

UNA PROPUESTA METODOLÓGICA PARA EL ESTUDIO DE COLOCACIONES EN UN CORPUS NO ETIQUETADO Y ESPECIALIZADO

A METHODOLOGICAL PROPOSAL FOR THE STUDY OF COLLOCATIONS IN AN UNTAGGED AND SPECIALIZED CORPUS

ELEONORA LOZANO BACHIOQUI
Universidad Autónoma de Baja California
eleonoralozano@uabc.edu.mx

Recibido: 02/09/2017

Aceptado: 07/10/2017

Resumen

El presente trabajo tiene como objetivo presentar una propuesta metodológica guiada por corpus para el estudio de unidades fraseológicas, específicamente las colocaciones, en los lenguajes de especialidad. Las unidades fraseológicas se identifican a través de los verbos clave de un área determinada en un corpus no etiquetado, especializado y con fines específicos. La metodología aquí propuesta se ha probado en la identificación de unidades fraseológicas en un corpus legal perteneciente al género del derecho civil mexicano. Para ello, se explican las diferentes etapas del proceso, asimismo se proporciona un resumen de los resultados obtenidos. La metodología propuesta resulta útil para la enseñanza-aprendizaje del español para propósitos específicos, la traducción legal y la formación de profesionales en derecho, así como cualquier otra persona interesada en el tema.

PALABRAS CLAVE: Lingüística de corpus, fraseología, colocaciones, lenguajes especializados, lenguaje jurídico.

Abstract

The aim of this paper is to present a corpus driven approach to study phraseological units, collocations in particular, in specialized languages. Phraseological units are identified through key verbs of a specific specialized language in an untagged specialized corpus, and for specific purposes. The methodology presented here has been tested for the identification of phraseological units on a legal corpus pertaining to Mexican civil law. The different stages of the process are explained and a summary of results is provided. The method is useful for teaching and learning Spanish for specific purposes, legal translation, legal education, and everyone interested in this subject.

KEYWORDS: Corpus linguistics, phraseology, collocations, specialized languages, legal language.

Para citar este artículo / To cite this article: Lozano Bachioqui, Eleonora (2017). Una propuesta metodológica para el estudio de colocaciones en un corpus no etiquetado y especializado. *ELUA*, 31: 233-250. doi: 10.14198/ELUA2017.31.12

Enlace / Link: <http://dx.doi.org/10.14198/ELUA2017.31.12>

1. INTRODUCCIÓN

La importancia de los lenguajes especializados o de especialidad está hoy en día fuera de toda duda. De hecho, el número de especialistas dedicados a su descripción y aplicación crece continuamente. Esto se debe a tres razones principalmente: en primer lugar, el papel que juega la lingüística aplicada en el marco de la lingüística general; en segundo lugar, las necesidades sociales en materia de plurilingüismo y, en tercer lugar, la importancia que la sociedad actual concede a las especialidades (Cabré, 2004).

Los estudios de los lenguajes especializados, entendidos como discursos especializados, han puesto de relieve la importancia de las colocaciones como elementos caracterizadores de un registro determinado. Por ejemplo, las colocaciones de la palabra *expediente* con verbos como *incoar*, *instruir*, *tramitar* o *archivar* ponen de manifiesto su pertenencia al registro jurídico y, dentro de este, al ámbito judicial (Corpas Pastor, 2001).

Además, L'Homme (2009) señala que desde 1980 se ha reconocido que las colocaciones resultan extremadamente útiles en la microestructura de los diccionarios de la lengua general, así como en los diccionarios especializados pues el acceso a las colocaciones ayuda a resolver problemas relacionados con la producción correcta de unidades léxicas y su uso con la combinación de un término en particular u otra unidad léxica. Así, las colocaciones resultan importantes en la aproximación a los lenguajes especializados, pero ¿cómo identificar y estudiar estas unidades?

Este trabajo muestra una propuesta metodológica guiada por corpus (McEnery y Hardie, 2011) para identificar, así como estudiar las colocaciones, a partir de los verbos clave, en un corpus no etiquetado, especializado y con fines específicos. El método aquí propuesto se ha probado en una investigación sobre unidades fraseológicas del lenguaje jurídico en el derecho civil mexicano. Así, el objetivo del presente trabajo es presentar la metodología, describir las diferentes etapas, así como explicar y justificar las decisiones tomadas en cada etapa. De esta manera, apoyar tanto a traductores e intérpretes, como a personas que se dedican a la enseñanza-aprendizaje de lenguajes especializados y cualquier persona interesada en el tema a resolver problemas relacionados con la correcta producción oral y escrita, así como la combinación correcta de un término en particular en el contexto de los lenguajes especializados.

2. LAS COLOCACIONES

Uno de los primeros lingüistas que introdujo la noción de significado por colocación fue Firth, quien estableció que la colocación es una combinación de palabras que muestra una atracción mutua (1957: 12-13). Firth creía que se podía conocer a una palabra por las palabras que la acompañan, es decir, el significado y el uso de una palabra se caracteriza por sus colocados. Desafortunadamente, el concepto permaneció vago durante algún tiempo y no fue sino hasta que apareció la lingüística de corpus que el concepto se difundió con trabajos como los de Sinclair (1991).

Para Sinclair (2004:141-142), la colocación es la coocurrencia de palabras en un espacio de no más de cuatro palabras y la describe en dos dimensiones: la sintagmática y la paradigmática. En la primera, la colocación es la relación más obvia y simple y está claramente descrita. En la segunda, se define de manera diferente porque los elementos

sólo se pueden colocar uno con el otro cuando están en un texto y dos elementos en un paradigma están clasificados como mutuamente excluyentes precisamente por la disposición de los mismos.

Corpas Pastor (2001: 101) establece que las colocaciones son un tipo de unidad fraseológica y que están formadas generalmente por dos unidades léxicas que suelen aparecer juntas en el discurso. Una de ellas constituye la base, que es la que determina la elección de la segunda, y la otra, el colocativo. Asimismo, la autora (2003: 185-186) la define como la tendencia de los hablantes a producir ciertas combinaciones de palabras entre una gran cantidad de combinaciones teóricamente posibles, que se caracteriza por la alta frecuencia de coaparición de sus elementos constitutivos o colocados, así como por presentar una gran variedad de patrones sintácticos y diversos grados de restricción interna e institucionalización. Según la misma autora, las colocaciones “constituyen la piedra angular de toda traducción, ya que son las responsables de la idiomática y naturalidad de la misma” (2003: p. 185).

Por su parte, Alonso Ramos señala que “las colocaciones pueden ser más o menos impredecibles, más o menos transparentes y más o menos restringidas semánticamente” (2010: 63). Además, argumenta que desde su perspectiva “no hay problema en incluir bajo el mismo término de *colocación* sintagmas completamente transparentes como *muy cansado* hasta sintagmas completamente opacos como *armado hasta los dientes*” (2010: 58). De acuerdo con el autor, la diferencia entre los sintagmas libres y las colocaciones es que estas últimas violan la propiedad paradigmática, que concierne a la selección; mientras que conservan la propiedad sintagmática, que concierne a la composicionalidad. Asimismo, establece que las colocaciones son importantes pues el uso incorrecto de estas delata a las personas no nativas (2010: 63, 57).

