



Revista Española de Lingüística

Órgano de la Sociedad Española de Lingüística

RSEL

45|2

Julio-Diciembre 2015

Edita
SeL

REVISTA ESPAÑOLA DE LINGÜÍSTICA
(RSEL)
45/2

Edita

SeL

REVISTA ESPAÑOLA DE LINGÜÍSTICA (RSEL)

ISSN: 0210-1874 · eISSN: 2254-8769

Depósito Legal: M-24.769-1971

DIRECTOR DE HONOR: D. Francisco Rodríguez Adrados (RAE y RAH).

DIRECTOR: Juan Antonio Álvarez-Pedrosa Núñez (UCM).

SECRETARIO: Luis Unceta Gómez (UAM).

CONSEJO DE REDACCIÓN: Montserrat Benítez (CSIC), José Antonio Berenguer (CSIC), Joaquín Garrido (UCM), Juana Gil Fernández (CSIC), Salvador Gutiérrez Ordóñez (U. León y RAE), Antonio Hidalgo (U. Valencia), Patricia Infante (CSIC), Manuel Leonetti (U. Alcalá), Eugenio Luján (UCM), Victoria Marrero (UNED), Ventura Salazar (U. Jaén), Esperanza Torrego (UAM).

CONSEJO ASESOR: Alberto Bernabé (UCM), Margarita Cantarero (SEL), Ramón Cerdá (UB), Victoria Escandell (UNED), Marina Fernández Lagunilla (UAM), José Manuel González Calvo (U. Extremadura), Emma Martinell (UB), Juan Carlos Moreno Cabrera (UAM), Gregorio Salvador (RAE), José Carlos de Torres (SEL), Jesús de la Villa (UAM).

A partir del número 38 (2008) la *Revista Española de Lingüística* ha recuperado el formato de dos fascículos al año, con periodicidad semestral. Los trabajos enviados para su publicación han de dirigirse al Secretario de la revista. Deberán ser originales e inéditos y ajustarse a las normas que aparecen en el número 38/2, así como en la página web de la Sociedad Española de Lingüística. Todos los trabajos son sometidos al dictamen de al menos dos evaluadores designados por el Consejo de Redacción, mediante informes de carácter confidencial.

Los derechos de publicación y difusión, bajo cualquier forma, son propiedad de *RSEL*. Todo texto publicado en la revista obliga a sus autores a no cederlo a terceros, sin autorización previa de la revista, quien sí queda autorizada a comercializarlo, debiendo entregar, en este caso, el 50% de los beneficios obtenidos a sus autores.

REDACCIÓN: Sociedad Española de Lingüística, Centro de Ciencias Humanas y Sociales del CSIC, c/ Albasanz, 26 - 28, 28037 Madrid.

Correo electrónico: secretarioRSEL@gmail.com. <<http://www.sel.edu.es/>>

DISEÑO Y COMPOSICIÓN: Carmen Chíncoa & Carlos Curiá (produccionRSEL@gmail.com).

SERVICIOS DE INFORMACIÓN: Los contenidos de la *RSEL* son recogidos sistemáticamente en *Bibliographie Linguistique / Linguistic Bibliography*, *CINDOC – Base de datos Sumarios ISOC, CSA – Linguistic and Language Behavior Abstracts, Dialnet, Francis, Modern Language Association (MLA) Bibliography*.

ÍNDICE 45/2 (2015)

ARTÍCULOS

<i>Conexión discursiva en aprendices anglohablantes de ELE</i>	7
JOSÉ MANUEL BUSTOS GÍSBERT	
<i>Grupos adverbiales comparativos y locuciones adverbiales focales en español</i>	41
P. PABLO DEVÍS MÁRQUEZ	
<i>Las preposiciones de tema/asunto en español: un análisis de sinonimia prepositiva</i>	69
ANTON GRANVIK	
<i>Intensividad y preposiciones de trayectoria: la sintaxis dialectal de hasta y desde</i>	95
PABLO RICO RAMA	
<i>La modalidad en la gramática: las construcciones saber/poder + infinitivo</i>	115
AXELLE VATRICAN	
<i>Pragmática, gramática e inferencias escalares</i>	143
BEGOÑA VICENTE CRUZ	
NOTAS E INFORMACIÓN.....	173
RESEÑAS	177

LAS PREPOSICIONES DE TEMA/ASUNTO EN ESPAÑOL: UN ANÁLISIS DE SINONIMIA PREPOSITIVA

ANTON GRANVIK

HANKEN School of Economics, Helsinki

Universidad de Gotemburgo

RESUMEN

Este trabajo investiga el uso de las preposiciones *de*, *sobre*, *acerca de* y *en torno a* para expresar el tema o asunto de una conversación, de donde procede la denominación de preposiciones de ‘tema/asunto’. El objetivo es averiguar qué diferencias hay en el uso de estas preposiciones, partiendo del llamado Principio-de-no-sinonimia (Goldberg 1995, p. 67). Basándome en la idea de que el contexto de uso puede revelar diferencias entre estas expresiones sinónimas dentro del ámbito de tema/asunto, he anotado 100 casos de cada preposición extraídos del *Corpus de español* de Davies 2002, según 22 factores gramaticales. Estos datos han sido analizados cuantitativamente mediante análisis de regresión logística (multinomial y binomial por pares) y sometidos a un agrupamiento jerárquico. Los resultados indican que especialmente *de* se distingue de las demás, mientras que *sobre*, *acerca de* y *en torno a* presentan usos tan semejantes que resulta difícil diferenciarlas nítidamente.

Palabras clave: sinonimia; preposiciones; tema/asunto; análisis de corpus; métodos cuantitativos.

ABSTRACT

This paper focuses on the Spanish prepositions *de*, *sobre*, *acerca de* and *en torno a*, which are all used to express the subject or topic of a conversation. Hence the label topic-marking prepositions. The aim is to find out what differences there are in the usage of these prepositions, taking as a starting point Goldberg’s 1995, p. 67, so-called Principle-of-No-Synonymy. Building on the assumption that the analysis of the usage context of each preposition can bring forth differences among these otherwise synonymous expressions, I annotated 100 cases of each preposition extracted from Davies’ 2002 *Corpus del español* for a total of 22 grammatical factors. The data was analyzed quantitatively by running a number of logistic regression analyses (multinomial and pair-wise binomial), as well as a hierarchical cluster analysis. The results show some clear differences, especially between *de* and the other three, while *sobre*, *acerca de* and *en torno a* are highly similar and hard to distinguish from one another.

Keywords: synonymy; topic-marking prepositions; Spanish; corpus analysis; quantitative methods.

RECIBIDO: 12/12/2014

APROBADO: 09/07/2015

1. INTRODUCCIÓN

El objetivo del presente estudio es investigar la relación semántica que se establece, típicamente, entre determinados verbos, sustantivos y adjetivos de habla o de pensamiento y su tema, asunto, o tópico. Esta relación la he llamado tema/asunto en una serie de trabajos anteriores (cf. Granvik 2012, 2014, Granvik y Taimitarha 2014), y las expresiones preposicionales que la establecen pueden, pues, denominarse marcadores de tema/asunto o, muchas veces simplemente, preposiciones de tema/asunto (PTA). En español, los marcadores de tema/asunto más frecuentes son las preposiciones *de* y *sobre*, a las que pueden añadirse, al menos, las siguientes locuciones prepositivas: *acerca de*, *en torno a* y *(con) respecto a/de*. Ahora bien, especialmente los usos de *con respecto a/de* se alejan sintácticamente de los usos claramente prepositivos de los marcadores mencionados antes, y se asemejan a otro grupo de expresiones que Pons 2001-2002, 2005 ha llamado «tematizadores», como *en cuanto a*, *en lo que se refiere a*, *en lo tocante a*, *en relación a/con*, *relacionado con*, etc.

El punto de partida de este trabajo es la relación de tema/asunto, que constituye una noción semántica de gran importancia para nuestra comunicación diaria. ¿Qué hacemos en nuestras conversaciones sino «hablar *de* algo»? La Figura 1 supone una esquematización de esta noción en términos visuales, donde el tema/asunto se representa como el ámbito temático que tienen en mente los participantes en la situación comunicativa, y con respecto al cual se expresan.

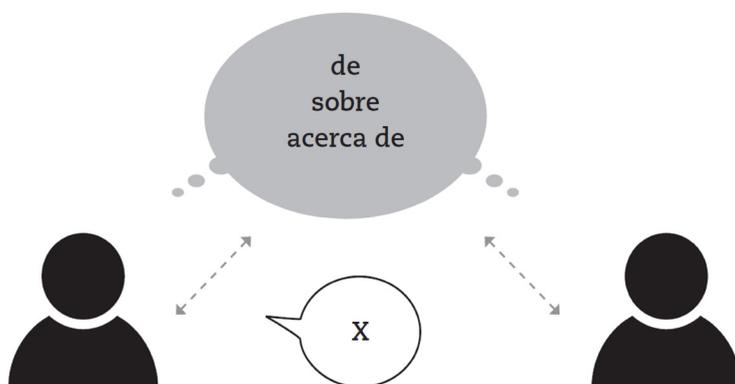


Figura 1. Representación esquemática de la noción de tema/asunto (Granvik 2014, p. 83).

