 speaking|tintured  
 speaking|superstition  
 speaking|made  
 speaking|frequent  
 speaking|allusion  
 speaking|ancient  
 speaking|popular  
 speaking|notion  
 speaking|regarded  
 speaking|all  
 speaking|black  
 speaking|cats  
 speaking|witches  
 speaking|disguise  
 his|intelligence 1  
 his|my  
 his|wife  
 his|heart  
 his|was  
 his|little  
 his|tintured  
 his|superstition  
 his|made  
 his|frequent  
 his|allusion  
 his|ancient  
 his|popular  
 his|notion  
 his|regarded  
 his|all  
 his|black  
 his|cats  
 his|witches  
 his|disguise  
 intelligence|my  
 intelligence|wife  
 intelligence|heart  
 intelligence|was  
 intelligence|little

intelligence|tintured  
 intelligence|superstition  
 intelligence|made  
 intelligence|frequent  
 intelligence|allusion  
 intelligence|ancient  
 intelligence|popular  
 intelligence|notion  
 intelligence|regarded  
 intelligence|all  
 intelligence|black  
 intelligence|cats  
 intelligence|witches  
 intelligence|disguise  
 my|wife 1  
 my|heart  
 my|was  
 my|little  
 my|tintured  
 my|superstition  
 my|made  
 my|frequent  
 my|allusion  
 my|ancient  
 my|popular  
 my|notion  
 my|regarded  
 my|all  
 my|black  
 my|cats  
 my|witches  
 my|disguise  
 wife|heart 1  
 wife|was  
 wife|little  
 wife|tintured  
 wife|superstition  
 wife|made  
 wife|frequent  
 wife|allusion  
 wife|ancient  
 wife|popular  
 wife|notion  
 wife|regarded  
 wife|all  
 wife|black

wife|cats  
 wife|witches  
 wife|disguise  
 heart|was 1  
 heart|little  
 heart|tintured  
 heart|superstition  
 heart|made  
 heart|frequent  
 heart|allusion  
 heart|ancient  
 heart|popular  
 heart|notion  
 heart|regarded  
 heart|all  
 heart|black  
 heart|cats  
 heart|witches  
 heart|disguise  
 was|little 1  
 was|tintured  
 was|superstition  
 was|made  
 was|frequent  
 was|allusion  
 was|ancient  
 was|popular  
 was|notion  
 was|regarded  
 was|all  
 was|black  
 was|cats  
 was|witches  
 was|disguise  
 little|tintured 1  
 little|superstition  
 little|made  
 little|frequent  
 little|allusion  
 little|ancient  
 little|popular  
 little|notion  
 little|regarded  
 little|all  
 little|black  
 little|cats

little|witches  
 little|disguise  
 tintured|superstition 1  
 tintured|made  
 tintured|frequent  
 tintured|allusion  
 tintured|ancient  
 tintured|popular  
 tintured|notion  
 tintured|regarded  
 tintured|all  
 tintured|black  
 tintured|cats  
 tintured|witches  
 tintured|disguise  
 superstition|made  
 superstition|frequent  
 superstition|allusion  
 superstition|ancient  
 superstition|popular  
 superstition|notion  
 superstition|regarded  
 superstition|all  
 superstition|black  
 superstition|cats  
 superstition|witches  
 superstition|disguise  
 made|frequent 1  
 made|allusion  
 made|ancient  
 made|popular  
 made|notion  
 made|regarded  
 made|all  
 made|black  
 made|cats  
 made|witches  
 made|disguise  
 frequent|allusion 1  
 frequent|ancient  
 frequent|popular  
 frequent|notion  
 frequent|regarded  
 frequent|all  
 frequent|black  
 frequent|cats

frequent|witches  
 frequent|disguise  
 allusion|ancient 1  
 allusion|popular  
 allusion|notion  
 allusion|regarded  
 allusion|all  
 allusion|black  
 allusion|cats  
 allusion|witches  
 allusion|disguise  
 ancient|popular 1  
 ancient|notion  
 ancient|regarded  
 ancient|all  
 ancient|black  
 ancient|cats  
 ancient|witches  
 ancient|disguise  
 popular|notion 1  
 popular|regarded  
 popular|all  
 popular|black  
 popular|cats  
 popular|witches  
 popular|disguise  
 notion|regarded 1  
 notion|all  
 notion|black  
 notion|cats  
 notion|witches  
 notion|disguise  
 regarded|all 1  
 regarded|black  
 regarded|cats  
 regarded|witches  
 regarded|disguise  
 all|black  
 all|cats  
 all|witches 1  
 all|disguise  
 black|cats 1  
 black|witches  
 black|disguise  
 cats|witches 1  
 cats|disguise

witches|disguise 1  
 23. that she was ever serious  
 this point i mention matter  
 all better reason that it  
 happens just now  
 remembered  
 that|she 1  
 that|was  
 that|ever  
 that|serious  
 that|this  
 that|point  
 that|i  
 that|mention  
 that|matter  
 that|all  
 that|better  
 that|reason  
 that|that  
 that|it  
 that|happens  
 that|just  
 that|now  
 that|remembered  
 she|was 1  
 she|ever  
 she|serious  
 she|this  
 she|point  
 she|i  
 she|mention  
 she|matter  
 she|all  
 she|better  
 she|reason  
 she|that  
 she|it  
 she|happens  
 she|just  
 she|now  
 she|remembered  
 was|ever 1  
 was|serious  
 was|this  
 was|point  
 was|i

was|mention  
 was|matter  
 was|all  
 was|better  
 was|reason  
 was|that  
 was|it  
 was|happens  
 was|just  
 was|now  
 was|remembered  
 ever|serious 1  
 ever|this  
 ever|point  
 ever|i  
 ever|mention  
 ever|matter  
 ever|all  
 ever|better  
 ever|reason  
 ever|that  
 ever|it  
 ever|happens  
 ever|just  
 ever|now  
 ever|remembered  
 serious|this  
 serious|point 1  
 serious|i  
 serious|mention  
 serious|matter  
 serious|all  
 serious|better  
 serious|reason  
 serious|that  
 serious|it  
 serious|happens  
 serious|just  
 serious|now  
 serious|remembered  
 this|point 1  
 this|i  
 this|mention  
 this|matter  
 this|all  
 this|better

this|reason  
 this|that  
 this|it  
 this|happens  
 this|just  
 this|now  
 this|remembered  
 point|i  
 point|mention  
 point|matter  
 point|all  
 point|better  
 point|reason  
 point|that  
 point|it  
 point|happens  
 point|just  
 point|now  
 point|remembered  
 i|mention 1  
 i|matter  
 i|all  
 i|better  
 i|reason  
 i|that  
 i|it  
 i|happens  
 i|just  
 i|now  
 i|remembered  
 mention|matter 1  
 mention|all  
 mention|better  
 mention|reason  
 mention|that  
 mention|it  
 mention|happens  
 mention|just  
 mention|now  
 mention|remembered  
 matter|all 1  
 matter|better  
 matter|reason  
 matter|that  
 matter|it  
 matter|happens