Koike, agrega que las colocaciones se distinguen de las locuciones pues las últimas “se caracterizan por su sentido idiomático en el marco de una estructura fijada” (2001: 31). Además, menciona que “la composicionalidad semántica es relativa en la colocación, mientras que la locución carece de ella” (2001: 35). Es decir, como establece Alonso Ramos (1994: 16), en las colocaciones cada constituyente aporta una parte de sentido a la expresión conjunta. Sin embargo, en las locuciones el sentido total de la expresión no refleja el sentido de sus partes constituyentes.

Además de Corpas Pastor (1996, 2001, 2003), Alonso Ramos (1994, 2010), Koike (2001), otros autores han profundizado sobre el concepto de colocación en el ámbito en español (Bosque, 2001) y en el ámbito en inglés (Stubbs, 2002; Hoey, 2005, McEnery y Hardie, 2011). Sin embargo, este es uno de los conceptos más controvertidos en lingüística y, a pesar de las referencias bibliográficas existentes, algunos estudiosos señalan que, de hecho, la colocación dista mucho de tener una definición aceptada de manera unánime (Alonso Ramos, 2010: 55).

En este trabajo se entenderá como colocación la relación de coocurrencia sistemática entre dos o más palabras, no lemas, a una distancia de hasta cinco palabras y que, debido a su composicionalidad semántica, se caracteriza por tener un significado relativamente transparente.

Existen dos tipos de colocaciones: las colocaciones léxicas y las gramaticales. Las colocaciones léxicas se distinguen porque no tienen una palabra dominante y presentan estructuras como las siguientes: verbo + sustantivo, adjetivo + sustantivo, sustantivo + verbo, sustantivo + sustantivo, adverbio + adjetivo y adverbio + verbo.

Por otro lado, las colocaciones gramaticales consisten de una palabra dominante (un sustantivo, adjetivo, participio o verbo) y una preposición o construcción gramatical (Benson, Benson, e Ilson, 2010: xiii). Están formadas por una unidad léxica y otra gramatical. En este trabajo se consideran solo aquellas colocaciones entre cuyos elementos se encuentra un verbo.

3. PROPUESTA METODOLÓGICA PARA LA IDENTIFICACIÓN DE LAS COLOCACIONES

Gries (2008: 20) menciona que los procedimientos más completos para la identificación de unidades fraseológicas se encuentran en la lingüística de corpus. La propuesta que aquí se presenta es de alcance descriptivo y está fundamentada precisamente en la metodología de corpus. De acuerdo con McEnery y Hardie (2011: 6), existen dos tipos de trabajos: los tipo *corpus-based* y los *corpus-driven*, en inglés, o basados en corpus y guiados por corpus, en español. Los primeros, los basados en corpus, utilizan corpus para estudiar una teoría o hipótesis que ya había sido previamente establecida en la literatura, con el fin de refinarla, aceptarla o rechazarla. Estos trabajos basados en corpus utilizan el corpus como método. Los segundos, los guiados por corpus, consideran que el corpus, más que un método, es la única fuente de una hipótesis sobre la lengua. Por lo tanto, afirman que el corpus *per se* encarna su propia teoría del lenguaje. De aquí, el método que se propone es guiado por corpus pues más que confirmar una hipótesis o una teoría previamente establecida el método tiene por objeto observar las colocaciones.

3.1. Primera etapa: la selección de textos para el corpus (la muestra)

Existen distintos tipos de corpus según el objeto de estudio. Para la identificación y el análisis de colocaciones en lenguajes especializados, se propone un corpus no etiquetado, especializado y con fines específicos. Los corpus especializados son un conjunto de textos de una clase en particular que se utilizan para investigar un tipo de lenguaje en específico. Los investigadores compilan corpus especializados para reflejar la variedad del lenguaje que desean investigar (Hunston, 2002: 14). Los corpus con fines específicos o mini corpus son corpus pequeños que son útiles para estudiar aspectos específicos de la lengua, como la extracción de la terminología o el estudio de registro y estilo (Maia, 2002: 223). Maia (2002: 223) establece que, aunque el número de un millón de palabras se ha convertido en un número atractivo para muchos corpus, se ha realizado trabajo interesante con corpus mucho más pequeños que un millón de palabras.

Para el estudio de unidades fraseológicas en el lenguaje jurídico se compiló un corpus legal perteneciente al género del derecho civil mexicano y para este se consideraron textos que estuvieran en un contexto lo más natural posible, es decir que no fueran leyes o códigos civiles, sino que fueran documentos utilizados por las personas en procesos civiles y que pertenecieran a los distintos subgéneros del derecho civil: el derecho de las personas, el derecho de la familia, el derecho de la sucesión, el derecho de las obligaciones y el derecho de los bienes.

En el Cuadro 1, se muestran algunos ejemplos de los documentos que se consideraron para cada uno de los subgéneros del derecho civil.

Subgénero del derecho civil	Ejemplos de los textos
1. Derecho de las personas	Actas de nacimiento y actas de defunción
2. Derecho de la familia	Actas de matrimonio, demandas, sentencias y actas de divorcio
3. Derecho de la sucesión	Testamentos y demandas de sucesión intestada
4. Derecho de las obligaciones	Contratos de arrendamiento, de comodato, de donación y de permuta

Cuadro 1. Ejemplos de los textos que componen el corpus del derecho civil mexicano.

Así, y dadas las características de los documentos legales que se muestran en el Cuadro 1, para el estudio del lenguaje jurídico en el derecho civil se consideró una muestra no probabilística y por conveniencia de textos legales. Al final, se compiló un corpus de 78 textos, con un total de 73,214 palabras y 5,751 tipos, unidad que considera una palabra, sin contabilizar el número de ocasiones que aparece dentro del corpus.

3.2. Segunda etapa: la selección del instrumento

Una palabra de cada 10 utilizadas por un hablante del inglés es un verbo (Kennedy, 2008: 22) y esto normalmente ocurre en la mayoría de las lenguas. De aquí, esta propuesta se plantea partir de los verbos clave del género. Para la identificación de los mismos, es necesario el uso de software de análisis léxico o de gestión de corpus; específicamente las herramientas de lista de palabras y lista de palabras clave. Las listas de palabras representan la transformación radical de un texto para su análisis e implica la división del mismo en unidades más pequeñas (Barlow, 2004: 207). Las palabras clave son aquellas cuya frecuencia es inusualmente alta en comparación con alguna norma. Estas se obtienen al comparar dos listas de palabras ya existentes, una de ellas sirve como referencia de la otra. El objetivo es averiguar qué palabras caracterizan el texto de interés del usuario (Scott, 2015: 229). Para el corpus legal, se utilizaron las herramientas de lista de palabras y lista de palabras clave de *WordSmith Tools 5.0* (Scott, 2008).