Otro indicio de la importancia de esta noción para el sistema lingüístico es el hecho de que existan tantas expresiones que la invocan, como revelan los siguientes ejemplos:

- (1) Esto era muy difícil de recomendar. Yo *hablé*, por ejemplo, *de* esto mismo con uno de los del grupo de los doce y... (CE, Habla Culta, La Paz, M5)
- (2) El secretario de Régimen Interior y Policía, Hugo San Martín, postergó para hoy su asistencia a la Comisión de Derechos Humanos de la Cámara de Diputados donde debe brindar *un informe sobre* las acciones policiales contra la marcha de productoras de coca. (CE, Bolivia, ERBOL,10/31/96)
- (3) El ministro de Economía, Domingo Cavallo, aseguró el viernes que no *habrá dudas acerca de* que las metas de enero a marzo fueron cumplidas. (CE, Arg:Prensa, 21_PRRÉ)
- (4) hay *questionamientos* muy profundos de consejeras y consejeros *en torno a* lo que se ha hecho público (CE, Entrevista (PAN))
- (5) esa síntesis, de erudición, que se acrecienta con datos personalmente adquiridos, observaciones de crítica profunda *con respecto a* su obra. (CE, *El que vendrá*, XIX)
- (6) *En cuanto a* la economía, los liberales han luchado contra los monopolios y las políticas de Estado que han intentado someter la economía a su control. (CE, *Enc: Liberalismo*)
- (7) *En lo que se refiere a* la muerte de Mozart, es muy difícil hablar hoy día de «misterios». (CE, ABC)

Como revelan los ejemplos (1) a (7), se trata esencialmente de la misma noción semántica en todos ellos, pero solo los cinco primeros pueden considerarse claramente casos prepositivos, en el sentido de tener tanto un elemento regente como uno regido. Para completar la caracterización de la noción de tema/asunto, considérese el ejemplo (8), donde no aparece ya ningún marcador de tema/asunto explícito, pero aun así no hay dificultad en averiguar de qué trata la predicación.

- (8) Con esa, te digo que me he convencido de que es un cuadrado, ja... ja... ja... (CE, Habla Culta: La Paz: M29)

Considerando todo el ámbito semántico de tema/asunto como un *continuum*, puede establecerse una relación evidente entre la noción de tema/asunto tal y como la manejo en este trabajo, y la de tópico o tema en sentido más tradicional (cf. Zubizarreta 1999, p. 4218)¹.

Ahora bien, si puede hacerse referencia a la noción de tema/asunto de tantas maneras diferentes, usando tantas expresiones lingüísticas diferentes, ¿cuál es la diferencia entre ellas? ¿Son todas equivalentes? Estas son las preguntas iniciales que me propongo esclarecer en este trabajo, si bien me limito a analizar con más profundidad cuatro expresiones que

1. A este valor corresponden también la noción de *aboutness* de Langacker 2008, p. 513, y el tema (*theme*) de Halliday 1967, p. 212.

inequívocamente pueden caracterizarse como preposiciones de tema/asunto (PTA), y que son: *de*, *sobre*, *acerca de* y *en torno a*. Más específicamente, me interesa revelar qué diferencias hay en el uso de estas preposiciones. Dicho de otro modo: partiendo de la idea de que el contexto de uso puede revelar cosas interesantes sobre el significado de una palabra, mi objeto es dar cuenta de qué factores gramaticales pueden explicar por qué se usa una PTA y no otra. Por ejemplo, ¿hay algo en el contexto del ejemplo (1) que nos pueda dar una pista sobre por qué el hablante en este caso se decantó por la preposición *de* en lugar de *sobre*, por ejemplo, una vez que esta habría sido igualmente aceptable?

El trabajo está estructurado de la siguiente manera. En la Sección 2, paso revista a las bases teóricas más relevantes para emprender el estudio. Empezaré por comentar brevemente el sistema prepositivo español y su funcionamiento, aunando este aspecto con el marco teórico de la Gramática de construcciones (GCx, *Construction Grammar*) en la que se inserta este trabajo. En esta sección también se comenta la idea de sinonimia desde la perspectiva particular de las preposiciones. En la Sección 3 se presentan los métodos y el corpus empleados, y en la Sección 4 se emprende el análisis. Los resultados se resumen y discuten en la Sección 5. La Sección 6 cierra el trabajo con unas reflexiones finales sobre la realización del trabajo y posibles elaboraciones futuras.

2. LAS PREPOSICIONES ESPAÑOLAS Y LA CUESTIÓN DE LA SINONIMIA DESDE LA PERSPECTIVA DE LA GRAMÁTICA DE CONSTRUCCIONES

Como es bien sabido, el español cuenta con una veintena de preposiciones simples, las cuales se ven complementadas por un amplio inventario de locuciones prepositivas, que incluye cientos de unidades distintas (cf. *NGLE*, §29.2a; Cifuentes Honrubia 2003). De una forma simplificada, puede constatarse que mientras las preposiciones sencillas son elementos gramaticales fundamentales, las locuciones prepositivas se han venido formando a lo largo del tiempo según el sistema se ha visto necesitado de modos de expresión alternativos.

En este sentido, la introducción de las locuciones prepositivas puede verse como una respuesta a la creciente necesidad de concretizar, enfatizar o especificar ciertos significados, para desambiguar los variados usos de las preposiciones simples y para evitar su uso excesivo. La expresión de tema/asunto constituye aquí un buen ejemplo, pues mientras que *de* y *sobre* se usan desde los testimonios más tempranos para expresar esta

idea, *acerca de* y *en torno a/de* aparecen más tarde, en los siglos XIV y XVIII, respectivamente (cf. Granvik 2014)².

- (9) a. *de civitate dei* (Bassols de Climent 1967-I)
 b. *sed hac super re nimis dixi* (Bassols de Climent 1967-I)
 c. Andén de Venus e Mercurio *Acerca de* lo cual devezes saber que cada uno de éstos tiene en la su sobrefaz dos ecéntricos iguales en su grandez (*Tratado de astrología*, CE, XV)
 d. dicen que en las naciones turbulentas pasiones se levantan *en torno a* vuestra gloria. (*Poesías*, CE, XIX)

Otro aspecto relevante de las cuatro preposiciones de tema/asunto que constituyen el centro de interés de este trabajo, es que todas ellas tienen su uso particular en la lengua. Así, aunque *de* y *sobre* son las PTA más frecuentes del español (véase la Tabla 2, abajo), como revelan los datos reunidos en la Tabla 1, ambas tienen muchos otros significados y usos aparte de expresar tema/asunto (cf. Granvik 2012). *Acerca de*, en cambio, se usa únicamente como PTA en la lengua actual, pues el valor originalmente locativo lo ha asumido su casi homófono, *cerca de* (cf. Granvik 2014, p. 94). Finalmente, *en torno a/de* presenta principalmente usos locativos, si bien los valores abstractos de aproximación y tema/asunto son bastante patentes, alcanzando este último una tercera parte de sus usos actuales (Granvik 2014, p. 102).

Preposición	Estructura semántica	Significados
<i>de</i>	Polisémica	Tema/asunto uno de docenas de significados
<i>sobre</i>	Polisémica	Tema/asunto uno de muchos significados
<i>acerca de</i>	Monosémica	Tema/asunto
<i>en torno a</i>	Polisémica	Locativo, Aproximación, Tema/asunto

Tabla 1. Usos y significados de las cuatro preposiciones de tema/asunto.

Con respecto al funcionamiento de las preposiciones en el discurso cabe mencionar, en primer lugar, la necesidad de que toda preposición, como elemento de relación que es, vaya precedida por un elemento regente y seguida por uno regido. Esta presencia obligada de elementos sintácticamente adyacentes a las preposiciones tiene una relevancia especial desde la perspectiva de la Gramática de construcciones (GCx). Este

2. Granvik 2014 presenta un panorama diacrónico del establecimiento en la lengua de las locuciones prepositivas *acerca de*, *en torno a/de* y *con respecto a/de* con valor de tema/asunto, relacionando su aparición y extensión con las nociones de gramaticalización y lexicalización.

acercamiento a la gramática postula que las unidades lingüísticas básicas no son los lexemas o morfemas, sino las construcciones. Aunque estas se definen simplemente como combinaciones de forma y significado, con respecto a las interpretaciones tradicionales la novedad de la GCx es que la «forma» puede corresponder a una unidad sintáctica de complejidad variable: así, son construcciones por igual las desinencias casuales, las palabras, la marca de pluralidad, las relaciones preposicionales, las estructuras transitivas y las construcciones pasivas, etc. (cf. Croft 2001, pp. 14-18; González-García 2012, pp. 261-262)³.

Desde la perspectiva de las preposiciones de tema/asunto puede considerarse que la relación de tema/asunto constituye una construcción esquemática, en la que el elemento fijo es la PTA (o, una de ellas), mientras que son variables los elementos que funcionan como regente y regido de la relación preposicional (ejemplo (10)). Por otro lado, en un nivel más específico, es posible asimismo considerar que cada preposición constituye su propia construcción (ejemplos (11a-d)). Desde este punto de vista, la relación de tema/asunto expresada por *de* sería una construcción diferente a la que establecen *sobre*, *acerca de* y *en torno a*, respectivamente.