matter|just  
 matter|now  
 matter|remembered  
 all|better  
 all|reason  
 all|that  
 all|it  
 all|happens  
 all|just  
 all|now  
 all|remembered  
 better|reason 1  
 better|that  
 better|it  
 better|happens  
 better|just  
 better|now  
 better|remembered  
 reason|that  
 reason|it  
 reason|happens 1  
 reason|just  
 reason|now  
 reason|remembered  
 that|it  
 that|happens  
 that|just  
 that|now  
 that|remembered  
 it|happens 1  
 it|just  
 it|now  
 it|remembered  
 happens|just  
 happens|now  
 happens|remembered  
 just|now  
 just|remembered  
 now|remembered  
 24. pluto this was cat is name  
 was my favorite pet playmate  
 pluto|this  
 pluto|was 1  
 pluto|cat  
 pluto|is  
 pluto|name

pluto|was 1  
 pluto|my  
 pluto|favorite  
 pluto|pet  
 pluto|playmate  
 this|was 1  
 this|cat  
 this|is  
 this|name  
 this|was  
 this|my  
 this|favorite  
 this|pet  
 this|playmate  
 was|cat 1  
 was|is  
 was|name  
 was|was  
 was|my 1  
 was|favorite  
 was|pet  
 was|playmate  
 cat|is 1  
 cat|name  
 cat|was  
 cat|my  
 cat|favorite  
 cat|pet  
 cat|playmate  
 is|name 1  
 is|was  
 is|my  
 is|favorite  
 is|pet  
 is|playmate  
 name|was  
 name|my  
 name|favorite  
 name|pet  
 name|playmate  
 was|my  
 was|favorite  
 was|pet  
 was|playmate  
 my|favorite 1  
 my|pet

my|playmate  
 favorite|pet 1  
 favorite|playmate 1  
 pet|playmate  
 25. i alone fed him he  
 attended me wherever i went  
 house  
 i|alone 1  
 i|fed  
 i|him  
 i|he  
 i|attended  
 i|me  
 i|wherever  
 i|i  
 i|went  
 i|house  
 alone|fed 1  
 alone|him  
 alone|he  
 alone|attended  
 alone|me  
 alone|wherever  
 alone|i  
 alone|went  
 alone|house  
 fed|him  
 fed|he  
 fed|attended  
 fed|me  
 fed|wherever  
 fed|i  
 fed|went  
 fed|house  
 him|he  
 him|attended  
 him|me  
 him|wherever  
 him|i  
 him|went  
 him|house  
 he|attended  
 he|me  
 he|wherever  
 he|i  
 he|went

he|house  
 attended|me 1  
 attended|wherever  
 attended|i  
 attended|went 1  
 attended|house  
 me|wherever  
 me|i  
 me|went  
 me|house  
 wherever|i  
 wherever|went 1  
 wherever|house  
 i|went 1  
 i|house  
 went|house 1  
 26. it was even difficulty that  
 i prevent him following me  
 through streets  
 it|was 1  
 it|even  
 it|difficulty  
 it|that  
 it|i  
 it|prevent  
 it|him  
 it|following  
 it|me  
 it|through  
 it|streets  
 was|even 1  
 was|difficulty  
 was|that  
 was|i  
 was|prevent  
 was|him  
 was|following  
 was|me  
 was|through  
 was|streets  
 even|difficulty 1  
 even|that  
 even|i  
 even|prevent  
 even|him  
 even|following

even|me  
 even|through  
 even|streets  
 difficulty|that 1  
 difficulty|i  
 difficulty|prevent 1  
 difficulty|him  
 difficulty|following  
 difficulty|me  
 difficulty|through  
 difficulty|streets  
 that|i  
 that|prevent  
 that|him  
 that|following  
 that|me  
 that|through  
 that|streets  
 i|prevent  
 i|him  
 i|following  
 i|me  
 i|through  
 i|streets  
 prevent|him 1  
 prevent|following 1  
 prevent|me  
 prevent|through  
 prevent|streets  
 him|following  
 him|me  
 him|through  
 him|streets  
 following|me 1  
 following|through 1  
 following|streets  
 me|through  
 me|streets  
 through|streets 1  
 27. our friendship lasted this  
 manner several years my  
 general temperament  
 character through  
 instrumentality fiend  
 intemperance had i blush  
 confess it experienced

radical alteration worse  
 our|friendship 1  
 our|lasted  
 our|this  
 our|manner  
 our|several  
 our|years  
 our|my  
 our|general  
 our|temperament  
 our|character  
 our|through  
 our|instrumentality  
 our|fiend  
 our|intemperance  
 our|had  
 our|i  
 our|blush  
 our|confess  
 our|it  
 our|experienced  
 our|radical  
 our|alteration  
 our|worse  
 friendship|lasted 1  
 friendship|this  
 friendship|manner  
 friendship|several  
 friendship|years  
 friendship|my  
 friendship|general  
 friendship|temperament  
 friendship|character  
 friendship|through  
 friendship|instrumentality  
 friendship|fiend  
 friendship|intemperance  
 friendship|had  
 friendship|i  
 friendship|blush  
 friendship|confess  
 friendship|it  
 friendship|experienced  
 friendship|radical  
 friendship|alteration  
 friendship|worse

lasted|this  
 lasted|manner 1  
 lasted|several  
 lasted|years 1  
 lasted|my  
 lasted|general  
 lasted|temperament  
 lasted|character  
 lasted|through  
 lasted|instrumentality  
 lasted|fiend  
 lasted|intemperance  
 lasted|had  
 lasted|i  
 lasted|blush  
 lasted|confess  
 lasted|it  
 lasted|experienced  
 lasted|radical  
 lasted|alteration  
 lasted|worse  
 this|manner 1  
 this|several  
 this|years  
 this|my  
 this|general  
 this|temperament  
 this|character  
 this|through  
 this|instrumentality  
 this|fiend  
 this|intemperance  
 this|had  
 this|i  
 this|blush  
 this|confess  
 this|it  
 this|experienced  
 this|radical  
 this|alteration  
 this|worse  
 manner|several  
 manner|years  
 manner|my  
 manner|general  
 manner|temperament



manner|character  
manner|through  
manner|instrumentality  
manner|fiend  
manner|intemperance  
manner|had  
manner|i  
manner|blush  
manner|confess  
manner|it  
manner|experienced  
manner|radical  
manner|alteration  
manner|worse  
several|years 1  
several|my  
several|general  
several|temperament  
several|character  
several|through  
several|instrumentality  
several|fiend  
several|intemperance  
several|had  
several|i  
several|blush  
several|confess  
several|it  
several|experienced  
several|radical  
several|alteration  
several|worse  
years|my  
years|general  
years|temperament  
years|character  
years|through  
years|instrumentality  
years|fiend  
years|intemperance  
years|had  
years|i  
years|blush  
years|confess  
years|it  
years|experienced

years|radical  
years|alteration  
years|worse  
my|general 1  
my|temperament  
my|character  
my|through  
my|instrumentality  
my|fiend  
my|intemperance  
my|had  
my|i  
my|blush  
my|confess  
my|it  
my|experienced  
my|radical  
my|alteration  
my|worse  
general|temperament 1  
general|character  
general|through  
general|instrumentality  
general|fiend  
general|intemperance  
general|had  
general|i  
general|blush  
general|confess  
general|it  
general|experienced  
general|radical  
general|alteration  
general|worse  
temperament|character 1  
temperament|through  
temperament|instrumentality  
temperament|fiend  
temperament|intemperance  
temperament|had  
temperament|i  
temperament|blush  
temperament|confess  
temperament|it  
temperament|experienced  
temperament|radical