Para la identificación y el análisis de las colocaciones, también es imperativo el uso de herramientas de análisis léxico, en particular las concordancias (para el análisis cualitativo) y las colocaciones (para el análisis cuantitativo). McEnery y Hardie (2011: 35) señalan que las concordancias son sin duda la herramienta más importante para los lingüistas pues permiten observar una palabra en contexto. Las concordancias organizan una gran cantidad de ejemplos de una palabra o frase para que el usuario pueda observar patrones que regularmente no son observables en un contexto normal (Hunston, 2002: 9). Sobre la herramienta de colocaciones, Barlow (2004: 213) establece que esta permite observar cuáles son las palabras que acompañan a otra con mayor frecuencia, los colocados, y cuáles son las combinaciones de palabras que coocurren con determinada frecuencia en el discurso. Las colocaciones son un ejemplo de cómo los programas computacionales de análisis de corpus pueden realizar cálculos y permiten observar, por ejemplo, pares de unidades léxicas (Hunston, 2002: 12). En el corpus legal, se utilizaron las herramientas de concordancias y colocaciones también de *WordSmith Tools 5.0* (Scott, 2008), en la primera fase, y *WordSmith Tools 6.0* (Scott, 2011), en la segunda.

3.3. Tercera etapa: El procesamiento y análisis del corpus

En este apartado, se señala el procedimiento para la compilación del corpus, la identificación de los verbos clave en un corpus, así como de las colocaciones relacionadas con estos.

Para la compilación del corpus, primero se debe seleccionar el conjunto de textos que lo conformarán para investigar un lenguaje en específico; para ello existen diversas fuentes. Actualmente, las publicaciones electrónicas han originado una gran oferta de textos disponibles en Internet de manera inmediata, que combinada con las herramientas de gestión y análisis de corpus ofrecen un sinfín de posibilidades para quienes se dedican a las lenguas. Una vez seleccionados los textos, se debe considerar que el formato de los mismos sea compatible con los programas computacionales de análisis de corpus.

Para la compilación del corpus legal, se hizo la selección y clasificación de los textos de acuerdo con los distintos subgéneros del derecho civil mexicano. Las fuentes fueron dos: Internet y traductores jurídicos que tenían en su poder documentos legales en español para su traducción. El formato de los textos del corpus fue *Unicode* (texto simple o *.txt*), que era el formato compatible con el programa seleccionado. Algunos de los textos que se recopilaron ya estaban digitalizados y fue sencillo convertirlos a ese formato. Sin embargo, muchos de ellos, por su carácter legal, se tuvieron que transcribir para tenerlos en versión electrónica y otros se digitalizaron utilizando un escáner y un programa de reconocimiento óptico de caracteres, *OCR* por sus siglas en inglés. Posteriormente, se eliminaron los datos personales y se editaron los textos pues algunos presentaron errores de ortografía o el programa de reconocimiento óptico dejó algunos caracteres equivocados y, finalmente, se convirtieron en archivo *Unicode* con formato *.txt*. Por último, para tener un control sobre los textos que componían el corpus, se realizó un inventario de los mismos en una hoja de cálculo de *Excel*. Después, de que se compiló el corpus, se procedió con la identificación de los verbos clave y las colocaciones relacionadas con estos.

3.3.1. Identificación de los verbos clave

Para la identificación de los verbos clave, primero se debe generar una lista de palabras del corpus objeto de estudio a través de un software de análisis léxico con la herramienta de lista de palabras. Después, se debe obtener otra lista de palabras de un corpus de referencia. Una vez que se cuente con las dos listas de palabras, ambas se deben contrastar y así generar una lista de palabras clave a través de la herramienta de lista de palabras claves del software seleccionado.

Para el corpus del derecho civil mexicano, se creó la lista de palabras. Después, se obtuvo otra lista de palabras de un corpus de referencia, el Corpus de Referencia del Español Actual (Real Academia Española, 2011). Cuando se obtuvieron las dos listas de palabras, se generó una lista de palabras clave del corpus de derecho civil mexicano, a través de la herramienta lista de palabras clave del mismo software.

Para calcular el índice de representatividad de una palabra, en el caso de *WordSmith Tools 5.0* (Scott, 2008), la herramienta de lista de palabras clave del software calcula: la frecuencia de una palabra en el corpus objeto de estudio, el número de palabras en el cor-

pus objeto de estudio, la frecuencia de la palabra en el corpus de referencia y el número de palabras en el corpus de referencia. El software realiza una tabulación cruzada con estos datos. Además, utiliza una medida estadística para realizar el cálculo de la representatividad de cada palabra clave. Una palabra se considera en la lista de palabras clave si resulta inusualmente frecuente (o excepcionalmente poco frecuente) en comparación con lo que se esperaría sobre la base de la lista de palabras más grande (Scott, 2015: 235). La medida que se utilizó en el corpus legal fue la medida de verosimilitud de log-likelihood con un valor de p máximo de 0, 000001, tal como lo considera *WordSmith Tools 5.0* y se consideraron aquellas palabras clave con un valor positivo¹.

Una vez que se obtiene la lista de palabras clave del corpus objeto de estudio, entonces se identifican en esta los verbos clave, incluyendo las distintas formas del verbo. Esta propuesta metodológica considera fundamental el análisis colocacional por palabras y no por lemas, ya que, de acuerdo con algunos autores, como Hoey (2005), establecen que el análisis por lemas puede distorsionar patrones colocacionales. Lo anterior se comprobó en el corpus legal, ya que antes de iniciar con el análisis de colocaciones se realizó una prueba con el plural y el singular del sustantivo *domicilio*. Se encontró, por ejemplo, que este en singular se coloca con los adjetivos: *actual* (IM² 7.80), *conyugal* (IM 6.01) y *legal* (IM 5.59). Es decir, se encontraron las combinaciones *domicilio actual*, *domicilio conyugal* y *domicilio legal* pero estas no se encontraron en plural en el corpus. Esto quiere decir que el sustantivo *domicilio* se comporta de manera distinta en singular y en plural; sería un error analizarlo como uno solo.

No obstante, después de identificar los verbos clave, se sugiere lematizar³ los verbos solo con el fin de considerar en el análisis las distintas formas de los verbos que aparecen en la lista de palabras clave. Por ejemplo, en el caso del corpus legal, *cubrir* resultó un verbo clave y al analizar las colocaciones se consideraron además las formas *cubrir* y *cubrirá*, que también aparecieron en la lista de palabras clave. La lematización manual ofrece la posibilidad de ordenar los datos y después estudiarlos, ya que sin la lematización resulta muy difícil buscar las colocaciones relacionadas con los verbos sin tener un punto de partida, por lo menos, una lista de palabras de palabras clave y sus lemas identificados.

La lista de palabras clave del derecho civil mexicano consistió de un total de 932 palabras y que después de lematizar manualmente los sustantivos y los verbos, se convirtió en una lista de 558 palabras clave. Algunos ejemplos son los verbos *celebrar*, *acreditar* y *sujetarse*, véase Cuadro 2, entre otros. Cabe mencionar que, para efectos del corpus legal, se consideraron sólo aquellos verbos que presentaron una frecuencia = 10, tomando en cuenta las diferentes formas del lema que se encontraron en la lista de palabras clave con el fin de poder observar las colocaciones.

1 De acuerdo con Michael Scott (2015: 236), en una lista de palabras clave, un índice de representatividad o valor *keyness* positivo significa que una palabra ocurre con más frecuencia que la frecuencia esperada atribuida al azar, en comparación con el corpus de referencia. En consecuencia, un valor positivo es significativo con relación al valor de p máximo de 0, 000001.