(10) Cx de tema/asunto	Rgdo	PTA	Rte
(11) a. Cx- <i>de</i>	Rgdo	<i>de</i>	Rte
b. Cx- <i>sobre</i>	Rgdo	<i>sobre</i>	Rte
c. Cx- <i>acerca de</i>	Rgdo	<i>acerca de</i>	Rte
d. Cx- <i>en torno a</i>	Rgdo	<i>en torno a</i>	Rte

En realidad, lo que aquí está en juego es si es lícito considerar que cada preposición establece una construcción propia. Este sería el caso si las características contextuales fueran diferentes de una preposición a otra. Por otro lado, si los contextos de uso de una y otra preposición coincidieran (perfectamente), habría que considerar que se trata simplemente de una construcción esquemática superior al nivel del lexema, y que las preposiciones de tema/asunto son sinónimas.

En términos de sinonimia, el estudio se enfoca en el nivel textual, pues lo que me interesa no son ni las posibles diferencias sociolingüísticas o estilísticas (como el género textual), ni la denotación semántica, sino el uso contextual de las cuatro preposiciones⁴. Siguiendo la línea de investigación iniciada por Stefanowitsch y Gries 2003, y Gries, y Stefanowitsch

3. Véase asimismo la reciente obra colectiva editada por Boas y González-García 2014 para una serie de estudios sobre las lenguas románicas desde distintas perspectivas construccionistas.

4. En la terminología de González Martínez 1988-1989, se trata de sinonimia de sentido: «al considerar este tipo de contenido lingüístico entran en juego factores pragmáticos de toda índole, tales como los derivados de la influencia de los contextos (con-textos y co-textos) en los que se sitúan los actos de habla que van a dar como producto el texto que intentemos estudiar».

2004, que establecen el llamado *análisis colostrucciona*l –etiqueta que pretende combinar las nociones de *colocación* y de *construcción*–, parto de la hipótesis, originalmente firthiana, de que las palabras adquieren su sentido en su contexto de uso⁵. Así, entre las preguntas específicas que intentaré responder se encuentran, al menos, las siguientes: ¿Cuáles son las diferencias de uso entre las cuatro preposiciones cuando establecen la relación de tema/asunto? ¿Son sinónimas o no? y, en caso afirmativo, ¿hasta qué punto?

Antes de pasar al apartado dedicado a la metodología empleada para responder estas preguntas, cabe señalar que en consonancia con la GCx y la Lingüística cognitiva, las preguntas que acabo de plantear incluyen, en realidad, una suposición teórica importante con respecto a la respuesta. Goldberg 1995, p. 67, una de las fundadoras de la vertiente «cognitiva» de las Gramáticas de construcciones (cf. González-García 2012, pp. 250-253), formula explícitamente un Principio-de-no-sinonimia, mediante el cual niega la existencia de una sinonimia absoluta: «Si dos construcciones son sintácticamente distintas, deben ser semántica o pragmáticamente distintas»⁶. Esta idea subyace también en la noción langackeriana de «estructuración conceptual o conceptualización» (*construal*)⁷ (Langacker 1987, 1992, 2008), pues implica que cada forma lingüística, especialmente los elementos gramaticales tradicionalmente considerados desprovistos de significado, contribuye con su propia estructuración conceptual al significado contextual. Es decir, toda la metodología que se introduce en el apartado siguiente está diseñada, en principio, para descubrir diferencias entre las preposiciones, con el fin ulterior de destacar que las forma(s) no son realmente sinónimas, sino que todas ellas tienen sus propios contextos de uso. Queda por ver hasta qué punto esto se cumple en el caso concreto de las preposiciones de tema/asunto.

3. CORPUS Y MÉTODOS EMPLEADOS

A fin de dar cuenta de las posibles diferencias y semejanzas entre las preposiciones *de*, *sobre*, *acerca de* y *en torno a* con valor de tema/asunto, he recurrido al *Corpus del español* de Davies 2002, del cual he extraído 100 casos de cada preposición correspondientes al siglo XX. Como revelan los

5. El original lee: «You shall know the meaning of a word by the company it keeps» (Firth 1968, p. 179).

6. El original lee: «[i]f two constructions are syntactically distinct, they must be semantically or pragmatically distinct» (Goldberg 1995, p. 67).

7. Sigo la traducción al español de este término que ofrecen Ibarretxe-Antuñano y Valenzuela 2012 en el índice su reciente obra *Lingüística Cognitiva*, así como la terminología de Maldonado 2012, pp. 214-15 en su capítulo de la misma.

datos de la Tabla 2 (segunda línea), ha sido necesario extraer un número variable de contextos para alcanzar el número final de 100 casos. Como indican los datos de la última línea, según este recuento *de* es claramente la preposición de tema/asunto más frecuente, si bien *sobre* quizá sea un mejor candidato a considerar la PTA prototípica de la lengua actual.

	<i>de</i>	<i>sobre</i>	<i>acerca de</i>	<i>en torno a</i>	Total
usos temáticos	100	100	100	100	400
casos revisados	1334	255	100	290	1978
porcentaje	7.5 %	39.2 %	100 %	34.5 %	20.2 %
nº total de casos de cada preposición	1 163 904	33 572	782	1006	1 199 264
nº calculado de casos temáticos	87 315	13 165	782	347	242 521

Tabla 2. Datos generales de la muestra manejada.

Los 400 casos de la relación de tema/asunto fueron anotados manualmente según una serie de factores que tienen la función de caracterizar de la manera más completa posible el contexto de uso de cada una de las preposiciones. Los factores empleados así como sus diferentes subtipos se presentan en la Tabla 3. La Figura 2, por su parte, recoge ejemplos concretos de la anotación de las muestras extraídas del CE.

Como puede deducirse de la Tabla 3, se trata de 17 factores sintácticos y cinco factores semánticos en total, los cuales proponen una caracterización individual de cada uno de los usos de las PTA. Sin embargo, para poder interpretar esta categorización y evaluar qué factores realmente son distintivos, es necesario recurrir al análisis cuantitativo. En este caso me he decantado por una regresión logística multinomial, pues lo que hace este análisis es crear un modelo que describe cuál(es) de los factores incluidos puede(n) usarse para diferenciar las cuatro PTA. La ventaja de la regresión logística multinomial es que considera, simultáneamente, el peso relativo de todos los factores incluidos en el análisis, determinando su contribución relativa al modelo como un todo. Con ello revela qué factores son significativos para cada PTA, creando un modelo general de clasificación basado en los datos. La capacidad de predicción del modelo, por su parte, ilustra hasta qué punto los factores usados son capaces de predecir el uso de las PTA⁸.

8. En el siguiente apartado, dedicado al análisis de los datos, se presentará con ejemplos el funcionamiento de este análisis. Véase Glynn 2010, pp. 257-258, para una caracterización general de la regresión logística.

Factores sintácticos o formales	Factores semánticos
<i>Clase sintáctica</i> de los elementos <i>regente</i> y <i>regido</i> : N(ombre), V(erbo), A(djetivo), 0 (sin núcleo)/N, pron(ombre), V, conj(unción)	<i>Clase semántica (cs)</i> del elemento <i>regente</i> (Comunicación/proceso mental/general) y del elemento <i>regido</i> (acción, entidad concreta, entidad abstracta)
<i>Complejidad</i> de los elementos <i>regente</i> y <i>regido</i> : SN = 1-3) y N = 0	<i>Animacidad</i> (de los elementos <i>regente</i> y <i>regido</i>) = humano/no-humano/colectivo/sin especificar
<i>Número</i> de los elementos <i>regente</i> y <i>regido</i> : singular o plural	<i>Presencia</i> de otra expresión de tema/asunto (<i>de, sobre, etc.</i>) en la misma oración
<i>Definitud</i> de los elementos <i>regente</i> y <i>regido</i> : N definido o indefinido; forma activa, pasiva o impersonal de <i>regente verbal</i>	Interpretación <i>metafórica</i> o <i>concreta</i> del objeto <i>regente</i> y <i>regido</i>
<i>Determinación</i> de los elementos <i>regente</i> y <i>regido</i> : det(erminado) o indet(erminado); indicativo, subjuntivo o infinitivo de <i>regente verbal</i>	<i>Valoración</i> semántica del elemento <i>regente</i> (neutra, negativa)
<i>Modificación</i> de los elementos <i>regente</i> y <i>regido</i> : modif(icado) vs. no modif(icado)	
<i>Atributo</i> junto a los elementos <i>regente</i> y <i>regido</i> : atributo o sin atributo	
<i>Posición</i> del elemento <i>regente</i> : antes o después de la PTA	
<i>Tipo</i> de oración (declarativa, interrogativa)	
<i>Negación</i> (oración negativa, afirmativa)	

Tabla 3. Los factores usados en la anotación de los ejemplos extraídos del corpus. El subrayado hace referencia a las abreviaturas usadas en la anotación (cf. la Figura. 2).