temperament|alteration  
 temperament|worse  
 character|through 1  
 character|instrumentality  
 character|fiend  
 character|intemperance  
 character|had  
 character|i  
 character|blush  
 character|confess  
 character|it  
 character|experienced  
 character|radical  
 character|alteration  
 character|worse  
 through|instrumentality 1  
 through|fiend  
 through|intemperance  
 through|had  
 through|i  
 through|blush  
 through|confess  
 through|it  
 through|experienced  
 through|radical  
 through|alteration  
 through|worse  
 instrumentality|fiend  
 instrumentality|intemperance  
 1  
 instrumentality|had  
 instrumentality|i  
 instrumentality|blush  
 instrumentality|confess  
 instrumentality|it  
 instrumentality|experienced  
 instrumentality|radical  
 instrumentality|alteration  
 instrumentality|worse  
 fiend|intemperance 1  
 fiend|had  
 fiend|i  
 fiend|blush  
 fiend|confess  
 fiend|it  
 fiend|experienced

fiend|radical  
 fiend|alteration  
 fiend|worse  
 intemperance|had 1  
 intemperance|i  
 intemperance|blush  
 intemperance|confess  
 intemperance|it  
 intemperance|experienced  
 intemperance|radical  
 intemperance|alteration  
 intemperance|worse  
 had|i  
 had|blush  
 had|confess  
 had|it  
 had|experienced 1  
 had|radical  
 had|alteration  
 had|worse  
 i|blush 1  
 i|confess  
 i|it  
 i|experienced  
 i|radical  
 i|alteration  
 i|worse  
 blush|confess 1  
 blush|it  
 blush|experienced  
 blush|radical  
 blush|alteration  
 blush|worse  
 confess|it 1  
 confess|experienced  
 confess|radical  
 confess|alteration  
 confess|worse  
 it|experienced  
 it|radical  
 it|alteration  
 it|worse  
 experienced|radical  
 experienced|alteration 1  
 experienced|worse 1  
 radical|alteration 1

radical|worse  
 alteration|worse  
 28. i grew day day more  
 moody more irritable more  
 regardless feelings others  
 i|grew 1  
 i|day  
 i|day  
 i|more  
 i|moody  
 i|more  
 i|irritable  
 i|more  
 i|regardless  
 i|feelings  
 i|others  
 grew|day 1  
 grew|day  
 grew|more  
 grew|moody  
 grew|more  
 grew|irritable  
 grew|more  
 grew|regardless  
 grew|feelings  
 grew|others  
 day|day 1  
 day|more  
 day|moody  
 day|more  
 day|irritable  
 day|more  
 day|regardless  
 day|feelings  
 day|others  
 day|more  
 day|moody  
 day|more  
 day|irritable  
 day|more  
 day|regardless  
 day|feelings  
 day|others  
 more|moody 1  
 more|more  
 more|irritable

more|more  
 more|regardless  
 more|feelings  
 more|others  
 moody|more  
 moody|irritable  
 moody|more  
 moody|regardless  
 moody|feelings 1  
 moody|others  
 more|irritable 1  
 more|more  
 more|regardless  
 more|feelings  
 more|others  
 irritable|more  
 irritable|regardless  
 irritable|feelings 1  
 irritable|others  
 more|regardless 1  
 more|feelings  
 more|others  
 regardless|feelings 1  
 regardless|others  
 feelings|others 1  
 29. i suffered myself use  
 intemperate language my  
 wife  
 i|suffered 1  
 i|myself  
 i|use  
 i|intemperate  
 i|language  
 i|my  
 i|wife  
 suffered|myself 1  
 suffered|use 1  
 suffered|intemperate  
 suffered|language  
 suffered|my  
 suffered|wife  
 myself|use  
 myself|intemperate  
 myself|language  
 myself|my  
 myself|wife

use|intemperate  
 use|language 1  
 use|my  
 use|wife  
 intemperate|language 1  
 intemperate|my  
 intemperate|wife  
 language|my  
 language|wife 1  
 my|wife 1  
 30. length i even offered her  
 personal violence  
 length|i  
 length|even 1  
 length|offered  
 length|her  
 length|personal  
 length|violence  
 i|even 1  
 i|offered  
 i|her  
 i|personal  
 i|violence  
 even|offered 1  
 even|her  
 even|personal  
 even|violence  
 offered|her 1  
 offered|personal  
 offered|violence  
 her|personal 1  
 her|violence  
 personal|violence 1  
 31. my pets course were  
 made feel change my  
 disposition  
 my|pets 1  
 my|course  
 my|were  
 my|made  
 my|feel  
 my|change  
 my|my  
 my|disposition  
 pets|course  
 pets|were

pets|made 1  
 pets|feel  
 pets|change  
 pets|my  
 pets|disposition  
 course|were  
 course|made 1  
 course|feel  
 course|change  
 course|my  
 course|disposition  
 were|made 1  
 were|feel  
 were|change  
 were|my  
 were|disposition  
 made|feel 1  
 made|change  
 made|my  
 made|disposition  
 feel|change 1  
 feel|my  
 feel|disposition  
 change|my  
 change|disposition 1  
 my|disposition  
 32. i only neglected ill used  
 them  
 i|only 1  
 i|neglected  
 i|ill  
 i|used  
 i|them  
 only|neglected 1  
 only|ill  
 only|used  
 only|them  
 neglected|ill  
 neglected|used 1  
 neglected|them  
 ill|used 1  
 ill|them  
 used|them 1  
 33. pluto i retained sufficient  
 regard restrain me  
 maltreating him i made

scruple maltreating rabbits  
 monkey even dog accident  
 through affection they came  
 my way  
 pluto|i  
 pluto|retained 1  
 pluto|sufficient  
 pluto|regard  
 pluto|restrain  
 pluto|me  
 pluto|maltreating  
 pluto|him  
 pluto|i  
 pluto|made  
 pluto|scruple  
 pluto|maltreating  
 pluto|rabbits  
 pluto|monkey  
 pluto|even  
 pluto|dog  
 pluto|accident  
 pluto|through  
 pluto|affection  
 pluto|they  
 pluto|came  
 pluto|my  
 pluto|way  
 i|retained 1  
 i|sufficient  
 i|regard  
 i|restrain  
 i|me  
 i|maltreating  
 i|him  
 i|i  
 i|made  
 i|scruple  
 i|maltreating  
 i|rabbits  
 i|monkey  
 i|even  
 i|dog  
 i|accident  
 i|through  
 i|affection  
 i|they

i|came  
 i|my  
 i|way  
 retained|sufficient  
 retained|regard 1  
 retained|restrain  
 retained|me  
 retained|maltreating  
 retained|him  
 retained|i  
 retained|made  
 retained|scruple  
 retained|maltreating  
 retained|rabbits  
 retained|monkey  
 retained|even  
 retained|dog  
 retained|accident  
 retained|through  
 retained|affection  
 retained|they  
 retained|came  
 retained|my  
 retained|way  
 sufficient|regard 1  
 sufficient|restrain  
 sufficient|me  
 sufficient|maltreating  
 sufficient|him  
 sufficient|i  
 sufficient|made  
 sufficient|scruple  
 sufficient|maltreating  
 sufficient|rabbits  
 sufficient|monkey  
 sufficient|even  
 sufficient|dog  
 sufficient|accident  
 sufficient|through  
 sufficient|affection  
 sufficient|they  
 sufficient|came  
 sufficient|my  
 sufficient|way  
 regard|restrain 1  
 regard|me

