

# Análisis de la diversidad morfosintáctica en las locuciones verbales

Belém Priego Sánchez

Laboratoire Lexiques, Dictionnaires, Informatique, CNRS (UMR 7187),  
Université Paris 13, Sorbonne Paris Cité,  
Francia

Benemérita Universidad Autónoma de Puebla,  
Facultad de Ciencias de la Computación,  
México

belemps@gmail.com

**Resumen.** La lengua permanece en constante cambio, disponiendo de una gran cantidad de palabras, combinación y unión de éstas, que permiten expresar un concepto determinado y que son utilizadas en diversos contextos. La fusión de las palabras permite construir frases; la inmensa mayoría de frases están formadas por un verbo y una o varias variables. El verbo exige y realiza una rigurosa selección de los sujetos y de los componentes que pueden acompañarle. A partir del análisis realizado a un tipo particular de frases, en este artículo se presentan los patrones morfosintácticos de las locuciones verbales, cuyo elemento principal es un sintagma verbal y presentan una gran diversidad morfosintáctica. El análisis realizado a las locuciones verbales fue desarrollado para el idioma español.

**Palabras clave:** Locución verbal, morfosintaxis, patrones morfosintácticos.

## 1. Introducción

La comunicación es un sistema complejo dado que para llevarla a cabo se necesita de un lenguaje (oral, escrito, señas, braille, etc.) que permita expresar una situación concreta y de manera diferente en cada circunstancia determinada tomando como referente el horizonte cultural compartido. A lo largo de la vida se adquiere, capta y memoriza información que complementa la base de conocimiento léxica personal y que permite desarrollar la habilidad de comunicación. Si se toma como base el lenguaje oral, esta información es representada en forma de frases, expresiones, dichos, la cual permite formular una idea o un concepto y visualizarlas como un todo capaz de expresar variadamente una situación en contextos diferentes. Dicha información, en ocasiones, se llega a plasmar de forma escrita en lenguaje natural (como español, francés, inglés o cualquier otro idioma) en libros, periódicos, revistas u otros textos y en formato digital. Siendo

ésta un recurso importante para la humanidad, la cual al ser almacenada es posible analizarla y obtener resultados benéficos como herramientas que ayuden y faciliten la explotación de esta información y que permitan obtener conocimiento de una lengua a través del acervo lingüístico y cultural que poseen.

La lengua permanece en constante cambio, disponiendo de una gran cantidad de palabras y unión de éstas que a veces no se considera lógica pero que permite expresar un concepto determinado y que no son unidas libremente, sino que son utilizadas y repetidas durante años como bloque de combinaciones. Ya desde la antigüedad, el hombre dedicado a la ciencia sintió un enorme interés por determinadas expresiones cotidianas, las cuales quiso agrupar y estudiar. Este tipo de expresiones que regularmente son fácilmente dominadas por los hablantes nativos de una lengua, plantean un gran desafío para los sistemas computacionales en términos de su interpretación debido a su naturaleza flexible y heterogénea. Estas expresiones no son tan frecuentes en los recursos léxicos como en los textos del mundo real y por lo tanto presentan un gran reto de estudio para diversas tareas de la lingüística computacional.

Este trabajo se enfoca principalmente en un tipo de expresiones denominadas unidades fraseológicas, las cuales no constituyen enunciados completos, tienen la principal característica de ser una combinación estable de dos o más términos y pueden constituir el núcleo de sintagmas verbales. Se analizan expresiones como por ejemplo: *chuparse el dedo*, *costar un ojo de la cara*, *dar de sí*, *mover cielo y tierra*, *ser el vivo retrato de alguien*, entre otras; es decir, expresiones denominadas locuciones verbales. A partir del análisis realizado se muestran los patrones morfosintácticos de estas construcciones lingüísticas y se identifican en un corpus periodístico.

El presente artículo está organizado de la siguiente manera. La Sección 2 expone de manera general a la unidades fraseológicas, se centra en la definición y características de las locuciones, particularmente de las locuciones verbales. Así como, algunos trabajos sobresalientes encontrados en la literatura. La Sección 3 describe la metodología llevada a cabo para la identificación de los patrones morfosintácticos. La Sección 4 muestra los resultados obtenidos al buscar los patrones en un corpus. Finalmente en la Sección 5 se presentan las conclusiones y perspectivas de este trabajo.