Ag_anim	Presencia	Rgte_metaf	Rgte_cs	Rgte_lex	Rgte	Rgte_num	Rgte_def	Rgte_det	Rgte_modif	Rgte_atr	Rgte_compl	Rgte_pos	PTA
humano	0	conc	cog	conocimiento	N	sg		1	1	0	0	1 pre	de
colectivo	1	metaf	gral	informe	N	sg		0	1	0	0	1 pre	sobre
sin espec	0	conc	gral	datos	N	pl		0	1	0	0	1 pre	acerca
sin espec	1	conc	gral	investigación	N	sg		0	1	0	0	1 pre	en torn

EJEMPLO

Generalmente, el narrador decimonónico es un narrador omnisciente, es decir, su conocimiento de la realidad es completo, es total.

la Cámara de Diputados donde debe brindar un informe sobre las acciones policiales contra la marcha de productoras de coca.

Las epístolas de Pablo y Hechos ofrecen al lector algunos datos acerca de la vida de estas primitivas comunidades cristianas y sobre su relación con las culturas Menchú dijo que se ha iniciado una investigación en torno a la primera edición porque ella no ha recibido ninguna regalía.

Rdo_lex	Rdo	Rdo_num	Rdo_def	Rdo_det	Rdo_modif	Rdo_atr	Rdo_compl	Rdo_metaf	Rdo_cs	Fuente	S_tipo	Negación	Valoración
realidad	N	sg	1	1	0	0	1	metaf	abs_obj	Habla Culta: Santiago: decl		0	0
acciones	N	sg	1	1	1	1	3	metaf	act	Bolivia:ERBOL:10/31/9 decl		0	0
vida	N	sg	1	1	0	1	2	metaf	abs_obj	Enc: Biblia	decl	0	0
edición	N	sg	1	1	1	0	2	metaf	abs_obj	Guat:Gerencia:98MA? decl		1	0

Figura 2. Anotación de los ejemplos incluidos en la muestra.

Al lado del análisis de regresión logística también he empleado otro análisis cuantitativo con una finalidad más bien descriptiva. El agrupamiento jerárquico (*hierachical clustering*) es una forma muy práctica de visualizar la relación existente entre grupos de sinónimos, pues permite construir imágenes del tipo de árboles genealógicos, donde cada ramificación corresponde a una diferencia notable entre los elementos separados en ese nodo (cf. Diyjak y Gries 2006; Janda y Solovyev 2013).

4. ANÁLISIS DE LOS DATOS

Como pudo observarse anteriormente, existen obvias diferencias entre *de*, *sobre*, *acerca de* y *en torno a* en cuanto a su frecuencia de uso como PTA. Siguiendo en el plano general, pueden observarse algunas diferencias fundamentales también en los rasgos formales de los elementos regentes y regidos de la relación prepositiva. Así, las Tablas 4a y 4b incluyen las distribuciones relativas de las diferentes clases de palabras que funcionan como elemento regente (4a) y regido (4b) de las PTA.

Regente	<i>de</i>	<i>sobre</i>	<i>acerca de</i>	<i>en torno a</i>	promedio
Nombre	37 %	73 %	51 %	78 %	60 %
Verbo	54 %	27 %	42 %	20 %	36 %
Adjetivo	9 %	0 %	2 %	0 %	3 %
0	0 %	0 %	5 %	2 %	2 %
Suma	100 %	100 %	100 %	100 %	100 %

Tabla 4a. Clase sintáctica de los elementos regentes.
0 = sin elemento regente.

Regido	<i>de</i>	<i>sobre</i>	<i>acerca de</i>	<i>en torno a</i>	promedio
Nombre	61 %	91 %	80 %	87 %	79 %
Pronombre	24 %	9 %	15 %	10 %	15 %
Verbo	10 %	0 %	1 %	0 %	3 %
Conjunción	5 %	0 %	4 %	3 %	3 %
Suma	100 %	100 %	100 %	100 %	100 %

Tabla 4b. Clase sintáctica de los elementos regidos.
'Conjunción' = oración subordinada introducida por conjunción.

Como revelan las cifras de estas dos tablas, existen diferencias notables en la frecuencia con que las preposiciones se combinan con diferentes tipos de elementos regentes y regidos. Es decir, al menos en este nivel las preposiciones parecen tener preferencias distribucionales distintas.

Puede verse, por ejemplo, que *sobre* y *en torno a* se usan preferentemente después de un elemento regente nominal, mientras que *de* y *acerca de* presentan frecuencias más elevadas con regentes verbales, y son las únicas PTA que se combinan con regentes adjetivales. Por otro lado, *de* y *sobre* no aparecen como PTA en estructuras sin elemento regente, algo que sí ocurre con *acerca de* y *en torno a*⁹. Con respecto al elemento regido, se nota que *sobre* es la PTA con el uso más restringido, combinándose únicamente con sustantivos y pronombres. *De* y *acerca de*, por su parte, son las únicas que rigen verbos en infinitivo, si bien la frecuencia de *acerca de* en este contexto es baja; delante de oraciones subordinadas pueden usarse *de*, *acerca de* y *en torno a*, pero esta construcción es siempre marginal.

4.1. Análisis estadístico: la regresión logística

Ahora, si las clases sintácticas de los elementos regentes y regidos marcan cierta diferencia entre las PTA, ¿cuál es la situación cuando hacemos entrar los demás factores? ¿Qué factores favorecen el uso de una u otra PTA cuando se toman en consideración conjuntamente? Para responder a estas preguntas, un procedimiento práctico es el uso de la regresión logística, pero ¿en qué consiste la construcción de un modelo matemático sobre la interacción de todos los factores incluidos en la anotación?

Tomándose como punto de comparación la distribución de las clases sintácticas (nombre, pronombre, verbo, adjetivo, conjunción y 0) por las cuatro PTA que se presentaron en las Tablas 4a y 4b, lo que hace la regresión logística es tomar la idea de coocurrencia a otro nivel, considerando no solo un factor, como la clase del elemento núcleo, sino todo el grupo de factores concurrentes para ver de qué manera se relacionan con cada una de las preposiciones. Así, el análisis de regresión logística es una comparación matemáticamente sólida de 22 factores (17 sintácticos y cinco

9. Este parece ser uno de los casos en los que se ha sentido la necesidad de sustituir la preposición *de* por expresiones más explícitas, pues en épocas anteriores *de* se usaba frecuentemente en títulos de capítulos o de trabajos expositivos, como ocurre en «Capítulo III. De lo que sucedió a nuestro caullero cuando salio de la venta», del *Quijote*. En la actualidad este uso de *de* es marginal, mientras abundan los trabajos académicos con títulos introducidos por *sobre*, *acerca de* y *en torno a*.

Cabe señalar en este punto que, en este trabajo, considero que los usos temáticos que carecen de un elemento regente explícito, como ocurre en el ejemplo del *Quijote*, siguen constituyendo una relación preposicional, pese a su anomalía formal. Esta decisión puede motivarse por dos factores: en primer lugar, muchas veces es natural suponer que existe un elemento regente implícito, elidido, cuya expresión explícita resultaría superflua. Este parece ser el caso del ejemplo del *Quijote*, donde cabe intuir una lectura del tipo: «Capítulo III. [Que trata] de los que sucedió...». En segundo lugar, en muchos otros casos, cuando la preposición de tema/asunto se encuentra en posición inicial, el elemento regente puede encontrarse más adelante en la oración, donde puede ser más o menos explícito. Este es el caso del ejemplo: «En torno a la religión, los griegos también tenían cuatro festivales nacionales, llamados juegos...» (CE, Enc: Grecia). Aquí todo el enunciado «los griegos también tenían festivales nacionales» puede considerarse elemento regente.

semánticos) en relación con las PTA. En lugar de hacer cada análisis por separado, en la regresión logística se consideran los 22 factores, que suman más de 30 valores diferentes, al mismo tiempo y relacionando el uno con el otro. El grado en el que un factor dado determina el uso de una u otra PTA se calcula tomando en consideración todos los demás factores simultáneamente.

Los resultados que produce un análisis de regresión logística es, poco sorprendentemente, una serie de valores numéricos presentados en diversas tablas (Tablas 5, 6 y 7). Empezando por la primera (Tabla 5), lo que esta indica es que los factores incluidos en la columna «Efecto» son los que contribuyen significativamente a la creación del modelo. Como es frecuente en los análisis estadísticos, lo determinante para considerar que un dato sea significativo es su valor de ji cuadrada (χ^2) y que el valor de probabilidad (valor p) correspondiente sea inferior a 0.05. En este caso, los valores de probabilidad se encuentran en la columna más a la derecha de la Tabla 5. En la segunda línea, se incluye el valor del modelo en total (Modelo final), mientras que los valores subsiguientes corresponden a los factores individuales. Entre los factores significativos se encuentran siete sintácticos y cuatro semánticos. Pero, como puede verse, tres de ellos corresponden a subtipos del factor «Definitud» del elemento regente. Además, si se siguiera estrictamente el valor p, con un valor de 0.16 el factor «regente_def_indef» (elemento regente indefinido) quizá no debería incluirse, pero, como podrá observarse más adelante, este valor contribuye significativamente a diferenciar *de* de *sobre* y *en torno a*.