2 El IM es el índice de Información Mutua, en las páginas siguientes se profundiza al respecto.

3 Lematizar consiste en agrupar por lemas. Un lema es un grupo de formas de palabras que se relacionan por ser las formas flexivas de una palabra base (McEnery y Hardie, 2011).

3.3.2. Identificación y análisis de las colocaciones

Para la identificación de las colocaciones, como ya se dijo antes, se requiere de las herramientas de concordancias y colocaciones. En el corpus legal, la primera se utilizó para ver las palabras en contexto, más concretamente. Por ejemplo, para observar cuáles son las palabras que se colocan con el verbo *suscribir* en el lenguaje jurídico y cuál es el contexto en el que aparecen, tal y como se puede ver en la Ilustración 1.

N	Concordancia	Set	Tag	Word	freq	perc	File	Fin
1	unión de los testigos que al final aparecen, suscriben el presente contrato en la ciudad	suscriben		545	1537%	037%	DC DO CONCOM03 ELB11.txt	93%
2	PRIMERA: Ambas partes convienen en suscribir el presente contrato, obligándose	suscribir		449	1516%	017%	DC DO CONCOM07 ELB11.txt	48%
3	CON FACULTADES SUFICIENTES PARA SUSCRIBIR EL PRESENTE INSTRUMENTO	suscribir		642	1823%	011%	DC DO CONDON10 ELB11.txt	41%
4	, cuenta con las facultades necesarias para suscribir este instrumento, de conformidad	suscribir		227	531%	022%	DC DO CONDEP01 ELB11.txt	22%
5	AMPLIAS Y BASTANTES PARA SUSCRIBIR EL PRESENTE CONTRATO.	suscribir		212	816%	014%	DC DO CONDON04 ELB11.txt	13%
6	cuenta con las facultades necesarias para suscribir este instrumento, de conformidad	suscribir		226	531%	023%	DC DO CONCOM04 ELB11.txt	22%
7	acredita con las facultades que ejerce para suscribir el presente instrumento en los	suscribir		266	1031%	031%	DC DO CONARR08 ELB11.txt	31%
8	PLENAMENTE FACULTADO PARA SUSCRIBIR EL PRESENTE CONTRATO.	suscribir		630	2131%	041%	DC DO CONCOM03 ELB11.txt	44%
9	Inmuebles y Servicios, está facultado para suscribir el presente contrato en atención a	suscribir		554	1542%	037%	DC DO CONDON09 ELB11.txt	36%
10	se encuentra facultado para suscribir el presente Contrato, de	suscribir		277	652%	028%	DC DO CONDON01 ELB11.txt	27%
11	UN PODER ESPECIAL PARA SUSCRIBIR CONTRATOS Y CONVENIOS,	suscribir		399	1356%	071%	DC DO CONDON04 ELB11.txt	36%
12	de las declaraciones anteriores, acuerdan suscribir el presente Contrato al tenor de las	suscribir		471	2047%	017%	DC DO CONDEP01 ELB11.txt	49%
13	LAS QUE SE ENCUENTRAN LAS DE SUSCRIBIR CONVENIOS Y CONTRATOS	suscribir		695	1836%	045%	DC DO CONDON10 ELB11.txt	44%
14	LAS FACULTADES SUFICIENTES PARA SUSCRIBIR EL PRESENTE CONTRATO.	suscribir		647	2111%	032%	DC DO CONCOM06 ELB11.txt	47%
15	con el mandato especial que le autoriza para suscribir el presente contrato, otorgado por	suscribir		192	1230%	021%	DC DO CONDON08 ELB11.txt	21%
16	EN FORMA ALGUNA POR LO QUE PUEDE SUSCRIBIR EL PRESENTE CONTRATO.	suscribir		306	957%	011%	DC DO CONDON10 ELB11.txt	20%

Ilustración 1. El verbo *suscribir* en la herramienta de concordancias.

En la Ilustración 1, se puede observar que en el corpus legal *instrumento*, *convenio* y *contrato* son sustantivos que se colocan con el verbo *suscribir*. Para confirmar esto, se empleó la herramienta de colocaciones. No obstante, primero se deben establecer algunos criterios de la lingüística de corpus para la identificación de las colocaciones.

El primero es el nivel superficial o el espacio colocacional, se dice que dos palabras coocurren si hay una distancia determinada entre una palabra y otra. Los valores más comunes son entre tres y cinco palabras de distancia. El segundo es el nivel textual, es decir la coocurrencia sólo se toma en cuenta si aparece dentro de una misma oración. El tercero es el nivel sintáctico; esto es se considera que una palabra está una cerca de la otra sólo si hay una relación sintáctica directa entre ellas (Evert, 2009: 1220-1223). En el caso del análisis de colocaciones en el corpus legal, se estableció un espacio colocacional de cinco palabras a la izquierda y cinco palabras a la derecha.

Otro de los criterios que se debe considerar en el análisis de colocaciones es la frecuencia de las colocaciones. Cualquier par de palabras que coocurra al menos dos veces en un corpus es una colocación potencial, aunque de hecho es común solicitar mayores niveles de frecuencia entre tres y cinco (Evert, 2009: 1215). En el corpus legal, se decidió incluir

aquellas colocaciones con una frecuencia = 2, por ser un corpus especializado y con fines específicos. Sin embargo, Hoey (2005), Sinclair (2004) y Stubbs (2002), entre otros, señalan que la coocurrencia no es suficiente para establecer si una combinación de palabras es o no una colocación; señalan que para ello es necesario hacer uso de la estadística y que existen índices estadísticos que se utilizan para la identificación de colocaciones, como en el caso de los índices de relación.

De acuerdo con Evert (2009: 1225-1226), los índices de relación asocian la frecuencia observada (O) con la frecuencia esperada (E) bajo la hipótesis de independencia y calculan un puntaje de relación para la atracción entre dos palabras. La forma más sencilla e intuitiva de relacionar O y E es el uso de la relación O / E . Por ejemplo, $O / E = 10$ significa que un par de palabras coocurre 10 veces más de lo que se esperaría por azar, lo que indica un cierto grado de colocabilidad; algunos de los índices de relación que existen son el de Información Mutua (IM) y de razón de verosimilitud log-likelihood (G2), este último presentado por Dunning (1993), entre otros.

De acuerdo con Manning, Raghavan y Schütze (2009: 272), el IM mide la manera en la que la presencia o ausencia de un elemento contribuye a la presencia de otro. Según McEnergy y Wilson (2001: 86), cuanto más fuerte sea la relación entre los dos elementos, mayor será el IM. Según Evert (2009: 1235), el G2 en realidad es una prueba de dos caras. Esto significa que la prueba no distingue entre $O \gg E$ y $O \ll E$. En consecuencia, asigna valores positivos en ambos casos. A continuación, en las Ilustraciones 2 y 3, se muestra un ejemplo del análisis colocacional del verbo *suscribir* en el corpus legal considerando estos dos índices.