regard|maltreating  
regard|him  
regard|i  
regard|made  
regard|scruple  
regard|maltreating  
regard|rabbits  
regard|monkey  
regard|even  
regard|dog  
regard|accident  
regard|through  
regard|affection  
regard|they  
regard|came  
regard|my  
regard|way  
restrain|me 1  
restrain|maltreating 1  
restrain|him  
restrain|i  
restrain|made  
restrain|scruple  
restrain|maltreating  
restrain|rabbits  
restrain|monkey  
restrain|even  
restrain|dog  
restrain|accident  
restrain|through  
restrain|affection  
restrain|they  
restrain|came  
restrain|my  
restrain|way  
me|maltreating  
me|him  
me|i  
me|made  
me|scruple  
me|maltreating  
me|rabbits  
me|monkey  
me|even  
me|dog  
me|accident

me|through  
me|affection  
me|they  
me|came  
me|my  
me|way  
maltreating|him 1  
maltreating|i  
maltreating|made  
maltreating|scruple  
maltreating|maltreating  
maltreating|rabbits  
maltreating|monkey  
maltreating|even  
maltreating|dog  
maltreating|accident  
maltreating|through  
maltreating|affection  
maltreating|they  
maltreating|came  
maltreating|my  
maltreating|way  
him|i  
him|made  
him|scruple  
him|maltreating  
him|rabbits  
him|monkey  
him|even  
him|dog  
him|accident  
him|through  
him|affection  
him|they  
him|came  
him|my  
him|way  
i|made 1  
i|scruple  
i|maltreating  
i|rabbits  
i|monkey  
i|even  
i|dog  
i|accident  
i|through

i affection	rabbits affection
i they	rabbits they
i came	rabbits came
i my	rabbits my
i way	rabbits way
made scruple 1	monkey even
made maltreating	monkey dog
made rabbits	monkey accident
made monkey	monkey through
made even	monkey affection
made dog	monkey they
made accident	monkey came
made through	monkey my
made affection	monkey way
made they	even dog 1
made came	even accident
made my	even through
made way	even affection
scruple maltreating 1	even they
scruple rabbits	even came
scruple monkey	even my
scruple even	even way
scruple dog	dog accident 1
scruple accident	dog through
scruple through	dog affection 1
scruple affection	dog they
scruple they	dog came
scruple came	dog my
scruple my	dog way
scruple way	accident through
maltreating rabbits 1	accident affection
maltreating monkey 1	accident they
maltreating even 1	accident came
maltreating dog	accident my
maltreating accident	accident way
maltreating through	through affection 1
maltreating affection	through they
maltreating they	through came
maltreating came	through my
maltreating my	through way
maltreating way	affection they
rabbits monkey	affection came 1
rabbits even	affection my
rabbits dog	affection way
rabbits accident	they came 1
rabbits through	they my

they|way  
came|my  
came|way 1  
my|way 1  
34. my disease grew me  
disease alcohol length even  
pluto was now becoming old  
consequently somewhat  
peevish even pluto began  
experience effects my ill  
temper  
my|disease 1  
my|grew  
my|me  
my|disease  
my|alcohol  
my|length  
my|even  
my|pluto  
my|was  
my|now  
my|becoming  
my|old  
my|consequently  
my|somewhat  
my|peevish  
my|even  
my|pluto  
my|began  
my|experience  
my|effects  
my|my  
my|ill  
my|temper  
disease|grew 1  
disease|me  
disease|disease  
disease|alcohol  
disease|length  
disease|even  
disease|pluto  
disease|was  
disease|now  
disease|becoming  
disease|old  
disease|consequently

disease|somewhat  
disease|peevish  
disease|even  
disease|pluto  
disease|began  
disease|experience  
disease|effects  
disease|my  
disease|ill  
disease|temper  
grew|me 1  
grew|disease  
grew|alcohol  
grew|length  
grew|even  
grew|pluto  
grew|was  
grew|now  
grew|becoming  
grew|old  
grew|consequently  
grew|somewhat  
grew|peevish  
grew|even  
grew|pluto  
grew|began  
grew|experience  
grew|effects  
grew|my  
grew|ill  
grew|temper  
me|disease  
me|alcohol  
me|length  
me|even  
me|pluto  
me|was  
me|now  
me|becoming  
me|old  
me|consequently  
me|somewhat  
me|peevish  
me|even  
me|pluto  
me|began



me|experience  
me|effects  
me|my  
me|ill  
me|temper  
disease|alcohol 1  
disease|length  
disease|even  
disease|pluto  
disease|was  
disease|now  
disease|becoming  
disease|old  
disease|consequently  
disease|somewhat  
disease|peevish  
disease|even  
disease|pluto  
disease|began  
disease|experience  
disease|effects  
disease|my  
disease|ill  
disease|temper  
alcohol|length  
alcohol|even  
alcohol|pluto  
alcohol|was  
alcohol|now  
alcohol|becoming  
alcohol|old  
alcohol|consequently  
alcohol|somewhat  
alcohol|peevish  
alcohol|even  
alcohol|pluto  
alcohol|began  
alcohol|experience  
alcohol|effects  
alcohol|my  
alcohol|ill  
alcohol|temper  
length|even 1  
length|pluto  
length|was  
length|now

length|becoming  
length|old  
length|consequently  
length|somewhat  
length|peevish  
length|even  
length|pluto  
length|began  
length|experience  
length|effects  
length|my  
length|ill  
length|temper  
even|pluto 1  
even|was  
even|now  
even|becoming  
even|old  
even|consequently  
even|somewhat  
even|peevish  
even|even  
even|pluto  
even|began  
even|experience  
even|effects  
even|my  
even|ill  
even|temper  
pluto|was 1  
pluto|now  
pluto|becoming  
pluto|old  
pluto|consequently  
pluto|somewhat  
pluto|peevish  
pluto|even  
pluto|pluto  
pluto|began  
pluto|experience  
pluto|effects  
pluto|my  
pluto|ill  
pluto|temper  
was|now  
was|becoming 1

was|old  
was|consequently  
was|somewhat  
was|peevish  
was|even  
was|pluto  
was|began  
was|experience  
was|effects  
was|my  
was|ill  
was|temper  
now|becoming 1  
now|old  
now|consequently  
now|somewhat  
now|peevish  
now|even  
now|pluto  
now|began  
now|experience  
now|effects  
now|my  
now|ill  
now|temper  
becoming|old 1  
becoming|consequently 1  
becoming|somewhat  
becoming|peevish  
becoming|even  
becoming|pluto  
becoming|began  
becoming|experience  
becoming|effects  
becoming|my  
becoming|ill  
becoming|temper  
old|consequently  
old|somewhat  
old|peevish  
old|even  
old|pluto  
old|began  
old|experience  
old|effects  
old|my