Efecto	Criterios de ajuste del modelo	Test de razón de verosimilitud (Likelihood Ratio Test)			
		verosimilitud del -2 log	ji cuadrada	Grados de libertad	Valor p
Solo Intercept	803.035				
Modelo final	599.172	203.862	33	0.000	
Modelo reducido					
Compl_regente	613.306	14.134	3	0.003	
Compl_regido	629.490	30.318	3	0.000	
Regente_N	622.554	23.382	3	0.000	
Regido_V	612.063	12.891	3	0.005	
Ag_anim_humano	622.978	23.805	3	0.000	
Regente_def_indef	604.395	5.222	3	0.156	
Regente_def_def	607.878	8.705	3	0.033	
Regente_def_activa	612.337	13.165	3	0.004	
Valoración	610.139	10.966	3	0.012	
Regente_cs_com	607.846	8.674	3	0.034	
Regido_cs_objetos	612.620	13.448	3	0.004	

Tabla 5. Información sobre el ajuste del modelo y el test de razón de verosimilitud del modelo de regresión logística multinomial.

La Tabla 6, por su parte, entra en un nivel más detallado y otorga datos relativos a la importancia específica de los distintos factores para diferenciar entre las PTA. Las cifras más importantes de la Tabla 6 son las que aparecen en las columnas «Valor p» y «Exp(B)». Como es habitual, un valor p por debajo del 0.05 es indicio de una importancia significativa, mientras que el valor del exponente B indica la fuerza y la direccionalidad de la significancia.

	PTA	B	Error estándar	Wald	GdL	Valor p	Exp(B)
<i>sobre</i>	Intercept	-2.328	0.717	10.541	1	0.001	
	Compl_regente	0.618	0.273	5.148	1	0.023	1.856
	Compl_regido	0.774	0.213	13.260	1	0.000	2.169
	Regente_N	18.583	0.751	612.909	1	0.000	117 646 637
	Regido_V	-17.480	2 848	.000	1	0.995	2.563E-8
	Ag_anim_humano	-1.209	0.386	9.801	1	0.002	0.298
	Regente_def_indef	-16.225	0.497	1063.824	1	0.000	8.989E-8
	Regente_def_def	-16.823	0.000	.	1	.	4.940E-8
	Regente_def_activa	1.834	0.694	6.988	1	0.008	6.257
	Valoración	0.198	0.706	.079	1	0.779	1.219
	Regente_cs_com	0.044	0.397	.012	1	0.912	1.045
	Regido_cs_objetos	-0.760	0.406	3.508	1	0.061	0.467
<i>acerca de</i>	Intercept	-1.990	0.618	10.362	1	0.001	
	Compl_regente	0.949	0.274	11.993	1	0.001	2.583
	Compl_regido	1.005	0.221	20.781	1	0.000	2.733
	Regente_N	1.974	0.881	5.019	1	0.025	7.199
	Regido_V	-2.685	1.195	5.047	1	0.025	0.068
	Ag_anim_humano	-0.787	0.386	4.158	1	0.041	0.455
	Regente_def_indef	-1.143	0.985	1.347	1	0.246	0.319
	Regente_def_def	-2.269	1.063	4.557	1	0.033	0.103
	Regente_def_activa	1.011	0.568	3.174	1	0.075	2.748
	Valoración	1.277	0.651	3.848	1	0.050	3.588
	Regente_cs_com	-0.112	0.389	.084	1	0.772	0.894
	Regido_cs_objetos	-1.091	0.436	6.271	1	0.012	0.336
<i>en torno a</i>	Intercept	-2.763	0.830	11.088	1	0.001	
	Compl_regente	0.864	0.292	8.724	1	0.003	2.372
	Compl_regido	1.124	0.239	22.125	1	0.000	3.078
	Regente_N	18.343	0.839	477.439	1	0.000	92 542 844
	Regido_V	-17.570	2822	.000	1	0.995	2.341E-8
	Ag_anim_humano	-1.917	0.423	20.519	1	0.000	0.147
	Regente_def_indef	-16.315	0.524	969.486	1	0.000	8.214E-8
	Regente_def_def	-16.642	0.000	.	1	.	5.925E-8
	Regente_def_activa	2.373	0.803	8.728	1	0.003	10.731
	Valoración	1.462	0.675	4.688	1	0.030	4.314
	Regente_cs_com	-1.080	0.468	5.336	1	0.021	0.340
	Regido_cs_objetos	-1.780	0.547	10.575	1	0.001	0.169

Tabla 6. Valores estimados de los parámetros con respecto a la significancia y contribución relativas de los once factores significativos del análisis de regresión logística multinomial. El grado de significancia de cada factor se ha calculado teniendo como punto de comparación la preposición *de*. Los factores que presentan valores p significativos están marcados con negrita.

Por ejemplo, visualizando la línea de «regente_N» en el bloque de *sobre*, el valor de 117 646 637 indica que la probabilidad de usar *sobre* es tanto más alta que la de usar *de* cuando el regente es nominal. El valor de 7.199 para este factor en el caso de *acerca de*, indica la misma tendencia pero en un grado mucho más moderado: la probabilidad de usar *acerca de* sobre *de* es solo 7:1 con elementos regentes nominales. Continuando con *acerca de*, en la línea siguiente, «regido_V», que corresponde a elementos regidos infinitivos (= verbo), la cifra de 0.068, por su parte, indica que la probabilidad de usar *de* es más elevada cuando la PTA rige un infinitivo. Alternativamente, invirtiendo este valor el resultado es de un 14.7, que corresponde a la probabilidad de usar *acerca de* sobre *de* cuando el elemento regido no es un infinitivo. Es decir, la direccionalidad de la significancia depende de si el valor de «Exp(B)» es superior o inferior a 1. Valores superiores a 1 indican una mayor probabilidad de usar la PTA en cuestión con el factor indicado, mientras que valores entre 0 y 1 indican una mayor probabilidad de uso de *de*.

Casos observados	Casos esperados				Porcentaje correcto
	<i>de</i>	<i>sobre</i>	<i>acerca de</i>	<i>en torno a</i>	
<i>de</i>	64	15	19	2	64.0 %
<i>sobre</i>	16	28	23	33	28.0 %
<i>acerca de</i>	19	21	27	28	28.4 %
<i>en torno a</i>	4	22	10	62	63.3 %
Porcentaje total	26.2 %	21.9 %	20.1 %	31.8 %	46.1 %

Tabla 7. Capacidad de clasificación del modelo estadístico.

Sobre la base de estos datos cuantitativos es, en principio, posible caracterizar cada una de las preposiciones de tema/asunto y establecer en qué aspectos difieren unas de otras. Sin embargo, la regresión logística incluye una tabla más, donde lo que hace es autoevaluarse. Es decir, siguiendo los efectos de los factores significativos incluidos en el modelo, el análisis establece una comparación entre la distribución analizada y una situación hipotética creada por el modelo. Esta comparación se presenta en la Tabla 7, y, como puede verse, los resultados dejan mucho que desear. Es decir, el porcentaje de clasificación de que es capaz el modelo no alcanza sino el 46.1 por ciento, lo que significa que, usando los criterios de clasificación que otorga el análisis, se clasifica mal más de la mitad de los casos. Más concretamente, viendo las cifras de la Tabla 7, lo que parece ocurrir es que *de* y *en torno a* constituyen dos polos en cierto modo opuestos, y estos dos polos sí los logra identificar el modelo en un grado aceptable. *Sobre* y *acerca de*, en cambio, se clasifican más frecuentemente como *en torno a* que como sí mismas. Las clasificaciones cruzadas de *sobre* y *acerca de* también son demasiado frecuentes. Todo esto significa que los factores identificados no pueden considerarse tan buenos predictores del uso de una

u otra preposición como habría sido deseable. Por este motivo he decidido no entrar en la interpretación detallada de los resultados en este punto.

4.2. *Más análisis cuantitativos*

¿Qué hacer cuando los resultados del análisis no son tan conclusivos como habría sido deseable? La solución más evidente sería intentar encontrar fallos en el análisis y posibles faltas en la anotación. Sin embargo, dado que no parece que el análisis presentado contenga errores y/o carencias enmendables, en este caso he decidido avanzar por otra vía. Es decir, para poder aprovechar al máximo los datos que están al alcance, he optado simplemente por aceptar que el modelo creado por la regresión logística es incompleto, lo que equivale a decir que la realidad lingüística elude el análisis. Así, en primer lugar he sustituido la variante multinomial por versiones binomiales de la regresión logística, con el fin de contrastar las preposiciones por pares. Si bien hay que, recordar que al abandonar el análisis paralelo de las cuatro PTA, también se pierde la perspectiva comparativa global, es de esperar que, al proseguir por pares, los resultados adquiridos arrojen alguna luz adicional sobre las diferencias entre aquellas preposiciones que menos era capaz de diferenciar el modelo inicial. Los resultados resumidos de dos rondas separadas de análisis de regresión logística binomial se presentan en la Tabla 8.