N	Word	With	Relation	Texts	Total	Total	Total	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	SUSCRIBIR	suscribir	286 603	11	15	0	0					15				
2	PRESENTE	suscribir	94 208	9	11	0	11							11		
3	PARA	suscribir	87 013	10	11	11	0					11				
4	CONTRATO	suscribir	65 338	8	9	0	9									9
5	FACULTADES	suscribir	57 793	5	5	5	0	1	4							
6	EL	suscribir	52 788	9	11	0	11							11		
7	FACULTADO	suscribir	45 035	3	3	3	0					3				
8	INSTRUMENTO	suscribir	36 455	4	4	0	4								2	2
9	CONVENIOS	suscribir	27 630	2	2	0	2							1		1
10	LAS	suscribir	26 163	4	5	5	0	4		1						
11	SUFICIENTES	suscribir	25 364	2	2	2	0							2		
12	CONTRATOS	suscribir	24 354	2	2	0	2							1		1
13	NECESARIAS	suscribir	21 807	2	2	2	0							2		
14	TAL	suscribir	18 595	2	2	0	2									2
15	ENCUENTRA	suscribir	16 162	2	2	2	0	1	1							
16	CONFORMIDAD	suscribir	15 835	2	2	0	2									2
17	LO	suscribir	14 742	2	3	2	1							2		1
18	DE	suscribir	9 079	5	5	1	4					1			2	2
19	ESTE	suscribir	9 071	2	2	0	2							2		
20	Y	suscribir	7 330	2	3	1	2			1					2	
21	QUE	suscribir	7 103	3	3	3	0	1	1	1						
22	EN	suscribir	6 306	3	3	1	2					1				2
23	SE	suscribir	5 748	2	2	2	0	2								

Ilustración 2. El verbo *suscribir* en la herramienta de colocaciones considerando el índice log-likelihood (G2)

En la Ilustración 2, se observa que las palabras que se colocan con el verbo *suscribir* están ordenadas del índice G2 más alto al más bajo y se estima que la probabilidad conjunta

de *contrato*, *convenio* e *instrumento* con el verbo *suscribir* los hace colocados del mismo. Por contraste, el artículo definido *el*, que pertenece a una categoría funcional y no léxica, tiene un índice G2 alto, incluso más alto que el de *instrumento* o *convenio* y se encuentra entre los 10 primeros colocados. Ahora obsérvese la Ilustración 3, en donde se muestran las colocaciones con el verbo *suscribir* con el índice IM.

N	Word	With	Reacion	Tests	Total	Total Left	Total Right	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	SUSCRIBIR	suscribir	12.292	11	15	0	0					15				
2	FACULTADO	suscribir	11.555	3	3	3	0			3						
3	CONVENIOS	suscribir	10.970	2	2	0	2						1		1	
4	SUFICIENTES	suscribir	10.292	2	2	2	0			2						
5	CONTRATOS	suscribir	9.970	2	2	0	2						1		1	
6	FACULTADES	suscribir	9.404	5	5	5	0	1	4							
7	NECESARIAS	suscribir	9.122	2	2	2	0			2						
8	TAL	suscribir	6.000	2	2	0	2									2
9	INSTRUMENTO	suscribir	7.768	4	4	0	4							2	2	
10	ENCUENTRA	suscribir	7.142	2	2	2	0	1	1							
11	CONFORMIDAD	suscribir	7.025	2	2	0	2									2
12	PRESENTE	suscribir	6.866	9	44	0	44							44		
13	PARA	suscribir	6.283	10	11	11	0					11				
14	CONTRATO	suscribir	6.100	8	9	0	9								9	
15	LAS	suscribir	4.908	4	5	5	0	4		1						
16	LO	suscribir	4.287	2	3	2	1		2							1
17	ESTE	suscribir	4.557	2	2	0	2						2			
18	EL	suscribir	4.130	9	11	0	11							11		
19	SE	suscribir	3.895	2	2	2	0	2								
20	Y	suscribir	2.892	2	3	1	2		1					2		
21	QUE	suscribir	2.830	3	3	3	0	1	1	1						
22	EN	suscribir	2.611	3	3	1	2					1				2
23	DE	suscribir	2.273	5	5	1	4					1			2	2

Ilustración 3. El verbo *suscribir* en la herramienta de colocaciones con el índice IM.

En la Ilustración 3, se estima una vez más que *contrato*, *instrumento* y *convenio* se colocan con el verbo *suscribir*, pero en este caso el artículo determinante *el* tiene un índice por debajo del que tienen esas palabras. Además, este no aparece entre los 10 primeros colocados sino que ocupa las últimas posiciones junto con otros elementos funcionales como los artículos definidos e indefinidos, conjunciones y preposiciones.

Así, al comparar ambos índices, se observa que el índice G2 es influenciado por la alta frecuencia, que fue lo que ocurrió con el artículo definido. Este índice podría resultar útil en trabajos para los que sean relevantes los elementos de categorías funcionales de alta frecuencia como los artículos o las preposiciones, por ejemplo, y que se enfoquen más en los aspectos sintácticos que los léxicos. Por lo anterior, en el análisis de colocaciones en el corpus con fines específicos (Maia, 2002) se sugiere el índice IM. En el corpus legal se consideró el índice IM, como está implementado en *WordSmith Tools*, para la identificación de las colocaciones pues proporcionaba datos más interesantes. Para el corpus legal, sólo se consideraron aquellas colocaciones que tenían un $IM = 3$. Esto está en consonancia con lo establecido por Hunston (2002: 71), quien señala que un IM de tres o más es significativo.

Para llevar un control sobre las colocaciones se realizó un control de datos en Excel que contiene, entre otros, los siguientes datos: el lema, la frecuencia del lema, el tipo de verbo, el colocado, la frecuencia del colocado, la frecuencia total de la colocación, el índice IM, y el tipo de colocación. A continuación, se muestra un resumen de los resultados obtenidos sobre la identificación y el análisis de colocaciones en el corpus legal del derecho civil mexicano.

4. RESULTADOS

A través de la metodología propuesta, en el corpus del derecho civil mexicano, se encontraron un total de 101 verbos con un índice de representatividad positivo en la lista de palabras clave. De estos, 60 tuvieron una frecuencia de =10, que son los verbos que se consideraron. En el Cuadro 2, se muestran estos verbos ordenados según el índice de representatividad, o *keyness*. Además, se presenta su frecuencia y las formas del lema que fueron clave en el corpus.

No.	Verbo o lema	Índice de representatividad	Frecuencia	Formas del lema
1	Declarar	633.07	125	Declara[94] Declaran[25] Declare[6]
2	Escribir	470.89	148	Escriba[28] Escrito [120]
3	Denominar	394.96	51	Denominará[37] Denominara[14]
4	Firmar	298.64	37	Firman[37]
5	Deber	298.2	96	Deberá[81] Debiendo[15]
6	Poder	223.32	113	Podrá[79] Pudiera[27] Pudiere[7]
7	Comprometer	213.62	43	Compromete[34] Comprometiéndose[9]
8	Subarrendar	208.91	15	Subarrendar[15]
9	Corresponder	207.9	12	Corresponda [12]
10	Celebrar	184.74	94	Celebrar[42] Celebran[35] Celebrado[17]
11	Acordar	162.85	23	Acordar[23]
12	Recibir	149.64	98	Recibir[52] Recibe[34] Reciba[8] Recibir las[4]
13	Manifestar	130.18	93	Manifiesta[37] Manifiestan[27] Manifestamos[5] Manifestando[24]
14	Pagar	128.76	88	Pagar[48] Pagará[31] Pagada[9]
15	Ser	126.08	137	Será[107] Serán[30]
16	Disponer	104.58	41	Dispuesto[41]
17	Desocupar	104.18	21	Desocuparlo[10] Desocupare[3] Desocuparla[5] Desocupe[3]
18	Cubrir	101.1	41	Cubrir[31] Cubrirá[10]
19	Suscribir	97.3	62	Suscribir[15] Suscrita[21] Suscrito[10] Suscritos[16]
20	Estipular	96.09	22	Estipulado[14] Estipulada[8]
21	Leer	93.05	27	Leído[27]
22	Requerir	84.79	21	Requiera[14] Requieran[7]
23	Constar	83.38	42	Constar[16] Constandoles[5] Consta[21]
24	Expedir	83.11	29	Expedido[12] Expedida[12] Expedidos[5]
25	Sujetar	80.92	17	Sujetan[11] Sujetarse[6]
26	Ratificar	78.83	10	Ratifican[10]
27	Entregar	76.99	32	Entregar[22] Entregará[10]