old|ill  
old|temper  
consequently|somewhat  
consequently|peevish 1  
consequently|even  
consequently|pluto  
consequently|began  
consequently|experience  
consequently|effects  
consequently|my  
consequently|ill  
consequently|temper  
somewhat|peevish 1  
somewhat|even  
somewhat|pluto  
somewhat|began  
somewhat|experience  
somewhat|effects  
somewhat|my  
somewhat|ill  
somewhat|temper  
peevish|even  
peevish|pluto  
peevish|began  
peevish|experience  
peevish|effects  
peevish|my  
peevish|ill  
peevish|temper  
even|pluto 1  
even|began 1  
even|experience  
even|effects  
even|my  
even|ill  
even|temper  
pluto|began  
pluto|experience  
pluto|effects  
pluto|my  
pluto|ill  
pluto|temper  
began|experience 1  
began|effects  
began|my  
began|ill

began|temper  
 experience|effects 1  
 experience|my  
 experience|ill  
 experience|temper  
 effects|my  
 effects|ill  
 effects|temper 1  
 my|ill  
 my|temper  
 ill|temper 1  
 35. one night returning home  
 much intoxicated one my  
 haunts town i fancied that cat  
 avoided my presence  
 one|night 1  
 one|returning  
 one|home  
 one|much  
 one|intoxicated  
 one|one  
 one|my  
 one|haunts  
 one|town  
 one|i  
 one|fancied  
 one|that  
 one|cat  
 one|avoided  
 one|my  
 one|presence  
 night|returning 1  
 night|home  
 night|much  
 night|intoxicated  
 night|one  
 night|my  
 night|haunts  
 night|town  
 night|i  
 night|fancied  
 night|that  
 night|cat  
 night|avoided  
 night|my  
 night|presence

returning|home 1  
 returning|much  
 returning|intoxicated 1  
 returning|one  
 returning|my  
 returning|haunts 1  
 returning|town  
 returning|i  
 returning|fancied  
 returning|that  
 returning|cat  
 returning|avoided  
 returning|my  
 returning|presence  
 home|much  
 home|intoxicated  
 home|one  
 home|my  
 home|haunts  
 home|town  
 home|i  
 home|fancied  
 home|that  
 home|cat  
 home|avoided  
 home|my  
 home|presence  
 much|intoxicated 1  
 much|one  
 much|my  
 much|haunts  
 much|town  
 much|i  
 much|fancied  
 much|that  
 much|cat  
 much|avoided  
 much|my  
 much|presence  
 intoxicated|one  
 intoxicated|my  
 intoxicated|haunts  
 intoxicated|town  
 intoxicated|i  
 intoxicated|fancied  
 intoxicated|that

intoxicated|cat  
intoxicated|avoided  
intoxicated|my  
intoxicated|presence  
one|my  
one|haunts  
one|town  
one|i  
one|fancied  
one|that  
one|cat  
one|avoided  
one|my  
one|presence  
my|haunts 1  
my|town  
my|i  
my|fancied  
my|that  
my|cat  
my|avoided  
my|my  
my|presence  
haunts|town 1  
haunts|i  
haunts|fancied 1  
haunts|that  
haunts|cat  
haunts|avoided  
haunts|my  
haunts|presence  
town|i  
town|fancied  
town|that  
town|cat  
town|avoided  
town|my  
town|presence  
i|fancied 1  
i|that  
i|cat  
i|avoided  
i|my  
i|presence  
fancied|that  
fancied|cat

fancied|avoided 1  
fancied|my  
fancied|presence  
that|cat  
that|avoided  
that|my  
that|presence  
cat|avoided 1  
cat|my  
cat|presence  
avoided|my  
avoided|presence 1  
my|presence  
36. i seized him his fright my  
violence he inflicted slight  
wound my hand his teeth  
i|seized 1  
i|him  
i|his  
i|fright  
i|my  
i|violence  
i|he  
i|inflicted  
i|slight  
i|wound  
i|my  
i|hand  
i|his  
i|teeth  
seized|him 1  
seized|his  
seized|fright  
seized|my  
seized|violence  
seized|he  
seized|inflicted  
seized|slight  
seized|wound  
seized|my  
seized|hand  
seized|his  
seized|teeth  
him|his  
him|fright  
him|my

him|violence  
him|he  
him|inflicted  
him|slight  
him|wound  
him|my  
him|hand  
him|his  
him|teeth  
his|fright 1  
his|my  
his|violence  
his|he  
his|inflicted  
his|slight  
his|wound  
his|my  
his|hand  
his|his  
his|teeth  
fright|my  
fright|violence 1  
fright|he  
fright|inflicted 1  
fright|slight  
fright|wound  
fright|my  
fright|hand  
fright|his  
fright|teeth  
my|violence 1  
my|he  
my|inflicted  
my|slight  
my|wound  
my|my  
my|hand  
my|his  
my|teeth  
violence|he  
violence|inflicted  
violence|slight  
violence|wound  
violence|my  
violence|hand  
violence|his

violence|teeth  
he|inflicted 1  
he|slight  
he|wound  
he|my  
he|hand  
he|his  
he|teeth  
inflicted|slight  
inflicted|wound 1  
inflicted|my  
inflicted|hand  
inflicted|his  
inflicted|teeth  
slight|wound 1  
slight|my  
slight|hand  
slight|his  
slight|teeth  
wound|my  
wound|hand 1  
wound|his  
wound|teeth  
my|hand 1  
my|his  
my|teeth  
hand|his  
hand|teeth 1  
his|teeth 1  
37. fury demon instantly  
possessed me  
fury|demon 1  
fury|instantly  
fury|possessed  
fury|me  
demon|instantly  
demon|possessed 1  
demon|me  
instantly|possessed 1  
instantly|me  
possessed|me 1  
38. i knew myself longer  
i|knew 1  
i|myself  
i|longer  
knew|myself

knew|longer 1  
myself|longer  
39. my original soul seemed  
once take its flight my body  
more fiendish malevolence  
gin nurtured thrilled every  
fibre my frame  
my|original 1  
my|soul  
my|seemed  
my|once  
my|take  
my|its  
my|flight  
my|my  
my|body  
my|more  
my|fiendish  
my|malevolence  
my|gin  
my|nurtured  
my|thrilled  
my|every  
my|fibre  
my|my  
my|frame  
original|soul 1  
original|seemed  
original|once  
original|take  
original|its  
original|flight  
original|my  
original|body  
original|more  
original|fiendish  
original|malevolence  
original|gin  
original|nurtured  
original|thrilled  
original|every  
original|fibre  
original|my  
original|frame  
soul|seemed 1  
soul|once

soul|take  
soul|its  
soul|flight  
soul|my  
soul|body  
soul|more  
soul|fiendish  
soul|malevolence  
soul|gin  
soul|nurtured  
soul|thrilled  
soul|every  
soul|fibre  
soul|my  
soul|frame  
seemed|once 1  
seemed|take  
seemed|its  
seemed|flight  
seemed|my  
seemed|body  
seemed|more  
seemed|fiendish  
seemed|malevolence  
seemed|gin  
seemed|nurtured  
seemed|thrilled  
seemed|every  
seemed|fibre  
seemed|my  
seemed|frame  
once|take 1  
once|its  
once|flight  
once|my  
once|body  
once|more  
once|fiendish  
once|malevolence  
once|gin  
once|nurtured  
once|thrilled  
once|every  
once|fibre  
once|my  
once|frame