	<i>Ronda 1: Los once factores significativos en el análisis multinomial</i>		<i>Ronda 2: Todos los factores anotados</i>	
Par de PTA	Porcentaje de clasificación correcta	Factores significativos	Porcentaje de clasificación correcta	Factores significativos
<i>de ~ sobre</i>	77.5 %	Compl_regido Regente_N (Regido_V) Ag_anim_humano Regente_def_activa	78.0 %	Regente_modif Compl_regido Regente_N Regido_N Ag_anim_0 Ag_anim_humano Regente_def_activa (Regente_def_pasiva)
<i>de ~ acerca de</i>	71.8 %	Compl_regente Connotaciones Compl_regido Regente_N Regido_V Regente_def_def	75.6 %	Regente_modif Connotaciones Regido_num Regido_def Regido_attr Regido_fig Ag_anim_humano

<i>de ~ en torno a</i>	87.4 %	Regente_compl Regido_compl Regido_wc_objetos Regente_N Ag_anim_humano Regente_def_activa	87.3 %	Regente_modif Regido_compl Regido_wc_objetos Regente_N Ag_anim_humano Regente_def_activa (Regente_def_pasiva)
<i>sobre ~ acerca de</i>	61.0 %	Connotaciones Regente_N	63.7 %	Regente_compl Regido_modif Regido_attr Regente_N
<i>sobre ~ en torno a</i>	62.6 %	Regente_wc_com Regido_wc_objetos Connotaciones	66.0 %	Regido_def Regido_det Regido_N Ag_anim_humano
<i>acerca de ~ en torno a</i>	68.4 %	Regente_wc_com Ag_anim_humano Regente_def_def	72.6 %	Regente_posición Regido_num(1) Regido_attr Regente_wc_com Ag_anim_humano Regente_def_def Regente_def_impersonal

Tabla 8. Capacidad clasificatoria y factores significativos de dos rondas sucesivas de regresión logística binomial.

Al observar las cifras de la Tabla 8, lo primero que destaca es que sigue habiendo una clara diferencia entre unos pares y otros. Así, en las tres oposiciones en las que está presente la preposición *de*, el porcentaje de clasificación excede el 70%; en la ronda 2, que se basa en todos los factores anotados, la cuota supera incluso el 75%. Todo ello parece corroborar el resultado anterior, es decir, destaca la diferencia entre *de* y las demás PTA analizadas. Por otro lado, los análisis binomiales también logran diferenciar claramente las expresiones *acerca de* y *en torno a*, con un porcentaje por encima del 72% en la ronda 2. En cambio, las diferencias entre *sobre* y *acerca de* y *en torno a*, respectivamente, resultan más difíciles de establecer de forma definitiva. Más adelante, en la sección 5, habrá ocasión de intentar ofrecer una explicación de esta dificultad.

Otra observación importante que puede extraerse de los datos de la Tabla 8 es que, aunque hay una leve mejora en la capacidad de clasificación entre las dos rondas, la diferencia no es significativa. Esto quiere decir que no parece sensato prestar demasiada atención al hecho de que los factores significativos en una y otra mitad de la tabla sean parcialmente diferentes. Tomados en conjunto, su capacidad predictiva de la elección de una u otra PTA en un contexto dado es prácticamente la misma. Sin embargo,

examinando las dos columnas de factores, resulta lícito considerar que aquellos factores que aparecen en ambos lados pueden considerarse más significativos que los demás, motivo por el cual aparecen marcados con negrita.

Un método cuantitativo adicional aplicable a una muestra de casos anotados es el agrupamiento jerárquico (*hierarchical clustering*), que crea una matriz de distancia o similitud entre los elementos comparados (cf. Gries 2009, pp. 306 y ss.). La aplicación de este análisis sobre la muestra de las PTA da como resultado el árbol jerárquico que se presenta en la Figura 3.

Leyendo la Figura 3 de arriba abajo, el agrupamiento jerárquico separa claramente la preposición *de* de las demás preposiciones, lo cual está en perfecta consonancia con los resultados de los análisis anteriores. En el siguiente nodo, la PTA que se separa de las restantes es *acerca de*, lo que también resulta conforme con las cifras presentadas en la Tabla 8, donde *acerca de* tiene la cifra más reducida de separación de *de* comparado con *sobre* y *en torno a*. El último, y único verdadero grupo que incluye el dendrograma de la Figura 3 lo constituyen *sobre* y *en torno a*, lo cual puede considerarse un tanto extraño ya que fueron *sobre* y *acerca de* las que menos se distinguían en el modelo de regresión logística inicial (la multinomial); a su vez, el porcentaje de clasificación correcta de *sobre* en relación con *acerca de* es más bajo que en relación con *en torno a*. Así, la agrupación que establece el último nodo del dendrograma parece obedecer más que nada a la gran distancia que separa *de* de *en torno a*, por una parte, y a la relativa similitud entre *sobre* y *acerca de*, por un lado, y entre *de* y *acerca de*, por otro.

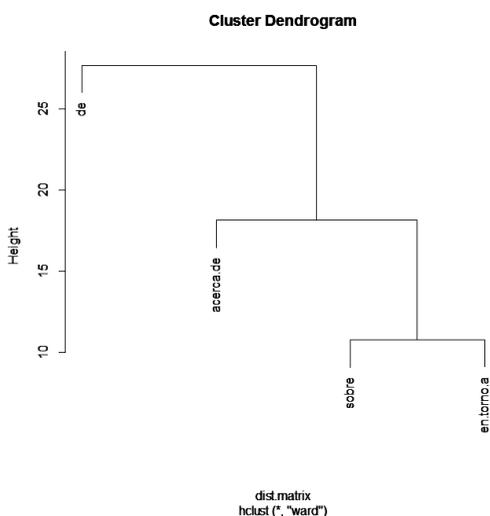


Figura 3. Dendrograma de agrupación jerárquica de las preposiciones *de*, *sobre*, *acerca de* y *en torno a*.

5. DISCUSIÓN DE LOS RESULTADOS

En este apartado se resumen los resultados más importantes que arroja el análisis, con énfasis especial en la cuestión de qué factores contribuyen significativamente al uso de cada preposición de tema/asunto. Los resultados se encuentran reunidos en las Tablas 6 y 8, que incluyen los factores significativos. Habiendo desaprovechado la oportunidad de presentar la significancia de los factores incluidos en el modelo en el apartado de análisis (Sección 4.1) por preferir explorar análisis alternativos, en este apartado voy a explorar los datos del análisis hasta donde me parece lícito, a fin de construir una caracterización lo más completa posible de cada una de las PTA. En la práctica, esto puede hacerse de dos formas distintas: por un lado, pueden comentarse uno por uno todos los factores significativos de la regresión logística haciendo hincapié en qué preposiciones son diferenciadas por cada factor (cf. Granvik y Taimitarha 2014). Por otro lado, pueden usarse los factores para establecer el contexto de uso típico de cada preposición. Dado que la capacidad diferenciadora del modelo regresivo es relativamente baja, no parece tener mucho sentido explorar la importancia individual de los factores a la hora de marcar una diferencia entre una y otra PTA, motivo por el cual he decidido aprovechar la segunda opción.

Ahora bien, aunque he optado por la segunda alternativa en este trabajo, permítaseme iniciar este apartado con el comentario detallado de uno de los factores que parece ser de importancia fundamental para diferenciar todas las preposiciones, a saber, que el elemento regente tenga un agente animado humano. La importancia de este factor, *ag_anim_humano* en las Tablas 6 y 8, distingue la preposición *de* de todas las demás preposiciones con un amplio margen. Según los datos de la Tabla 6, la ratio de probabilidades es de 0.298 para la pareja *de ~ sobre*; 0.455 para la pareja *de ~ acerca de* y 0.147 para la pareja *de ~ en torno a*. Lo que implican estas cifras es que la probabilidad de usar *de* frente a *sobre* cuando el agente es humano es de más de 3:1; la razón de probabilidad es de 2:1 a favor de *de* en relación con *acerca de*, y con respecto a *en torno a* la probabilidad es de casi 7:1 a favor de *de*. Todo ello se traduce, pues, en que *de* favorece los contextos donde se relaciona un agente humano con el elemento regente, lo cual puede verse en la Tabla 9 y la Figura 4.

Preposición de tema/asunto		<i>de</i>	<i>sobre</i>	<i>acerca de</i>	<i>en torno a</i>	Total
Agente humano	Nº	66	40	51	21	178
	Frecuencia	37.1 %	22.5 %	28.7 %	11.8 %	100 %
Agente no humano	Nº	34	60	49	79	222
	Frecuencia	15.3 %	27.0 %	22.1 %	35.6 %	100 %
Nº totales		100	100	100	100	400

Tabla 9. Distribución de Agentes humanos y no humanos por las cuatro PTA.
 $P < 0.001$, $\chi^2 = 43.61$ (3 GdL).

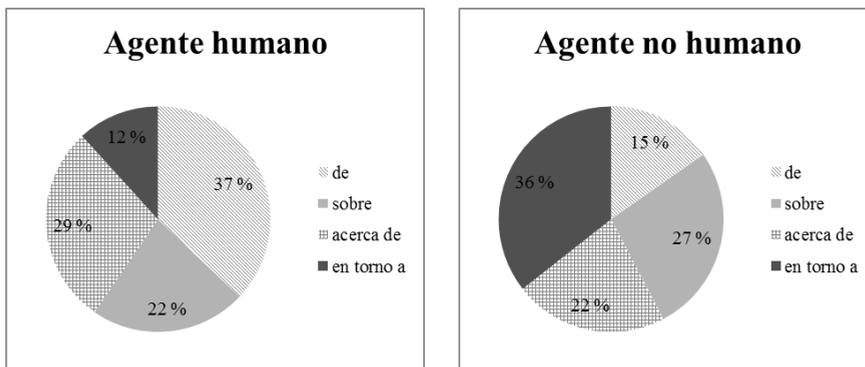


Figura 4. Visualización gráfica de la distribución de las cuatro PTA por agentes humanos y no humanos.