## 2. Unidades fraseológicas

La fraseología, disciplina lingüística encargada del estudio de la combinación de palabras características de una lengua, ha incrementado su estudio en los últimos años, dado que ha sido un espacio de interés para investigadores atraídos hacia estas construcciones lingüísticas por el hecho de ser un modo particular de habla de los pueblos como material de estudio científico e implica diferentes dimensiones del lenguaje: lingüísticas, pragmáticas, culturales y muchas más. Este hecho refuerza el interés de estudio de este trabajo.

En la lingüística española se emplean diferentes denominaciones a la combinación de estas palabras; a manera de ilustración se citan en este trabajo los

nombres que servirán para dar una idea de la abundante nomenclatura que existe alrededor de estas denominaciones comúnmente identificadas como: unidad fraseológica, fraseologismo, modismos, locuciones, expresiones fijas, dichos, frases hechas, expresiones idiomáticas, expresiones, frases, refranes, proverbios, modos de decir, aforismos, entre otras (en [16] aparecen hasta 64 denominaciones diferentes). En este trabajo se emplea el término *unidad fraseológica* para referirse a las combinaciones de palabras que tienen un significado como un todo, debido a que esta denominación goza de gran aceptación en la filología española, es conocido en la fraseología internacional. Se hace hincapié de la existencia de diferentes tipos de unidades fraseológicas y se remarca que en este trabajo de investigación se analiza solamente un tipo particular de ellas.

Las clasificaciones de las unidades fraseológicas, de manera cronológica, han sido establecida por lingüistas como [6,8,25,27,13,5], entre otros. A partir de ellas, en [7] se propone realizar una clasificación, que parte de [12] y que toma en cuenta algunas características de los trabajos mencionados. Por tanto, en este trabajo, se toma como base esta clasificación que divide en dos grupos a las unidades fraseológicas. El primer grupo, corresponde a las unidades que no constituyen un enunciado completo, las cuales incluyen a las *colocaciones* y a las *locuciones*. El segundo grupo, lo forman las unidades que constituyen un enunciado completo, denominadas como *enunciados fraseológicos*.

Las colocaciones, desde el punto de vista del sistema de la lengua son sintagmas completamente libres que presentan cierto grado de restricción combinatoria determinada por el uso. Éstas son generadas a partir de reglas y generalmente son de base semántica. Algunos ejemplos son: *asumir una responsabilidad, conciliar el sueño, banco de peces, correr un rumor, declararse una epidemia, desempeñar un cargo, estallar una guerra, fuente fidedigna, relación estrecha, negar rotundamente, rematadamente loco, tomar una decisión, una rebanada de pan, viaje relámpago, zarpar un barco*, entre otras.

Las locuciones, son definidas por [6] como una combinación estable de dos o más términos, que funciona como elemento oracional y cuyo sentido unitario consabido no se justifica, sin más, como una suma del significado normal de los componentes. Las diferentes definiciones de locución en español han seguido esta caracterización. Las locuciones se han dividido según la función oracional que desempeñen, independientemente de que sean conmutables por palabras simples o por sintagmas. En [7] se distinguen los siguientes tipos:

- Locuciones nominales: *el qué dirán, mosca muerta, patas de gallo, santo y seña*, entre otras.
- Locuciones adjetivas: *corto de medios, de armas tomar, más suave que el algodón, sano y salvo*, entre otras.
- Locuciones adverbiales: *boca con boca, con el corazón en la mano, de par en par, más de la cuenta, por lo pronto*, entre otras.
- Locuciones verbales: *cargársela, dar de sí, meter a alguien en cintura, ir y venir, saber de qué pie cojea*, entre otras.
- Locuciones prepositivas: *a pesar de, delante de, en lugar de, gracias a*, entre otras.

- Locuciones conjuntivas: *así que, como si, dado que, mientas tanto, puesto que, tan pronto como*, entre otras.
- Locuciones clausales: *como quien dice, como Dios manda, hacérsele a alguien agua la boca, subírsele a alguien, revolvérsele a alguien las tripas, ...*).

Los enunciados fraseológicos, de acuerdo a [14], constituyen cadenas autónomas del habla y, en cuanto a tales, se formulan con entonación independiente, como corresponde a su carácter de unidades mínimas de comunicación. En los enunciados fraseológicos se distingue entre las *paremias* y las *fórmulas rutinarias*. En la primera diferenciación tenemos ejemplos como *errar es humano perdonar es divino, las paredes oyen, poderoso caballero es don dinero, por la boca muere el pez*, entre otras; y en la segunda por ejemplo *a eso voy, ¿qué hay?, ¿qué te digo yo?, ¿puedo ayudarle?, hasta luego, ya lo creo*, entre otras.