take|its 1  
take|flight 1  
take|my  
take|body  
take|more  
take|fiendish  
take|malevolence  
take|gin  
take|nurtured  
take|thrilled  
take|every  
take|fibre  
take|my  
take|frame  
its|flight  
its|my  
its|body  
its|more  
its|fiendish  
its|malevolence  
its|gin  
its|nurtured  
its|thrilled  
its|every  
its|fibre  
its|my  
its|frame  
flight|my  
flight|body 1  
flight|more  
flight|fiendish  
flight|malevolence  
flight|gin  
flight|nurtured  
flight|thrilled  
flight|every  
flight|fibre  
flight|my  
flight|frame  
my|body 1  
my|more  
my|fiendish  
my|malevolence  
my|gin  
my|nurtured  
my|thrilled

my|every  
my|fibre  
my|my  
my|frame  
body|more  
body|fiendish  
body|malevolence  
body|gin  
body|nurtured  
body|thrilled  
body|every  
body|fibre  
body|my  
body|frame  
more|fiendish 1  
more|malevolence  
more|gin  
more|nurtured  
more|thrilled  
more|every  
more|fibre  
more|my  
more|frame  
fiendish|malevolence 1  
fiendish|gin  
fiendish|nurtured  
fiendish|thrilled  
fiendish|every  
fiendish|fibre  
fiendish|my  
fiendish|frame  
malevolence|gin 1  
malevolence|nurtured  
malevolence|thrilled 1  
malevolence|every  
malevolence|fibre  
malevolence|my  
malevolence|frame  
gin|nurtured 1  
gin|thrilled  
gin|every  
gin|fibre  
gin|my  
gin|frame  
nurtured|thrilled  
nurtured|every

nurtured|fibre  
nurtured|my  
nurtured|frame  
thrilled|every 1  
thrilled|fibre  
thrilled|my  
thrilled|frame  
every|fibre 1  
every|my  
every|frame  
fibre|my  
fibre|frame 1  
my|frame  
40. i took my waistcoat  
pocket pen knife opened it  
grasped poor beast throat  
deliberately cut one its eyes  
socket  
i|took 1  
i|my  
i|waistcoat  
i|pocket  
i|pen  
i|knife  
i|opened  
i|it  
i|grasped  
i|poor  
i|beast  
i|throat  
i|deliberately  
i|cut  
i|one  
i|its  
i|eyes  
i|socket  
took|my  
took|waistcoat 1  
took|pocket  
took|pen  
took|knife  
took|opened  
took|it  
took|grasped  
took|poor  
took|beast

took|throat  
took|deliberately  
took|cut  
took|one  
took|its  
took|eyes  
took|socket  
my|waistcoat 1  
my|pocket  
my|pen  
my|knife  
my|opened  
my|it  
my|grasped  
my|poor  
my|beast  
my|throat  
my|deliberately  
my|cut  
my|one  
my|its  
my|eyes  
my|socket  
waistcoat|pocket 1  
waistcoat|pen  
waistcoat|knife  
waistcoat|opened  
waistcoat|it  
waistcoat|grasped  
waistcoat|poor  
waistcoat|beast  
waistcoat|throat  
waistcoat|deliberately  
waistcoat|cut  
waistcoat|one  
waistcoat|its  
waistcoat|eyes  
waistcoat|socket  
pocket|pen  
pocket|knife 1  
pocket|opened  
pocket|it  
pocket|grasped  
pocket|poor  
pocket|beast  
pocket|throat



pocket|deliberately  
pocket|cut  
pocket|one  
pocket|its  
pocket|eyes  
pocket|socket  
pen|knife 1  
pen|opened  
pen|it  
pen|grasped  
pen|poor  
pen|beast  
pen|throat  
pen|deliberately  
pen|cut  
pen|one  
pen|its  
pen|eyes  
pen|socket  
knife|opened 1  
knife|it  
knife|grasped  
knife|poor  
knife|beast  
knife|throat  
knife|deliberately  
knife|cut  
knife|one  
knife|its  
knife|eyes  
knife|socket  
opened|it 1  
opened|grasped  
opened|poor  
opened|beast  
opened|throat  
opened|deliberately  
opened|cut  
opened|one  
opened|its  
opened|eyes  
opened|socket  
it|grasped  
it|poor  
it|beast  
it|throat

it|deliberately  
it|cut  
it|one  
it|its  
it|eyes  
it|socket  
grasped|poor  
grasped|beast 1  
grasped|throat 1  
grasped|deliberately  
grasped|cut  
grasped|one  
grasped|its  
grasped|eyes  
grasped|socket  
poor|beast 1  
poor|throat  
poor|deliberately  
poor|cut  
poor|one  
poor|its  
poor|eyes  
poor|socket  
beast|throat  
beast|deliberately  
beast|cut  
beast|one  
beast|its  
beast|eyes  
beast|socket  
throat|deliberately  
throat|cut 1  
throat|one  
throat|its  
throat|eyes  
throat|socket  
deliberately|cut 1  
deliberately|one  
deliberately|its  
deliberately|eyes  
deliberately|socket  
cut|one  
cut|its  
cut|eyes 1  
cut|socket  
one|its

one|eyes  
one|socket  
its|eyes  
its|socket  
eyes|socket 1  
41. i blush i burn i shudder i  
pen damnable atrocity  
i|blush 1  
i|i  
i|burn  
i|i  
i|shudder  
i|i  
i|pen  
i|damnable  
i|atrocity  
blush|i  
blush|burn 1  
blush|i  
blush|shudder  
blush|i  
blush|pen  
blush|damnable  
blush|atrocity  
i|burn 1  
i|i  
i|shudder  
i|i  
i|pen  
i|damnable  
i|atrocity  
burn|i  
burn|shudder 1  
burn|i  
burn|pen  
burn|damnable  
burn|atrocity  
i|shudder 1  
i|i  
i|pen  
i|damnable  
i|atrocity  
shudder|i  
shudder|pen 1  
shudder|damnable  
shudder|atrocity

i|pen 1  
i|damnable  
i|atrocity  
pen|damnable  
pen|atrocity 1  
damnable|atrocity 1  
42. reason returned morning  
i had slept fumes night is  
debauch i experienced  
sentiment half horror half  
remorse crime i had guilty it  
was best feeble equivocal  
feeling soul remained  
untouched  
reason|returned 1  
reason|morning 1  
reason|i  
reason|had  
reason|slept  
reason|fumes  
reason|night  
reason|is  
reason|debauch  
reason|i  
reason|experienced  
reason|sentiment  
reason|half  
reason|horror  
reason|half  
reason|remorse  
reason|crime  
reason|i  
reason|had  
reason|guilty  
reason|it  
reason|was  
reason|best  
reason|feeble  
reason|equivocal  
reason|feeling  
reason|soul  
reason|remained  
reason|untouched  
returned|morning  
returned|i  
returned|had