<i>de</i>	<i>sobre</i>	<i>acerca de</i>	<i>en torno a</i>
Complejidad del elemento regente = 0	Complejidad del elemento regente = 1	Complejidad del elemento regido = 2, 3	Complejidad del elemento regente = 2, 3
Complejidad del elemento regido = 0	Complejidad del elemento regido = 3	Valoración negativa (Elemento regido no objetos)	Complejidad del elemento regido = 2, 3
Regente no-nominal	Regente nominal		Regente nominal
Regido verbal (= infinitivo)	(Elemento regente no indefinido)		Elemento regente indefinido
Agente humano	(Agente animado no humano)		Valoración negativa (Agente animado no humano)
Objetos como elemento regido			(No voz activa en el elemento regente)
			(Elemento regente no de comunicación)
			(Elemento regido no objeto)

Tabla 10. Rasgos típicos de las cuatro preposiciones de tema/asunto¹⁰.

Volviendo la atención a la caracterización típica de los contextos de uso de las cuatro PTA, cabe empezar por constatar que el agente humano será, evidentemente, un rasgo contextual importante de la preposición *de*. Los

10. Los rasgos colocados entre paréntesis corresponden a valores negativos de los factores/subtipos en cuestión, con lo cual, en realidad, hacen referencia a todos los demás subtipos. Por eso, no pueden interpretarse de manera tan directa como los valores positivos.

otros rasgos distintivos de *de* y de las demás preposiciones se presentan en la Tabla 10.

Aunque en teoría sería posible construir, con estos factores, ejemplos que correspondan por completo a estas caracterizaciones, he optado por extraer de la muestra dos ejemplos concretos de cada preposición en los que se actualiza el mayor número posible de factores. Así, los ejemplos (12) a (15) constituyen una muestra de los casos del corpus más típicos de cada PTA analizada.

- (12) a. Si *se trata de* hablar de autores que más me llegan –en esto también posiblemente soy un poco anacrónico–, el que más me llega entre todos es Lawrence Durrell y Albert Camus. (CE, Habla Culta: Santiago: M18)
 b. Se postula no más. Entonces resulta que si uno *habla de* lo sicógeno, bueno, lo sicógeno viene de lo síquico, ¿no es cierto? (CE, Habla Culta: Santiago, M17)
- (13) a. Estos trabajos revolucionan actualmente *los conocimientos sobre* los cambios evolutivos y las relaciones filogenéticas entre las especies. (CE, España:, ABC)
 b. Se pensaba que era posible aplicar, «de golpe», *teorías generales y abstractas sobre* la mejor forma de gobierno. (CE, España, ABC)
- (14) a. Consultado el alcalde cruceño Johnny Fernández, se mostró primero sorprendido por la publicación y aseveró luego que la misma «corrobora *las dudas de la familia*» *acerca de* las verdaderas causas del desastre en Uncía. (CE, Bolivia, ERBOL, 01/31/96)
 b. En su reunión con Bill Clinton, el presidente Leonel Fernández tuvo la oportunidad de hacerle llegar a Bill Clinton *las preocupaciones acerca de* la deuda bilateral (de 700 millones de dólares), la paridad textil y el creciente incremento de los deportados dominicanos desde Estados Unidos. (CE, DR, Listin, 98Jun11)
- (15) a. En tanto *la indefinición judicial en torno a* la cuestión de competencia planteada en Buenos Aires por el juez federal Ricardo Weschler no estuvo al juez Burad, reveló que varios organismos oficiales de esta ciudad no respondieron a sus pedidos de informes. (CE, Arg:Prensa, 67_OTNA)
 b. Por otra parte, *el pronunciamiento general* que escuchamos aquí *en torno a* la cuestionada economía neoliberal, esa economía de mercado, ¿es posible que se vuelva a esa economía? (CE, Entrevista (PRI))

Lo que revelan los factores listados en la Tabla 10, que caracterizan el contexto de uso prototípico de cada PTA, es que parece posible establecer una cierta diferencia entre estas preposiciones. Sin embargo, estas diferencias deben verse como tendencias distribucionales y colocacionales más que como una distinción semántica entre unas y otras. Por otro lado, cabe volver a insistir en que unas se diferencian de las demás más que las otras. Así, *de*, por ejemplo, tiene seis rasgos diferenciadores, y *en torno a*, cinco;

por su parte, *sobre* y *acerca de* solo tienen dos factores caracterizadores propiamente dichos, así como algunos rasgos contextuales desfavorecidos.

Finalmente, mirando con atención los factores distintivos presentados en la Tabla 10, es posible encontrar una explicación a por qué los modelos matemáticos no son capaces de diferenciar nítidamente *sobre* de *acerca de* y *en torno a*. Por ejemplo, de los tres factores distintivos (positivos) de *sobre*, *Complejidad del elemento regente* = 1, *Complejidad del elemento regido* = 2 y 3 y *Regente nominal*, dos son compartidos por *en torno a*, a saber la complejidad del elemento regido y el regente nominal. También *acerca de* se caracteriza por preferir elementos regidos complejos. *Acerca de*, además, comparte tanto la *Valoración negativa* como la *Complejidad del elemento regido* con *en torno a*. Todo esto implica que al caracterizar el contexto «prototípico» de estas PTA siguiendo estos factores, el resultado acabará siendo el de contextos oracionales muy semejantes.

Ello puede observarse también en los ejemplos (12) a (15). Así, la gran semejanza entre *sobre*, *acerca de* y *en torno a* es clara en los ejemplos (13), (14) y (15), donde parece posible la sustitución de *sobre* por *acerca de* o *en torno a* y viceversa. Por otro lado, parece intuitivamente más aceptable sustituir *sobre* y *acerca de* por *de* en los ejemplos (13) y (14), que introducir esta preposición en (15a o b), lo cual está en perfecta consonancia con el dendrograma de la Figura 3. Además, viendo los ejemplos (12a y b) resulta bastante evidente que *de* es la preposición que más claramente se diferencia de las demás, pues no parece intuitivamente muy natural sustituirla por otra. Las demás, en cambio, resultan sustituibles entre sí en un grado considerable.

Queda por ver si en trabajos posteriores será posible extender estas interpretaciones, y determinar si los contextos de uso de *sobre*, *acerca de* y *en torno a* realmente son tan semejantes que las PTA resultan intercambiables, o si, por el contrario, un análisis más amplio y profundo logra precisar con mayor detalle las diferencias en el contexto de uso.

6. CONCLUSIONES

Tras examinar y anotar 400 ejemplos prepositivos extraídos de un amplio corpus digital, y empleando una serie de métodos cuantitativos para analizar los datos, ¿qué puede decirse acerca de la relación de sinonimia existente entre las preposiciones *de*, *sobre*, *acerca de* y *en torno a*, cuando estas se usan para establecer una relación de tema/asunto? ¿Son sinónimos o no? ¿Sus respectivos contextos de uso permiten establecer diferencias entre ellas? Empezaré por responder a las últimas dos preguntas con un sí, que, sin embargo, necesita algunas matizaciones. Con esto, mi esperanza es que esta discusión llegue, de algún modo al menos, a constituir una respuesta también a la primera.

En primer lugar, las preposiciones investigadas pueden considerarse sinónimas desde el momento en que se constata que se usan para expresar el mismo valor de tema/asunto (cf. Salvador 1985). Sin embargo, pese a esto, el plantearse siquiera la tercera pregunta implica en realidad que se (pre) supone que no son sinónimos completos. Así, la respuesta afirmativa a la tercera pregunta, en principio, confirma esta sospecha: hay diferencias en el uso contextual de las cuatro preposiciones, lo cual significa que no son completamente sinónimas. Sin embargo, creo que los resultados obtenidos permiten constatar que hay algo más que meros síes y noes. La realidad semántica tal y como se ve desde la óptica de las PTA analizadas no puede reducirse a respuestas de este tipo, sino que es bastante más compleja.

Con los resultados del análisis a mano, es obvio que *de*, *sobre*, *acerca de* y *en torno a* no constituyen un grupo homogéneo de miembros mutuamente sustituibles en todos los casos y contextos. Especialmente *de* se sitúa en el margen de esta macrocategoría de preposiciones de tema/asunto. Pero los miembros restantes son muy semejantes y muchas veces sustituibles, como revela la incapacidad del modelo de regresión logística de diferenciarlas nítidamente (cf. Tabla 7) y la coincidencia en los rasgos/factores distintivos (cf. Tabla 10). Solo eliminando la perspectiva global y concentrándose en la oposición entre dos PTA en concreto, se revelan diferencias más nítidas, como el hecho de que, en la pareja *sobre*, ~ *acerca de*, la primera prefiera regentes nominales, mientras que *acerca de* se relaciona más estrechamente con regentes con una valoración negativa. Entre *sobre* y *en torno a*, esta prefiere contextos con una valoración negativa, mientras que aquella se combina con regentes de comunicación y regidos que pueden caracterizarse como objetos (cf. Tablas 8 y 10).