En este artículo se decidió trabajar con las locuciones verbales que serán descritas en la subsección 2.1.

## 2.1. Locuciones verbales

Las locuciones verbales constituyen el núcleo de sintagmas verbales, es decir, están formadas por un núcleo verbal, acompañado por sus complementos. Desde el punto de vista sintáctico expresan procesos y actúan como los predicados, con o sin complementos. Estas unidades fraseológicas, igual que los verbos, se combinan con el sujeto y los complementos para formar una oración.

Las locuciones verbales presentan las mismas características como el resto de las locuciones, las cuales son: fijación interna, unidad de significado y fijación externa parsamática. La fijación es la característica más importante de las locuciones. Así, en las locuciones verbales, la forma fija significa que excepto el verbo, el resto de los elementos que forman parte de la locución no pueden modificarse, tampoco pueden sustituirse por otras palabras ni es posible introducir algo nuevo entre ellos. El verbo que constituye la locución cambia según la persona, el número, el tiempo y el modo de la oración. Sin embargo, la fijación de las locuciones no es absoluta, más bien es relativa y es posible encontrar una locución que tiene dos o más formas en las que es posible sustituir una de las palabras que la constituyen o donde es posible introducir entre las palabras de la locución otra palabra.

En general, las locuciones no constituyen actos de habla ni enunciados, es decir, necesitan combinarse con otros signos lingüísticos y que equivalen a sintagmas. Estas unidades fraseológicas, no constituyen enunciados completos y, generalmente, funcionan como elementos oracionales. Las locuciones son diferenciadas de las combinaciones libres de palabras de la lengua por su institucionalización, su estabilidad sintáctico-semántica y su función denominativa. La estabilidad es la característica esencial para delimitar unos fenómenos lingüísticos de otros; se incluyen tanto aspectos léxicos-semánticos como los morfosintácticos.

Por lo tanto, existen variadas pruebas para comprobar la cohesión semántica y morfosintáctica. El primer caso, se refleja en el carácter de unidad de significación en la lengua que presentan dichas unidades, ya tengan significado

compositivos (el significado se deduce de acuerdo a sus componentes de la locución) o traslaticio (el significado no se puede deducir de los componentes de la locución). Éste es complementado, con el segundo caso, cohesión morfosintáctica, traducido en la determinadas pruebas y operaciones formales, que comprueban no sólo la estabilidad formal de las locuciones, sino también su integridad semántica. Las principales pruebas aplicadas a los elementos de las locuciones son: a) Sustitución, b) Eliminación, c) Deficiencias transformativas.

Al hacer uso de las locuciones verbales, se debe de tener en cuenta que son propias de determinadas situaciones comunicativas. Esto significa que no todas las locuciones se pueden usar en cualquier situación comunicativa. Existen las que son propias de situaciones en las que existe mucha confianza entre los interlocutores (en los diccionarios se marcan como coloquiales, informales o familiares), por ejemplo: *aguar la fiesta*, *romper el hielo*, *ser harina de otro costal*, *traer cola*, y muchas más. Existen locuciones verbales que no tienen ninguna marca y por eso su uso es más amplio, por ejemplo *abrir paso*, *dar carpetazo*, *pasar la factura*, entre otras. Finalmente las locuciones verbales que tienen un significado vulgar, muchas veces contienen vulgarismos y su uso puede expresar la falta de respeto.

Cabe destacar que el uso principal de las locuciones verbales, y de cualquier locución, es su distribución geográfica, distinguiéndose así, locuciones de ámbito general y locuciones locales que se usan sólo en determinadas zonas.

## 2.2. Trabajo relacionado

Entre los autores interesados en el estudio de las unidades fraseológicas, podemos encontrar por un lado grupos vinculados a corrientes de lingüística teórica [23,24,4] y, por otro lado, corrientes vinculadas a la práctica terminográfica y la estandarización de unidades fraseológicas [3,1]. En las últimas décadas, sin embargo, ambas corrientes comparten el interés por las tecnologías de extracción automática de unidades fraseológicas. A partir del interés por la extracción de estas unidades, algunos autores se han centrado en identificar patrones sintácticos, morfológicos o la mezcla de ambos que ayuden a determinar la estructura interna de esta combinación de palabras.

En [10] se propone una herramienta basada en el etiquetado de las partes de la oración y el alineamiento de palabras para extraer candidatos a unidades fraseológicas y sus traducciones. La lista de unidades candidatas contiene frases de varias palabras y de una sola palabra. Las unidades fraseológicas de varias palabras coinciden con un conjunto de patrones sintácticos definidos por expresiones regulares y se identifican mediante una búsqueda en el documento etiquetado con las partes del discurso.