No.	Verbo o lema	Índice de representatividad	Frecuencia	Formas del lema
28	Comparecer	71.07	33	Comparece[9] Comparecen[8] Comparecemos[4] Comparecieron[4] Comparezco[8]
29	Otorgar	70.53	54	Otorgar[16] Otorga[29] Otorgan[9]
30	Quedar	69.26	20	Quedando[20]
31	Acreditar	67.28	54	Acreditar [11] Acredita[27] Acreditamos [4] Acreditan [7] Acredito[5]
32	Aceptar	66.24	21	Acepta[21]
33	Anteceder	64.32	13	Antecede[9] Antecedan[4]
34	Describir	61.58	40	Descrito[17] Descritos[15] Descritas[8]
35	Tener	58.07	33	Tendrá[25] Tenerme[8]
36	Inscribir	56.92	22	Inscrito[11] Inscrita[11]
37	Protestar	55.56	11	Protesto[7] Protestamos[4]
38	Causar	55.02	15	Cause[8] Causen[7]
39	Ubicar	52.62	15	Ubicada[15]
40	Contratar	51.34	13	Contratar[13]
41	Demandar	49.36	12	Demandar[9] Demando[3]
42	Expresar	48.34	21	Expresa[21]
43	Exhibir	47.24	10	Exhibir[10]
44	Retener	45.35	11	Retener[11]
45	Exponer	43.55	28	Exponer[13] Expuesto[15]
46	Estar	40.93	15	Estando[15]
47	Garantizar	39.44	17	Garantizar[17]
48	Constituir	38.98	11	Constituida[11]
49	Exigir	33.75	13	Exigir[13]
50	Disolver	31.85	13	Disuelta[6] Disuelto [7]
51	Convenir	31.67	79	Conviene[18] Convienen[54] Conviniendo[7]
52	Devolver	31.61	26	Devolver[10] Devolverá[9] Devuelta[7]
53	Autorizar	28.36	34	Autorizar[7] Autorizo[3] Autorizando[13] Autoriza[11]
54	Realizar	28.2	43	Realizar[30] Realice[13]
55	Oír	28.02	20	Oír[16] Oírlas[4]
56	Ceder	27.18	10	Ceder[10]
57	Mencionar	26.96	41	Mencionado[12] Mencionada[13] Mencionados[16]
58	Hacer	26.85	30	Hacerlo[26] Hiciere[4]
59	Originar	26.66	10	Origine [4] Originen [6]
60	Destinar	25.75	24	Destinar[6] Destinará[6] Destinado[12]

Cuadro 2. Verbos clave del corpus legal de derecho civil con una frecuencia = 10.

En el Cuadro 2, los lemas están ordenados según el índice de representatividad; se puede observar que aunque algunos verbos muestran un índice de representatividad alto, estos presentan una frecuencia baja y viceversa. Esto se debe a que los verbos clave no son aquéllos que tienen una mayor frecuencia dentro del corpus, sino aquéllos que, como ya se dijo antes, se producen con una frecuencia inusual en comparación con el corpus de referencia. Por ejemplo, el verbo *subarrendar* tiene una frecuencia de 15 y un índice de representatividad de 208.91, mientras que el verbo *entregar* tiene una frecuencia de 32 y un índice de representatividad de 71.07.

También, se aprecia que aunque hay verbos con un índice de representatividad muy alto como *declarar*, *denominar* y *firmar*, con índices de 633.07, 394.96 y 298.64 respectivamente, y que ocupan las primeras posiciones en el Cuadro 2, estos no son necesariamente los más representativos o específicos del género. Scott (2015: 237) señala que puede aparecer una palabra con un índice de representatividad de 50 y otra con uno de 500 dentro de una lista de palabras clave, pero esto no significa necesariamente que una sea más representativa que la otra dentro de un texto.

Por lo tanto, el autor menciona que generalmente no es seguro confiar en la posición que las palabras ocupan en una lista de palabras clave según el índice de representatividad. Así, se deberá considerar que los verbos que se presentan en el Cuadro 2 son los verbos clave en el corpus legal del derecho civil mexicano (con una frecuencia =10) sin considerar que unos son más representativos que otros. En lo que a las formas de los lemas concierne, se aprecian las formas clave de los verbos; que son el infinitivo, como en *subarrendar*, *celebrar* y *recibir*, el gerundio, como en *comprometiéndose*, *manifestando* y *quedando*, y el participio, como *pagada* o *mencionada*. Asimismo, se aprecian tiempos y modos verbales clave, como el presente y el futuro del indicativo; por ejemplo, *declara* y *comparecen*, para el primer caso, o *destinará* y *tendrá*, para el segundo. También, aparecen verbos en presente, pretérito imperfecto y futuro del subjuntivo; ejemplos de ello son *origine*, *requiera* y *desocupare*, respectivamente.

De estos 60 verbos, que aparecen en el Cuadro 2, algunos no se incluyeron en la base de datos de colocaciones; tal es el caso de los verbos *disolver*, *disponer* o *expresar* ya que se encontró que en el corpus estos tres en realidad son participios con funciones de adjetivo o sustantivo, como se puede observar en los ejemplos (1), (2) y (3).

- (1) a. sociedad conyugal vigente o disuelta
- (1) b. se declara disuelto el vínculo
- (2) de acuerdo con lo dispuesto
- (3) con expresa renuncia de cualesquier otro fuero

Se encontraron en el mismo caso los siguientes verbos: *describir*, *constituir* *escribir*, *estipular* y *expedir*. Cabe señalar que en (1 b) además *disuelto* resultó colocado de *declarar*; por lo que aparece bajo *declarar* y no *disolver*. Otros verbos que no se incluyeron en este trabajo son los que cumplen la función de modal como *deber* y *poder*, como en los ejemplos (4) y (5).

- (4) debiendo el arrendador reembolsar
- (5) la arrendadora podrá rescindir este instrumento jurídico

Sin embargo, el uso de los modales sí se consideró para el análisis de las estructuras y patrones de las colocaciones de los verbos clave, tal es el caso del ejemplo (6).

(6) por ningún motivo podrá el arrendatario retener rentas

Otro verbo que tampoco se incluyó fue *exponer* pues se encontró que en realidad era colocado del verbo *comparecer* como en (7).

(7) Comparezco a exponer lo siguiente.