Cabe recordar que, al tratarse de toda una macrocategoría, en la que todos los integrantes se usan para hacer referencia al mismo valor semántico, hablar de oposición bimembre no es muy natural, ya que no hay ningún motivo para suponer que un hablante no tenga acceso continuo a la totalidad de elementos léxicos pertenecientes a la categoría. Desde esta perspectiva creo que los resultados alcanzados en este trabajo, aunque no determinantes en algunos aspectos, permiten concluir que las PTA constituyen un buen caso de expresiones sinónimas que son, en un grado considerable, sustituibles.

Dicho esto, cabe subrayar que no creo que los resultados alcanzados ofrezcan una explicación definitiva de la relación entre *de*, *sobre*, *acerca de* y *en torno a*, sino que han de verse más bien como un paso inicial. En primer lugar, al concentrarse el análisis en los factores contextuales y, así, limitarse al nivel estrictamente gramatical, siempre existe la posibilidad de extender el análisis para incluir otro tipo de factores, sociolingüísticos, estilísticos, como géneros textuales y tradiciones discursivas, etc.

En segundo lugar, dentro del nivel gramatical también existen varias vías por donde ampliar el análisis. Una alternativa sería aprovechar los grandes

corpora para ver si las PTA tienen colocaciones léxicas específicas, algo que podría intentarse aplicando los métodos del análisis colostrucciona (cf. Stefanowitsch y Gries 2003; Gries y Stefanowitsch 2004; Granvik y Taimitarha 2014). Otra alternativa sería recurrir al uso de cuestionarios y tests empíricos destinados a averiguar hasta qué punto los diferentes factores gramaticales realmente afectan a un hablante a la hora de escoger una PTA en un contexto dado¹¹. Habiendo establecido los contextos prototípicos de las preposiciones, sería muy interesante intentar averiguar hasta qué punto estos afectan a la elección de preposición. ¿Hay mayor coincidencia entre un ejemplo de corpus y la elección de hablantes nativos en un contexto prototípico que en uno menos típico? ¿Los hablantes son más rápidos en escoger la preposición que quieren usar cuando el contexto es típico? Responder a preguntas como estas implicaría un tratamiento serio de la cuestión de la sustituibilidad de sinónimos, un aspecto al que solo he aludido superficialmente en este trabajo, pero que constituye una cuestión teórica muy importante del estudio de la sinonimia.

REFERENCIAS BIBLIOGRÁFICAS

- ARPE, A. y JÄRVIKIVI, J. (2007): «Every method counts: Combining corpus-based and experimental evidence in the study of synonymy», *Corpus Linguistics and Linguistic Theory* 3, 2, pp. 131-159.
- BASSOLS DE CLIMENT, M. (1967): *Sintaxis latina*, 2 vols, Madrid, CSIC.
- BOAS, H.C. y GONZÁLEZ-GARCÍA, F. (2014): *Romance perspectives on construction grammar*, Ámsterdam/Filadelfia, John Benjamins.
- CIFUENTES HONRUBIA, J.L. (2003): *Locuciones prepositivas. Sobre la gramaticalización preposicional en español*, Alicante, Universidad de Alicante.
- CROFT, W. (2001): *Radical construction grammar. Syntactic theory in typological perspective*, Oxford, Oxford University Press.
- DAVIES, M. (2002): *Corpus del español. 100 million words. 1200s-1900s*. En línea: <<http://www.corpusdelespanol.org>>.
- DIVJAK, D. y GRIES S.Th. (2006): «Ways of trying in Russian: Clustering behavioral profiles», *Corpus Linguistics and Linguistic Theory* 2, 1, pp. 23-60.
- GLYNN, D. (2010): «Testing the hypotheses: Objectivity and verification in usage-based cognitive semantics», en Glynn, D. y Fisher, K. (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, Berlín/Nueva York, De Gruyter Mouton, pp. 239-269.
- GOLDBERG, A. (1995): *Constructions: A construction grammar approach to argument structure*, Chicago (IL), The University of Chicago Press.

11. Los estudios de Arppe y Järvi­kivi 2007 y de Vanhatalo 2002, 2003 contrastan datos de corpus con datos recogidos mediante el uso de cuestionarios, destacando la importancia del uso de métodos diferentes en el estudio de sinonimia léxica. Granvik y Taimitarha (en elaboración) supone un paso en esta dirección.

- GONZÁLEZ-GARCÍA, F. (2012): «La(s) gramática(s) de construcciones», en Ibarretxe-Antuñano, I. y Valenzuela, J. (dirs.), *Lingüística cognitiva*, Barcelona, Anthropos, pp. 249-280.
- GONZÁLEZ MARTÍNEZ, J.M. (1988-1989): «La sinonimia. Problema metalingüístico», *Anales de Filología Hispánica* 4, pp. 193-210.
- GRANVIK, A. (2012): *De de. Estudio histórico-comparativo de los usos y la semántica de la preposición de en español*, Mémoires de la Société Néophilologique de Helsinki 84, Helsinki, Société Néophilologique.
- , (2014): «Hablando *de, sobre y acerca* de la gramaticalización y la lexicalización: panorama diacrónico de las relaciones entre preposiciones y locuciones prepositivas dentro del campo semántico de tema/asunto», en Girón Alconchel, J.L. y Sáez Rivera, D.M. (eds.), *Procesos de gramaticalización en la historia del español*, Madrid/Fránkfort del Meno, Iberoamericana/Vervuert, pp. 77-117.
- , y TAIMITARHA S. (2014): «Topic-marking prepositions in Swedish: A corpus-based analysis of adpositional synonymy», *Nordic Journal of Linguistics*, 37, 2, pp. 257-296.
- , (en elaboración): «Does true synonymy exist? Testing the mutual interchangeability of near-synonyms by means of corpus and experimental data», leída en el 13th International Cognitive Linguistics Conference, Universidad de Northumbria, Inglaterra, 20-25 de julio de 2015.
- GRIES, S.Th. (2009): *Statistics for linguistics with R. A practical introduction*, Berlín, De Gruyter Mouton.
- , y STEFANOWITSC, A. (2004): «Extending collocation analysis: A corpus-based perspective on ‘alternations’», *International Journal of Corpus Linguistics* 9, 1, pp. 97-129.
- HALLIDAY, M.A.K. (1967): «Notes on transitivity and theme in English. Part 2», *Journal of Linguistics* 3, pp. 199-244.
- IBARRETXE-ANTUÑANO, I y VALENZUELA, J. (dirs.) (2012): *Lingüística cognitiva*, Barcelona, Anthropos.
- JANDA, L. y SOLOVYEV, V. (2013): «What constructional profiles reveal about synonymy: A case study of Russian words for SADNESS and HAPPINESS», en Janda, L.A. (ed.), *Cognitive linguistics: The quantitative turn. The essential reader*, Berlín/Nueva York, De Gruyter Mouton, pp. 295-324.
- LANGACKER, R.W. (1987): *Foundations of cognitive grammar, vol. 1: Theoretical prerequisites*, Stanford (CA), Stanford University Press.
- , (1992): «Prepositions as grammatical(izing) elements», *Leuvense Bijdragen* 81, pp. 287-309.
- , (2008): *Cognitive grammar: A basic introduction*, Oxford, Oxford University Press.
- MALDONADO, R. (2012): «La Gramática cognitiva», en Ibarretxe-Antuñano, I. y Valenzuela, J. (dirs.), *Lingüística cognitiva*, Barcelona, Anthropos, pp. 213-247.
- NGLE = REAL ACADEMIA ESPAÑOLA Y AALE (2009): *Nueva gramática de la Lengua Española*, 2 vols, Madrid, Espasa.
- PONS RODRÍGUEZ, L. (2001-2002): «La historia de *en cuanto a* como tematizador», *Anuari de Filologia* 11-12, pp. 73-94.
- , (2005): «Los tematizadores en los manuales de ELE», *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y*

- realidad. Actas del XV Congreso Internacional de ASELE*, Sevilla, Universidad de Sevilla, pp. 683-692.
- SALVADOR, G. (1985): «Sí hay sinónimos», en *Semántica y lexicología del español. Estudios y lecciones*, Madrid, Paraninfo, pp. 51-66.
- STEFANOWITSCH, A. y GRIES, S.Th. (2003): «Collostructions: Investigating the interaction of words and constructions», *International Journal of Corpus Linguistics* 8, 2, pp. 209-243.
- VANHATALO, U. (2002): 'Naiset motkottaa aiheesta ja nalkuttaa syystä': Kyselytestit verbien semanttisten sisältöjen arvioinnissa» ['Las mujeres reniegan con razón y se quejan con motivos para ello': el uso de cuestionarios en el análisis del contenido semántico de verbos], *Virittäjä* 3, pp. 330-353.
- , (2003): «Kyselytestit vs. korpuslingvistiikka lähisynonyymien semanttisten sisältöjen arvioinnissa – Mitä vielä keskeisestä ja tärkeästä?» [Cuestionarios vs. lingüística de corpus en el análisis de los contenidos semánticos de sinónimos próximos], *Virittäjä* 3, pp. 351-369.
- ZUBIZARRETA, L^a M^a. (1999): «Las funciones informativas: tema y foco», en Bosque, I. y Demonte, V. (dirs.), *Gramática descriptiva de la lengua española*, Madrid, Espasa Calpe, pp. 4215-4244.

Edita
SeL