En [15], la indexación y recuperación de unidades fraseológicas se realiza mediante la combinación de la sintaxis y la morfosintaxis. Esencialmente lo que el sistema hace es tomar como entrada una lista de términos (unidades fraseológicas) y un corpus. La lista es precompilada manual o automáticamente, ésta es extendida detectando todas las variantes del término y marcando sus ocurrencias en el corpus.

En [18] se tiene como objetivo reconocer estas unidades lingüísticas en inglés, asignarles su significado y traducirlas en francés. Este proyecto incluye la identificación de las unidades fraseológicas, la construcción de un recurso léxico y la aplicación en la traducción. El sistema trata de mapear el término de una base de datos léxica que incluye la traducción, la información lingüística y metalingüística de las palabras para que de acuerdo a las partes de la oración, el estilo y dominio de la unidad fraseológica, las restricciones y demás características lingüísticas se identifique en el corpus la unidad y se proporcione su mejor traducción encontrada.

Existen en la literatura más trabajos relacionados con la temática de este artículo, sin embargo, en este trabajo de investigación no se pretende hacer un análisis exhaustivo sino presentar solamente aquellos considerados como los más relevantes partiendo de trabajos seminales. Para el caso del español, algunos trabajos dedicados al análisis morfosintáctico de las locuciones y que determinan diferentes tipos de ellas, se presentan en [6,8,27,13,5,7]. Adicionalmente, en [20,17,2,26,9,11,19] se analizan otro tipo de patrones (semántico, composicional, léxico), con el fin de extraer a estas unidades y determinar las características que podrían generalizarse en estas unidades lingüísticas.

En la siguiente sección se describe la metodología llevada a cabo para el análisis de la diversidad morfosintáctica de las locuciones verbales.

### **3. Metodología**

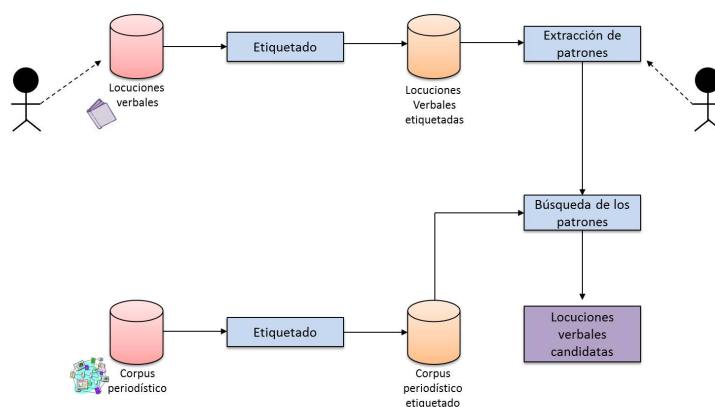
Con el fin de identificar los patrones morfosintácticos en las locuciones verbales, se parte de la taxonomía, de las locuciones realizada por [7], anteriormente descrita, en su clasificación de las unidades fraseológicas en español. A partir de dicha taxonomía, se decide centrarse en las locuciones verbales, debido a que la mayoría de frases está formada por un verbo y una o varias variables. El verbo exige y realiza una rigurosa selección de los sujetos y de los componentes que pueden acompañarle. Estas frases se encuentran fusionadas en la oración para enunciar algo de manera más amplia, pero al separarse de la oración tienen sentido completo, es decir, tienen información semántica por ellas mismas y constituyen el núcleo de sintagmas verbales.

En dicha taxonomía se clasifican los tipos de locuciones verbales de acuerdo a su variedad morfosintáctica, los cuales comprenden: a) Locuciones formadas por dos núcleos verbales unidos por conjunción, b) Locuciones compuestas de verbo y pronombre, c) Locuciones compuestas de verbo, pronombre y partícula, d) Locuciones de verbo más partícula asociada a éste, con complemento opcional, e) Locuciones formadas por verbo copulativo más atributo, f) Locuciones formadas por verbo más complemento circunstancial, g) Locuciones formadas por verbo más suplemento h) Locuciones formadas por verbo más objeto directo y i) Locuciones negativas. En este trabajo son denominadas como Tipo 1, Tipo 2, ..., Tipo 9; respectivamente. Con base en esta taxonomía, se prosigue a inquirir ejemplos de locuciones verbales que cumplan con la variedad morfosintáctica y que ayuden a determinar los patrones morfosintácticos.