Finalmente, se decidió no incluir el verbo *ser*, ya que “*ser* [...] no puede establecer colocaciones sustantivo-verbo por no poder introducir ningún complemento directo, frente a, p. ej., *saldar* [...]” (Koike, 2001, p. 41). El resto de los verbos, en total 46, fueron incluidos en la base de datos de colocaciones. A continuación, se muestran los resultados sobre las colocaciones que se encontraron con estos verbos.

En lo que a las colocaciones concierne, se encontraron un total de 248 colocaciones relacionadas con los verbos clave en el corpus legal de derecho civil, de las cuales el 88% fueron colocaciones léxicas simples, el 6% fueron colocaciones léxicas complejas y el otro 6% fueron colocaciones gramaticales. Tal como se observa en el Gráfico 1.

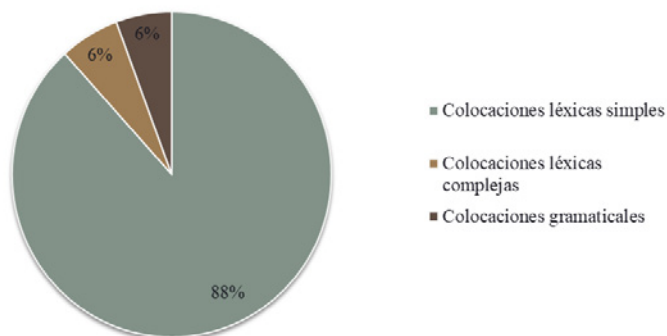


Gráfico 1. Colocaciones en el corpus del derecho civil mexicano.

Lo anterior se traduce en 219 colocaciones léxicas simples, 15 colocaciones léxicas complejas y 14 colocaciones gramaticales. A continuación, se presentan algunos ejemplos de este tipo las colocaciones que se encontraron en el corpus legal del derecho civil mexicano a través de esta propuesta metodológica. Ejemplos de las colocaciones léxicas simples del tipo verbo + sustantivo objeto son los que se muestran en el Cuadro 3.

En el Cuadro 3, específicamente en los ejemplos (1), (2) y (3) se observa la presencia de verbos auxiliares o modales. La aparición de modales es recurrente en el corpus legal del derecho civil mexicano para algunos verbos. Además, se puede apreciar que el verbo *cubrir* se combina con sustantivos relacionados con cuantía o cantidad. Asimismo, el verbo *suscribir* aparece con sustantivos relacionados con documentación. Otros ejemplos de colocaciones léxicas simples fueron las del tipo verbo + preposición + sustantivo como las que se observan en el Cuadro 4.

Verbo	Sustantivo (objeto)	IM	Ejemplos
Cubrir +	Honorarios	9.07	(1) <i>El Arrendatario se obliga de antemano a cubrir los gastos y honorarios [...]</i>
	Mensualidad	9.07	(2) <i>[...] deberá cubrir la mensualidad, aun cuando únicamente use la vivienda un día.</i>
	Importe	8.60	(3) <i>se obliga, a cubrir el importe de los servicios de la energía eléctrica, [...]</i>
Suscribir +	Convenio	10.97	(1) <i>Facultades para suscribir convenios</i>
	Instrumento	7.76	(2) <i>su carácter de [...], cuenta con las facultades necesarias para suscribir este instrumento</i>
	Contrato	6.10	(3) <i>[...] suscriben el presente contrato [...]</i>

Cuadro 3. Ejemplos de colocaciones léxicas simples del tipo verbo + sustantivo objeto.

Verbo	Preposición + sustantivo	IM	Ejemplos
Comparecer +	(Por) derecho	8.03	(1) <i>[...]comparece por derecho propio y se le denomina como “EL AVAL”</i>
Firmar +	(Al) calce	9.04	(2) <i>[...] los testigos que firman al calce[...]</i>
	(Al) margen	10.4	(3) <i>[...]lo firman al margen en cada foja con excepción de la última que se firma al calce.</i>
Quedar +	(En) aptitud	11.46	(4) <i>[...] quedando ambos cónyuges en aptitud de contraer nuevas nupcias [...]</i>
	(En) beneficio	8.53	(5) <i>quedando en consecuencia en beneficio de la casa.</i>

Cuadro 4. Ejemplos de colocaciones léxicas simples del tipo verbo + preposición + sustantivo.

Nótese que aunque estas combinaciones se han clasificado como colocaciones léxicas del tipo verbo + preposición + sustantivo, también pudieran ser gramaticales. Se decidió clasificarlas como léxicas considerando otros aspectos como que el verbo *firmar* regularmente no se coloca con la preposición *a* sino con la preposición *en*, y que esta última sería la colocación gramatical o que *quedar*, como verbo pseudo copulativo, normalmente va acompañado por un adjetivo en la lengua estándar y no una preposición. Estas consideraciones van más allá de las medidas meramente cuantitativas.

Verbo	Adverbio	IM	Ejemplos
Comparecer +	Respetuosamente	12.78	(1) <i>Ante su señoría respetuosamente comparezco a fin de [...]</i>
Convenir +	Voluntariamente	9.85	(2) <i>Las partes convienen voluntariamente y de común acuerdo [...]</i>

Cuadro 5. Ejemplos de colocaciones léxicas simples del tipo verbo + adverbio.

Cabe señalar que se descubrió que algunas de las colocaciones que se presentan en el Cuadro 5 no son fijas. Por ejemplo, para *comparecer + respetuosamente*, que se observa en el ejemplo (1), también se encontró *comparecer con respeto* por *comparecer respetuosamente*. Esto muestra que algunas colocaciones no presentan el mismo grado de fijación. Esto coincide con lo establecido por Alonso Ramos (2010: 57) quien establece que no todas las colocaciones tienen el mismo grado de idiosincrasia, transparencia semántica o de fijación.

Como ya se mencionó, el 6% del total de las colocaciones relacionadas con los verbos clave en el corpus legal del derecho civil mexicano fueron colocaciones léxicas complejas. Ejemplos de estas colocaciones se observan en el Cuadro 6.

Verbo	Locución verbal	IM	Colocativo	IM	Ejemplos
Autorizar	Autorizar y dar	13.30	Fe	8.45	(1) <i>La secretaria de acuerdos autoriza y da fe</i>
Oír	Oír y recibir	10.40	Notificaciones	10.24	(2) <i>para oír y recibir toda clase de notificaciones</i>
Leer	Leer y explicar	10.44	Instrumento	7.72	(3) <i>que leído y explicado este instrumento</i>

Cuadro 6. Ejemplos de colocaciones léxicas complejas del tipo locución verbal + sustantivo.

En los ejemplos del Cuadro 6, al igual que en otro tipo de colocaciones, se encontró que algunas de estas son más o menos fijas, tal es el caso de la locución *leer y explicar + instrumento* que también puede combinarse con *contrato*.

Las colocaciones gramaticales representaron el otro 6% del total de las colocaciones en el corpus legal del derecho civil mexicano. Las colocaciones que se encontraron de este tipo fueron las colocaciones de verbo + preposición, en particular del tipo verbo + preposición + verbo. Ejemplos de estas son las que se muestran en el Cuadro 7.