returned|slept  
 returned|fumes  
 returned|night  
 returned|is  
 returned|debauch  
 returned|i  
 returned|experienced  
 returned|sentiment  
 returned|half  
 returned|horror  
 returned|half  
 returned|remorse  
 returned|crime  
 returned|i  
 returned|had  
 returned|guilty  
 returned|it  
 returned|was  
 returned|best  
 returned|feeble  
 returned|equivocal  
 returned|feeling  
 returned|soul  
 returned|remained  
 returned|untouched  
 morning|i  
 morning|had  
 morning|slept  
 morning|fumes  
 morning|night  
 morning|is  
 morning|debauch  
 morning|i  
 morning|experienced  
 morning|sentiment  
 morning|half  
 morning|horror  
 morning|half  
 morning|remorse  
 morning|crime  
 morning|i  
 morning|had  
 morning|guilty  
 morning|it  
 morning|was  
 morning|best

morning|feeble  
 morning|equivocal  
 morning|feeling  
 morning|soul  
 morning|remained  
 morning|untouched  
 i|had 1  
 i|slept  
 i|fumes  
 i|night  
 i|is  
 i|debauch  
 i|i  
 i|experienced  
 i|sentiment  
 i|half  
 i|horror  
 i|half  
 i|remorse  
 i|crime  
 i|i  
 i|had  
 i|guilty  
 i|it  
 i|was  
 i|best  
 i|feeble  
 i|equivocal  
 i|feeling  
 i|soul  
 i|remained  
 i|untouched  
 had|slept 1  
 had|fumes  
 had|night  
 had|is  
 had|debauch  
 had|i  
 had|experienced  
 had|sentiment  
 had|half  
 had|horror  
 had|half  
 had|remorse  
 had|crime  
 had|i

had|had  
had|guilty  
had|it  
had|was  
had|best  
had|feeble  
had|equivocal  
had|feeling  
had|soul  
had|remained  
had|untouched  
slept|fumes 1  
slept|night  
slept|is  
slept|debauch  
slept|i  
slept|experienced  
slept|sentiment  
slept|half  
slept|horror  
slept|half  
slept|remorse  
slept|crime  
slept|i  
slept|had  
slept|guilty  
slept|it  
slept|was  
slept|best  
slept|feeble  
slept|equivocal  
slept|feeling  
slept|soul  
slept|remained  
slept|untouched  
fumes|night  
fumes|is  
fumes|debauch 1  
fumes|i  
fumes|experienced  
fumes|sentiment  
fumes|half  
fumes|horror  
fumes|half  
fumes|remorse  
fumes|crime

fumes|i  
fumes|had  
fumes|guilty  
fumes|it  
fumes|was  
fumes|best  
fumes|feeble  
fumes|equivocal  
fumes|feeling  
fumes|soul  
fumes|remained  
fumes|untouched  
night|is 1  
night|debauch  
night|i  
night|experienced  
night|sentiment  
night|half  
night|horror  
night|half  
night|remorse  
night|crime  
night|i  
night|had  
night|guilty  
night|it  
night|was  
night|best  
night|feeble  
night|equivocal  
night|feeling  
night|soul  
night|remained  
night|untouched  
is|debauch 1  
is|i  
is|experienced  
is|sentiment  
is|half  
is|horror  
is|half  
is|remorse  
is|crime  
is|i  
is|had  
is|guilty

is|it  
is|was  
is|best  
is|feeble  
is|equivocal  
is|feeling  
is|soul  
is|remained  
is|untouched  
debauch|i  
debauch|experienced  
debauch|sentiment  
debauch|half  
debauch|horror  
debauch|half  
debauch|remorse  
debauch|crime  
debauch|i  
debauch|had  
debauch|guilty  
debauch|it  
debauch|was  
debauch|best  
debauch|feeble  
debauch|equivocal  
debauch|feeling  
debauch|soul  
debauch|remained  
debauch|untouched  
i|experienced 1  
i|sentiment  
i|half  
i|horror  
i|half  
i|remorse  
i|crime  
i|i  
i|had  
i|guilty  
i|it  
i|was  
i|best  
i|feeble  
i|equivocal  
i|feeling  
i|soul

i|remained  
i|untouched  
experienced|sentiment 1  
experienced|half  
experienced|horror  
experienced|half  
experienced|remorse  
experienced|crime  
experienced|i  
experienced|had  
experienced|guilty  
experienced|it  
experienced|was  
experienced|best  
experienced|feeble  
experienced|equivocal  
experienced|feeling  
experienced|soul  
experienced|remained  
experienced|untouched  
sentiment|half  
sentiment|horror 1  
sentiment|half  
sentiment|remorse 1  
sentiment|crime  
sentiment|i  
sentiment|had  
sentiment|guilty  
sentiment|it  
sentiment|was 1  
sentiment|best 1  
sentiment|feeble 1  
sentiment|equivocal  
sentiment|feeling 1  
sentiment|soul  
sentiment|remained  
sentiment|untouched  
half|horror 1  
half|half  
half|remorse  
half|crime  
half|i  
half|had  
half|guilty  
half|it  
half|was

half|best  
half|feeble  
half|equivocal  
half|feeling  
half|soul  
half|remained  
half|untouched  
horror|half  
horror|remorse  
horror|crime  
horror|i  
horror|had  
horror|guilty  
horror|it  
horror|was  
horror|best  
horror|feeble  
horror|equivocal  
horror|feeling  
horror|soul  
horror|remained  
horror|untouched  
half|remorse 1  
half|crime  
half|i  
half|had  
half|guilty  
half|it  
half|was  
half|best  
half|feeble  
half|equivocal  
half|feeling  
half|soul  
half|remained  
half|untouched  
remorse|crime 1  
remorse|i  
remorse|had  
remorse|guilty  
remorse|it  
remorse|was  
remorse|best  
remorse|feeble  
remorse|equivocal  
remorse|feeling

remorse|soul  
remorse|remained  
remorse|untouched  
crime|i  
crime|had  
crime|guilty 1  
crime|it  
crime|was  
crime|best  
crime|feeble  
crime|equivocal  
crime|feeling  
crime|soul  
crime|remained  
crime|untouched  
i|had 1  
i|guilty  
i|it  
i|was  
i|best  
i|feeble  
i|equivocal  
i|feeling  
i|soul  
i|remained  
i|untouched  
had|guilty 1  
had|it  
had|was  
had|best  
had|feeble  
had|equivocal  
had|feeling  
had|soul  
had|remained  
had|untouched  
guilty|it  
guilty|was  
guilty|best  
guilty|feeble  
guilty|equivocal  
guilty|feeling  
guilty|soul  
guilty|remained  
guilty|untouched  
it|was

it|best  
 it|feeble  
 it|equivocal  
 it|feeling  
 it|soul  
 it|remained  
 it|untouched  
 was|best  
 was|feeble  
 was|equivocal  
 was|feeling  
 was|soul  
 was|remained  
 was|untouched  
 best|feeble  
 best|equivocal  
 best|feeling  
 best|soul  
 best|remained  
 best|untouched  
 feeble|equivocal  
 feeble|feeling 1  
 feeble|soul  
 feeble|remained  
 feeble|untouched  
 equivocal|feeling  
 equivocal|soul  
 equivocal|remained  
 equivocal|untouched  
 feeling|soul  
 feeling|remained  
 feeling|untouched  
 soul|remained 1  
 soul|untouched  
 remained|untouched 1  
 43. i again plunged excess  
 soon drowned wine all  
 memory deed  
 i|again 1  
 i|plunged  
 i|excess  
 i|soon  
 i|drowned  
 i|wine  
 i|all  
 i|memory