En cuanto a los ejemplos utilizadas, se emplearon las locuciones verbales presentadas en [22], debido a que éstas fueron recuperadas manualmente. Posteriormente, estas locuciones verbales se clasifican de acuerdo a los tipos de la taxonomía empleada según sus componentes. Una vez clasificadas se necesita saber su estructura morfosintáctica para de esta manera obtener los patrones, así que, las locuciones fueron etiquetadas con FreeLing<sup>1</sup>. En la tabla 1 se presenta una muestra de las locuciones verbales identificadas de acuerdo a su tipo y sus respectivas etiquetas morfosintácticas<sup>2</sup>.

Para la búsqueda de los patrones morfosintácticos identificados, se seleccionó un fragmento del corpus periodístico presentado en [21], el cual contiene aproximadamente 1,960,373 palabras. La identificación de los patrones morfosintácticos en el corpus se ha realizado de dos diferentes maneras, una tomándolo en cuenta el contexto y la otra sin tomarlo en cuenta. En la primera aproximación, se ha utilizado una ventana de cinco palabras a la izquierda de la locución verbal y cinco palabras a la derecha, denominándolas *contexto izquierdo* y *contexto derecho*, respectivamente.

Básicamente la metodología propuesta de este trabajo considera tener dos elementos esenciales: 1) Una lista de locuciones verbales, y 2) Un conjunto de textos, ambos etiquetados morfosintácticamente. Del primer recurso léxico se obtienen los patrones morfosintácticos, y éstos son buscados en el corpus de textos con la finalidad de obtener una lista de posibles locuciones verbales, las cuales concuerdan con los patrones morfosintácticos obtenidos de las locuciones semilla (ver figura 1).



**Fig. 1.** Metodología empleada para la identificación de patrones morfosintácticos en las locuciones verbales.

<sup>1</sup> Para más información de la herramienta, consultar <http://nlp.lsi.upc.edu/freeling/>

<sup>2</sup> Para una referencia del significado del etiquetado morfológico de FreeLing referirse a <http://nlp.lsi.upc.edu/freeling/doc/tagsets/tagset-es.html>

**Tabla 1.** Ejemplo de locuciones verbales identificadas de acuerdo a sus etiquetas morfosintácticas.

Tipo de locución verbal	Ejemplos	Etiquetas morfosintácticas (resultados de FreeLing)
Tipo 1	dar y tomar ir y venir llevar y traer	VMN0000 CC VMN0000 VMN0000 CC VMN0000 VMN0000 CC VMN0000
Tipo 2	apañársela arreglársela cargársela	VMN0000 PP3CN000 PP3FSA00 VMN0000 PP3CN000 PP3FSA00 VMN0000 PP3CN000 PP3FSA00
Tipo 3	brincarse la barda darse su taco tomarla con (alguien/algo)	VMN0000 PP3CN000 DA0FS0 NCFS000 VMN0000 PP3CN000 DP3CS0 NCMS000 VMN0000 PP3FSA00 SPS00 (PI0CS000/PI0CS000)
Tipo 4	dar de sí ir con (uno) tomar (algo/a alguien) por	VMN0000 SPS00 CS VMN0000 SPS00 PI0MS000 VMN0000 (PI0CS000/SPS00 PI0CS000) SPS00
Tipo 5	ser ajonjolí de todos los moles  ser el vivo retrato de alguien	VSN0000 AQ0CS0 SPS00 DI0MP0 NCMP000 VSN0000 DA0MS0 AQ0MS0 NCMS000 SPS00 PI0CS000 VSN0000 AQ0CS0
Tipo 6	ser gacho decir hasta la despedida dormir como un tronco meter a alguien en cintura	VMN0000 SPS00 DA0FS0 NCFS000 VMN0000 CS DI0MS0 NCMS000 VMN0000 SPS00 PI0CS000 SPS00 NCFS000
Tipo 7	meter las cuatro oler a cuero quemado	VMN0000 DA0FP0 Z VMN0000 SPS00 NCMS000 VMP00SM
Tipo 8	pagar el pato chuparse el dedo  mover cielo y tierra saber de qué pie cojea alguien	VMN0000 DA0MS0 NCMS000 VMN0000 PP3CN000 DA0MS0 NCMS000 VMN0000 NCMS000 CC NCFS000 VMN0000 SPS00 DT0CN0 NCMS000 VMIP3S0 PI0CS000
Tipo 9	no haber vuelta de hoja no poder ver ni en pintura a alguien  no tener un pelo de tonto	RN VMN0000 NCFS000 SPS00 NCFS000 RN VMN0000 VMN0000 CC SPS00 NCFS000 SPS00 PI0CS000 RN VMN0000 DI0MS0 NCMS000 SPS00 NCMS000

#### 4. Resultados

En este trabajo de investigación se han identificado 34 patrones morfosintácticos que sirven como semilla para encontrar posibles locuciones verbales dentro de un corpus de textos. Dichos patrones han sido obtenidos mediante el etiquetado



morfosintáctico de una lista semilla de 43 locuciones verbales. En la tabla 2 se presenta una muestra de los patrones morfosintácticos identificados como más frecuentes en el corpus de textos periodísticos.