En el Cuadro 7, se puede observar la preferencia de algunos verbos por ir acompañados por preposiciones en el género del derecho civil mexicano, cuando no lo hacen así regularmente en la lengua estándar; tal como se observa en el ejemplo (1).

Verbo	Preposición + verbo	IM	Ejemplos
Convenir	(En) someterse	9.85	(1) las partes convienen en someterse a la jurisdicción y competencia de los tribunales federales de la ciudad de México [...]
	(En) sujetarse	9.70	(2) [...] convienen en sujetarse a la jurisdicción de los tribunales federales competentes [...]
	(En) establecer	9.44	(3) las partes convienen en establecer como pena convencional para el caso de incumplimiento [...]
Venir	(A) demandar	12.09	(4) [...] por medio del presente del escrito vengo a demandar en controversia familiar [...]
Comparecer	(A) exponer		(5) [...] comparezco a exponer lo siguiente

Cuadro 7. Ejemplos de colocaciones gramaticales verbo + preposición + verbo.

5. CONCLUSIONES

La propuesta metodológica parece funcionar en estudios de alcance descriptivo, especialmente con corpus especializados y con fines específicos. Una de las ventajas de este método es que los resultados se pueden abordar de manera independiente. Por un lado, las palabras clave de un corpus siempre dicen mucho sobre el lenguaje objeto de estudio y, por el otro, las colocaciones sobre los patrones y la idiomatidad dentro del mismo.

Otra de las ventajas que ofrece esta propuesta es que, a pesar del tamaño del corpus, pues para algunos el tamaño es importante, este permitió identificar los verbos clave, así como las colocaciones relacionadas con estos e incluso observar algunos patrones colocacionales como los que se pudieron observar en los ejemplos. Además, permitió observar que el comportamiento de algunas de estas colocaciones es más o menos fijo y que, aunque la relación entre los colocados no es exclusiva, sí existe una asociación semántica entre los colocados que restringe la coocurrencia para los elementos que comparten rasgos semánticos.

Finalmente, y otra de las bondades de esta metodología, es que el trabajo de análisis sin etiquetas permite una mayor proximidad con los datos del corpus objeto de estudio, pues se analizan todos los datos y no sólo algunos de estos. Además, se considera que las herramientas de análisis cualitativo (como las concordancias) y cuantitativo (como las listas de palabras clave y colocaciones) son suficientes para la identificación y el análisis de colocaciones en el corpus, aunque claro está que el análisis de los datos consume tiempo.

El conocimiento sobre estas colocaciones es importante tanto para los aprendices de una lengua extranjera, como para quienes estudian una lengua con fines específicos y para quienes se dedican a la traducción pues, el desconocimiento sobre estas afecta la creatividad y la producción lingüística automática, es decir casi sin pensar. También, el mismo desconocimiento puede producir incompreensión por parte del receptor en el proceso de la comunicación, así como falsos sentidos.

Las colocaciones son sólo un eslabón en la cadena de la producción lingüística en la traducción y en la enseñanza-aprendizaje de una lengua extranjera o con fines específicos pues cada lengua, y dentro de ella los lenguajes especializados, tiene patrones específicos.

Así, se considera que la metodología aquí propuesta resulta de gran utilidad para la extracción de colocaciones y a una aproximación a los elementos caracterizadores de los lenguajes especializados, tal como se pudo apreciar en los ejemplos.

Referencias bibliográficas

- Alonso Ramos, M. (1994). "Hacia una definición del concepto de colocación: de J. R Firth a I. A. Mel'cuk", *Revista de Lexicografía*, 1, pp. 9-28.
- Alonso Ramos, M. (2010). "No importa si la llamas o no colocación, descríbela". En Mellado, C. et al. (eds.), *La fraseografía del S. XXI: nuevas propuestas para el español y el alemán*. Alemania: Frank & Timme, , pp. 55-80.
- Barlow, M. (2004). "Software for corpus access and analysis". En J. M. Sinclair. *How to use corpora in language teaching*. Amsterdam: John Benjamins Publishing Company, pp. 205-221.
- Benson, M., Benson, E., & Ilson, R. (2010). *The BBI Combinatory Dictionary of English*. Amsterdam: John benjamins.
- Bosque, I. (2001). "Sobre el concepto de 'colocación' y sus limitaciones", *Lingüística Española Actual* 23 (1), pp. 9-40.
- Cabré, M. T. (2004). Lenguajes especializados o lenguajes para propósitos específicos, *Revista Foro Hispánico*, 26, pp. 19-34.
- Corpas Pastor, G. (1996). *Manual de fraseología española*. España: Gredos
- Corpas Pastor, G. (2001). "En torno al concepto de colocación", *EUSKERA*, XLVI, pp. 90-107.
- Corpas Pastor, G. (2003). *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. España: Iberoamericana.
- Dunning, T. (1993). "Accurate Methods for the Statistics of Surprise and Coincidence". En *Computational Linguistics* vol. 19, no. 1, pp. 61-74.
- Evert, S. (2009). "Corpora and Collocations". En A. Lüdeling, & Kytö, M. *Corpus Linguistics: An International Handbook*, Alemania: Mouton Gruyter. Vol. 2, pp. 1212-1248.
- Firth, J. R. (1957). "A Synopsis of Linguistic Theory (1930-1955)". En *Studies in Linguistic Analysis*. Oxford: The Philological Society, pp. 1-32.
- Gries, S. (2008). "Phraseology and linguistic theory: A brief survey". En Granger, S., & Meunier, F. (Edits.), *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins, pp. 3-26.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Nueva York: Routledge.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Reino Unido: Cambridge University Press.
- Kennedy, G. (2008). "Phraseology and language pedagogy: Semantic preference associated with English verbs in the British National Corpus". En Meunier, F. y Granger, S. *Phraseology in Foreign Language Learning and Teaching*. Philadelphia: John Benjamins, pp. 21-41.
- Koike, K. (2001). *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*. Alcalá de Henares: Universidad de Alcalá.
- L'Homme, M.-C. (2009). "A methodology for describing collocations in a specialized dictionary". En S. Nielsen, & S. Tarp (Edits.). *Lexicography in the 21st Century: In honour of Henning Bergenholtz*. Amsterdam: John Benjamins, pp. 237-256.
- Maia, B. (2002). "Do it yourself, disposable, specialized mini corpora. Where next? reflections on teaching translation and terminology through corpora", *Cadernos de Tradução*, 1(9), pp. 221-235.
- Manning, C., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Nueva York: Cambridge University Press.
- McEnery, T. y Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Estados Unidos: Cambridge University Press.
- Real Academia Española. Corpus de Referencia del Español Actual. Recuperado el 25 de septiembre de 2011, de Banco de datos (CREA) [en línea]: <http://corpus.rae.es/creanet.html>
- Scott, M. (2008) *WordSmith Tools version 5*, Liverpool: Lexical Analysis Software.
- Scott, M. (2011) *WordSmith Tools version 6*, Liverpool: Lexical Analysis Software.
- Scott, M. (2015). *WordSmith Manual*. Stroud: Lexical Analysis Software.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the text: Language, Corpus and Discourse*. Nueva York: Routledge.
- Stubbs, M. (2002). *Words and Phrases: Corpus Studies of Lexical Semantics*. Gran Bretaña: Blackwell Publishers.