i|deed  
 again|plunged 1  
 again|excess  
 again|soon  
 again|drowned  
 again|wine  
 again|all  
 again|memory  
 again|deed  
 plunged|excess 1  
 plunged|soon  
 plunged|drowned  
 plunged|wine  
 plunged|all  
 plunged|memory  
 plunged|deed  
 excess|soon 1  
 excess|drowned  
 excess|wine  
 excess|all  
 excess|memory  
 excess|deed  
 soon|drowned 1  
 soon|wine  
 soon|all  
 soon|memory  
 soon|deed  
 drowned|wine 1  
 drowned|all  
 drowned|memory  
 drowned|deed  
 wine|all  
 wine|memory 1  
 wine|deed  
 all|memory 1  
 all|deed  
 memory|deed 1  
 44. meantime cat slowly  
 recovered  
 meantime|cat  
 meantime|slowly  
 meantime|recovered 1  
 cat|slowly 1  
 cat|recovered  
 slowly|recovered 1  
 45. socket lost eye presented

it true frightful appearance  
he longer appeared suffer  
any pain  
socket|lost  
socket|eye 1  
socket|presented 1  
socket|it  
socket|true  
socket|frightful  
socket|appearance  
socket|he  
socket|longer  
socket|appeared  
socket|suffer  
socket|any  
socket|pain  
lost|eye 1  
lost|presented  
lost|it  
lost|true  
lost|frightful  
lost|appearance  
lost|he  
lost|longer  
lost|appeared  
lost|suffer  
lost|any  
lost|pain  
eye|presented 1  
eye|it  
eye|true  
eye|frightful  
eye|appearance  
eye|he  
eye|longer  
eye|appeared  
eye|suffer  
eye|any  
eye|pain  
presented|it  
presented|true  
presented|frightful  
presented|appearance 1  
presented|he  
presented|longer  
presented|appeared

presented|suffer  
presented|any  
presented|pain  
it|true 1  
it|frightful  
it|appearance  
it|he  
it|longer  
it|appeared  
it|suffer  
it|any  
it|pain  
true|frightful  
true|appearance  
true|he  
true|longer  
true|appeared  
true|suffer  
true|any  
true|pain  
frightful|appearance 1  
frightful|he  
frightful|longer  
frightful|appeared  
frightful|suffer  
frightful|any  
frightful|pain  
appearance|he  
appearance|longer  
appearance|appeared 1  
appearance|suffer  
appearance|any  
appearance|pain  
he|longer  
he|appeared  
he|suffer  
he|any  
he|pain  
longer|appeared 1  
longer|suffer  
longer|any  
longer|pain  
appeared|suffer 1  
appeared|any  
appeared|pain  
suffer|any



suffer|pain 1  
any|pain  
46. he went house usual  
expected fled extreme terror  
my approach  
he|went 1  
he|house  
he|usual  
he|expected  
he|fled  
he|extreme  
he|terror  
he|my  
he|approach  
went|house 1  
went|usual  
went|expected  
went|fled  
went|extreme  
went|terror  
went|my  
went|approach  
house|usual 1  
house|expected 1  
house|fled  
house|extreme  
house|terror  
house|my  
house|approach  
usual|expected  
usual|fled  
usual|extreme  
usual|terror  
usual|my  
usual|approach  
expected|fled  
expected|extreme  
expected|terror  
expected|my  
expected|approach 1  
fled|extreme  
fled|terror 1  
fled|my  
fled|approach  
extreme|terror 1  
extreme|my

extreme|approach  
terror|my  
terror|approach  
my|approach  
47. i had much my old heart  
left first grieved this evident  
dislike part creature had once  
loved me  
i|had 1  
i|much  
i|my  
i|old  
i|heart  
i|left  
i|first  
i|grieved  
i|this  
i|evident  
i|dislike  
i|part  
i|creature  
i|had  
i|once  
i|loved  
i|me  
had|much  
had|my  
had|old  
had|heart 1  
had|left  
had|first  
had|grieved  
had|this  
had|evident  
had|dislike  
had|part  
had|creature  
had|had  
had|once  
had|loved  
had|me  
much|my  
much|old  
much|heart 1  
much|left  
much|first

much|grieved  
much|this  
much|evident  
much|dislike  
much|part  
much|creature  
much|had  
much|once  
much|loved  
much|me  
my|old 1  
my|heart  
my|left  
my|first  
my|grieved  
my|this  
my|evident  
my|dislike  
my|part  
my|creature  
my|had  
my|once  
my|loved  
my|me  
old|heart 1  
old|left  
old|first  
old|grieved  
old|this  
old|evident  
old|dislike  
old|part  
old|creature  
old|had  
old|once  
old|loved  
old|me  
heart|left 1  
heart|first  
heart|grieved  
heart|this  
heart|evident  
heart|dislike  
heart|part  
heart|creature  
heart|had

heart|once  
heart|loved  
heart|me  
left|first  
left|grieved 1  
left|this  
left|evident  
left|dislike  
left|part  
left|creature  
left|had  
left|once  
left|loved  
left|me  
first|grieved 1  
first|this  
first|evident  
first|dislike  
first|part  
first|creature  
first|had  
first|once  
first|loved  
first|me  
grieved|this  
grieved|evident  
grieved|dislike 1  
grieved|part  
grieved|creature 1  
grieved|had  
grieved|once  
grieved|loved  
grieved|me  
this|evident  
this|dislike  
this|part  
this|creature  
this|had  
this|once  
this|loved  
this|me  
evident|dislike 1  
evident|part  
evident|creature  
evident|had  
evident|once

evident|loved  
evident|me  
dislike|part 1  
dislike|creature  
dislike|had  
dislike|once  
dislike|loved  
dislike|me  
part|creature 1  
part|had  
part|once  
part|loved  
part|me  
creature|had  
creature|once  
creature|loved 1  
creature|me  
had|once 1  
had|loved  
had|me  
once|loved 1  
once|me  
loved|me 1  
48. this feeling soon gave  
place irritation  
this|feeling  
this|soon  
this|gave  
this|place  
this|irritation  
feeling|soon  
feeling|gave  
feeling|place  
feeling|irritation  
soon|gave  
soon|place  
soon|irritation  
gave|place  
gave|irritation  
place|irritation  
49. came if my final  
irrevocable overthrow spirit  
perverseness  
came|if 1  
came|my  
came|final

came|irrevocable  
came|overthrow  
came|spirit  
came|perverseness  
if|my  
if|final 1  
if|irrevocable  
if|overthrow  
if|spirit  
if|perverseness  
my|final 1  
my|irrevocable  
my|overthrow  
my|spirit  
my|perverseness  
final|irrevocable 1  
final|overthrow  
final|spirit  
final|perverseness  
irrevocable|overthrow 1  
irrevocable|spirit  
irrevocable|perverseness  
overthrow|spirit  
overthrow|perverseness 1  
spirit|perverseness 1  
50. this spirit philosophy  
takes account  
this|spirit  
this|philosophy  
this|takes  
this|account  
spirit|philosophy 1  
spirit|takes  
spirit|account  
philosophy|takes 1  
philosophy|account  
takes|account 1