**Tabla 2.** Muestra de patrones morfosintácticos de las locuciones verbales.

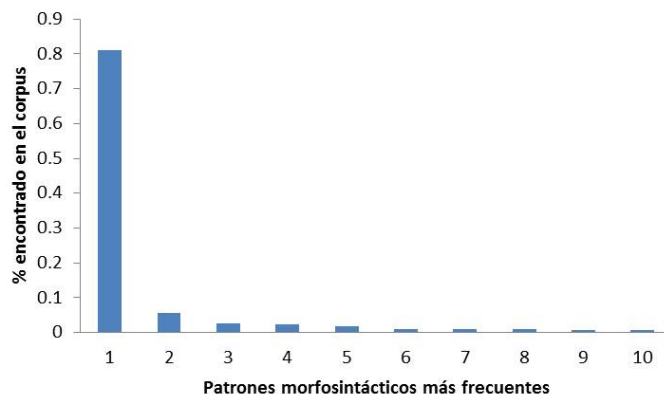
Estructura sintáctica	Patrón morfosintáctico
V + Prep	VMN0000 SPS00
V + Det + Nom + Adj	VMN0000 DA0FS0 NCFS000 AQ0CS0
V + Pron + Prep	VMN0000 PP3FSA00 SPS00
V + Prep + Conj	VMN0000 SPS00 CS
V + Pron + Det + Nom	VMN0000 PP3CN000 DA0MS0 NCMS000
V + Det + N + Prep + Det + Nom	VMN0000 DI0MS0 NCMS000 SPS00 DA0FS0 NCFS000
V + Det + Nom	VSN0000 DA0FS0 NCFS000
V + Prep + Pron	VMN0000 SPS00 PI0CS000
V + Prep + Pron	VMN0000 SPS00 PI0MS000
V + Pron + Prep	VMN0000 PI0CS000 SPS00

**Tabla 3.** Ejemplo de las locuciones verbales encontradas en el corpus periodístico.

Frecuencia de aparición	Locución verbal candidata
357	llegar/llegar/VMN0000 a/a/SPS00
201	contar/contar/VMN0000 con/con/SPS00
152	participar/participar/VMN0000 en/en/SPS00
117	tratar/tratar/VMN0000 de/de/SPS00
113	apoyar/apoyar/VMN0000 a/a/SPS00
110	cumplir/cumplir/VMN0000 con/con/SPS00
106	salir/salir/VMN0000 de/de/SPS00
99	ir/ir/VMN0000 a/a/SPS00
93	ver/ver/VMN0000 con/con/SPS00
90	acudir/acudir/VMN0000 a/a/SPS00

En la Tabla 3 se presenta un ejemplo de las 10 locuciones verbales encontradas como más frecuentes en el corpus y que empatan con el patrón morfosintáctico indicado en la misma Tabla. En total, se extrajeron 3,083 resultados coincidentes con los patrones registrados.

En la figura 2 se puede observar que de los 10 patrones morfosintácticos más frecuentes (ver Tabla 2), el primero obtiene un 80% de cobertura con respecto a los demás. Esto se encuentra derivado del hecho de ser un patrón demasiado general que parte de locuciones verbales semilla tales como: “ir con”.



**Fig. 2.** Porcentaje de las locuciones verbales más frecuentes encontradas en el corpus.

Cabe mencionar que de los 34 patrones morfosintácticos detectados a partir de las locuciones semilla, solamente se encontraron coincidencias sobre 18. Esto significa, que 16 patrones no han arrojado posibles locuciones verbales. En la Tabla 4 se muestran tales patrones; una discusión sobre los mismos sigue a continuación.

Observando los patrones que no encontraron coincidencias en el corpus de textos podemos ver que en general se trata de una secuencia no habitual de etiquetas morfológicas, cuya frecuencia es muy baja en los corpus textuales. El corpus utilizado tiene únicamente 5,000 noticias (361 palabras en promedio por noticia), y por tanto, la probabilidad de encontrar una de estas secuencias es muy baja.

## 5. Conclusiones y perspectivas

En este trabajo de investigación se presentan experimentos tendientes a la identificación automática de locuciones verbales a partir de textos planos. La metodología propuesta indica tomar como entrada un conjunto de locuciones verbales semilla que sirvan para encontrar un conjunto de patrones morfosintácticos, los cuales son posteriormente utilizados para encontrar coincidencias de los mismos sobre un corpus de textos (en nuestro caso, fue del género periodístico).

El experimento realizado sobre un conjunto inicial de 43 locuciones verbales, permitió encontrar 34 patrones morfosintácticos. De éstos, únicamente 18 encontraron coincidencias en el corpus de textos.

Como trabajo a futuro se desea ampliar el corpus de textos para determinar si es posible encontrar coincidencias de todos y cada uno de los patrones morfosintácticos detectados a partir de las locuciones verbales semilla. Adicionalmente, sería importante evaluar el filtrado de locuciones verbales candidatas para incrementar la precisión en la identificación de las mismas.

**Tabla 4.** Conjunto de patrones morfosintácticos de los cuales no se encontraron coincidencias en el corpus de textos.

---

**Patrón morfosintáctico**

---

VMN0000 PP3FSA00 SPS00 PI0CS000  
VSN0000 DA0MS0 AQ0MS0 NCMS000 SPS00 PI0CS000  
VMN0000 Z NCFP000 SPS00 PI0CS000  
VMN0000 SPS00 PI0CS000 SPS00 NCFS000  
VMN0000 SPS00 PI0CS000 CS VMIP3S0 SPS00 NCMS000  
VMN0000 SPS00 DT0CN0 NCMS000 VMIP3S0 PI0CS000  
VMN0000 PP3CN000 DA0MP0 NCMP000 SPS00 DA0FS0 NCFS000  
VMN0000 DA0FS0 NCFS000 SPS00 NCFS000 SPS00 DA0MS0 VMP00SM  
VMN0000 DA0FP0 SPS00 NP00000  
RN VMN0000 Z NCMP000 SPS00 RG  
RN VMN0000 VMN0000 CC SPS00 NCFS000 SPS00 PI0CS000  
RN VMN0000 SPS00 DT0CN0 NCFS000 VMN0000 PP3CN000  
RN VMN0000 SPS00 DA0FS0 NCFS000 DA0FS0 AQ0FS0  
RN VMN0000 PP3CSD00 CC VMN0000 PP3CSD00 PI0CS000 SPS00 PI0CS000 SPS00 PI0CS000  
RN VMN0000 NCMS000 CC VMIP1S0  
RN VMN0000 NCFS000 SPS00 NCFS000

---

## Referencias

1. Arntz, R., Picht, H.: Introducción a la terminología. In: Fundación Germán Sánchez Ruipérez. Barcelona (1988)
2. Baldwin, T.: Deep lexical acquisition of verb-particle constructions. *Comput. Speech Lang.* 19(4), 398–414 (Oct 2005), <http://dx.doi.org/10.1016/j.csl.2005.02.004>
3. Cabré, T., Estopá, R.: Introducción a la teoría general de la terminología y a la lexicografía. In: Institut Universitari de Lingüística Aplicada. Barcelona (1979)
4. Cabré, T., Estopá, R.: Unidades de conocimiento especializado, caracterización y tipología. In: Cabré, M. T.; Bach, C. (eds.) *Coneixement, llenguatge i discurs especialitzat*. Barcelona (2005)
5. Carneado Moré, Z., Tristán Pérez, A.M.: *Estudios de la fraseología*. La Habana: Academia de Ciencias de Cuba. Instituto de literatura y lingüística (1983)
6. Casares, J.: *Introducción a la lexicología moderna*. In: C.S.I.C. Madrid (1950)
7. Corpas Pastor, G.: *Manual de fraseología española*. Gredos, Madrid (1996)
8. Coseriu, E.: Structure lexicale et enseignement du vocabulaire. In: *Actes du premier colloque international de linguistique appliquée*. pp. 175–217 (1966)
9. Van de Cruys, T., Moirón, B.n.V.: Semantics-based multiword expression extraction. In: *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions*. pp. 25–32. MWE '07, Association for Computational Linguistics, Stroudsburg, PA, USA (2007), <http://dl.acm.org/citation.cfm?id=1613704.1613708>
10. Dagan, I., Church, K.W.: Termight: Identifying and translating technical terminology. In: *ANLP*. pp. 34–40 (1994), <http://dblp.uni-trier.de/db/conf/anlp/anlp1994.html#DaganC94>

11. Davis, A.R., Barrett, L.: Lexical semantic factors in the acceptability of english support-verb-nominalization constructions. *ACM Trans. Speech Lang. Process.* 10(2), 5:1–5:15 (Jun 2013), <http://doi.acm.org/10.1145/2483691.2483694>
12. Gramley, S., Pätzold, K.M.: *A survey of modern English*. Londres-Nueva York, Routledge (1992)
13. Haensch, Wolg, G., Ettinger, L., Werner, S.: *La lexicografía. De la lingüística teórica a la lexicografía práctica*. Gredos, Madrid (1982)
14. Hernando Cuadrado, L.: Sobre las unidades fraseológicas en español. In: *Actas de la Sociedad Española de Lingüística. XX Aniversario*. vol. 1, pp. 538–546 (1990)
15. Jacquemin, C., Klavans, J.L., Tzoukermann, E.: Expansion of multi-word terms for indexing and retrieval using morphology and syntax. In: *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics*. pp. 24–31. EAACL '97, Association for Computational Linguistics, Stroudsburg, PA, USA (1997), <http://dx.doi.org/10.3115/979617.979621>
16. Martins Baltar, M.: *La locution entre langue et usages*. In: ENS Editions, Fontenay-St. Cloud (1997)
17. McCarthy, D., Keller, B., Carroll, J.: Detecting a continuum of compositionality in phrasal verbs. In: *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment - Volume 18*. pp. 73–80. MWE '03, Association for Computational Linguistics, Stroudsburg, PA, USA (2003), <http://dx.doi.org/10.3115/1119282.1119292>
18. Michiels, A., Dufour, N.: Defi, a tool for automatic multi-word unit recognition, meaning assignment and translation selection. In: *Proceedings of the first international conference on language resources and evaluation*. pp. 1179–1186 (1998)
19. Nissim, M., Zaninello, A.: Modeling the internal variability of multiword expressions through a pattern-based method. *ACM Trans. Speech Lang. Process.* 10(2), 7:1–7:26 (Jun 2013), <http://doi.acm.org/10.1145/2483691.2483696>
20. Piao, S.S.L., Rayson, P., Archer, D., Wilson, A., McEnery, T.: Extracting multiword expressions with a semantic tagger. In: *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment - Volume 18*. pp. 49–56. MWE '03, Association for Computational Linguistics, Stroudsburg, PA, USA (2003), <http://dx.doi.org/10.3115/1119282.1119289>
21. Priego Sánchez, B., Pinto, D., Mejri, S.: Metodología para la identificación de secuencias verbales fijas. *Research in Computing Science* 85, 45–56 (2014), [http://rcs.cic.ipn.mx/2014\\_85/Metodologia%20para%20la%20identificacion%20de%20secuencias%20verbales%20fijas.pdf](http://rcs.cic.ipn.mx/2014_85/Metodologia%20para%20la%20identificacion%20de%20secuencias%20verbales%20fijas.pdf)
22. Priego Sánchez, B., Pinto, D., Mejri, S.: Towards the automatic identification of spanish verbal phraseological units. *Research in Computing Science* 96, 65–73 (2015), [http://rcs.cic.ipn.mx/2015\\_96/Towards%20the%20Automatic%20Identifi\\_cation%20of%20Spanish%20Verbal%20Phraseological%20Units.pdf](http://rcs.cic.ipn.mx/2015_96/Towards%20the%20Automatic%20Identifi_cation%20of%20Spanish%20Verbal%20Phraseological%20Units.pdf)
23. Sager, J.: *A practical course in terminology processing*. In: Amsterdam/Philadelphia: John Benjamins (1990)
24. Sager, J.: *La terminología: representación y comunicación*. In: Institut Universitari de Lingüística Aplicada. Barcelona (1999)
25. Thun, H.: Probleme der phraseologie. In: *Beihefte zur zeitschrift für romanische philologie* 168. Tubinga, Max Niemeyer (1978)
26. Zhang, Y., Kordoni, V., Villavicencio, A., Idiart, M.: Automated multiword expression prediction for grammar engineering. In: *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*. pp. 36–44. MWE '06, Association for Computational Linguistics, Stroudsburg, PA, USA (2006), <http://dl.acm.org/citation.cfm?id=1613692.1613700>

27. Zuluaga, A.: La función del diminutivo en español. In: *Thesaurus XXV*. pp. 23-48 (1980)