

Entropía y significado. La gradualidad semántica en las preposiciones del español

Entropy and Meaning. Semantic Gradualness in Spanish Prepositions

Francesc Reina González

UNIVERSIDAD DE BARCELONA

ESPAÑA

francesc.reina@ub.edu

<http://orcid.org/0000-0002-5720-8422>

Lluís Padró

UNIVERSIDAD POLITÉCNICA DE CATALUÑA

ESPAÑA

padro@cs.upc.edu

<http://orcid.org/0000-0003-4738-5019>

Recibido: 27-4-2020 / **Aceptado:** 7-12-2022

DOI: 10.4151/S0718-0934202401160563

Resumen

El significado que aportan las preposiciones se puede observar desde una perspectiva distribucional, es decir, qué clase de elementos semánticos anteceden y preceden un elemento lingüístico y cómo determina ese contexto su propio significado. Uno de los objetivos de nuestro trabajo fue clasificar semánticamente, a través del valor de la entropía procedente de la Teoría de la Información (Shannon, 1948), un conjunto de sustantivos que son comandados por preposiciones en español. Suponemos que los valores semánticos de las preposiciones se organizan en clases: funcionales, semifuncionales, semiléxicas y léxicas. Una vez clasificados dichos sustantivos observaremos cómo es la entropía de las clases preposicionales a la que pertenecen (el examen se realizará para tres de esos grupos). De este modo, justificaremos la denominada Hipótesis de la Gradualidad de la Similitud Semántica para esos grupos preposicionales.

Palabras clave: preposición, funcionalidad, gradualidad, entropía y semántica

Abstract

The meaning that prepositions provide can be observed from a distributional perspective, that is, what kind of semantic elements precede and follow a linguistic element and how that context determines their meaning. One of the objectives of our research was the semantic classification of a set of nouns used with prepositions in Spanish, using the entropy value obtained from the Information Theory (Shannon, 1948). We assume that the prepositions' semantic values are organized into classes: functional, semi-functional, semi-lexical, and lexical. Once the nouns are classified, we

will examine the entropy of the prepositional classes to which they belong (the observation will be carried out for three of these groups). In this way, we will justify our hypothesis concerning the semantic similarity degree for these prepositional groups.

Keywords: preposition, functionality, graduality, entropy and semantics

INTRODUCCIÓN

El estatus categorial de la preposición en las teorías lingüísticas de la segunda mitad del siglo XX, fundamentalmente las que proceden de la tradición generativista, ha estado presidido por la negación de los rasgos clasificatorios del componente categorial: [-N – V]. La propuesta de Noam Chomsky inicia un largo camino sujeto a controversias sobre hasta dónde era posible justificar esa doble negación para una de las cuatro categorías léxicas mayores. Así, y durante más de dos décadas, el objetivo explicativo se centró en la determinación de sus tareas sintáctico-estructurales a través de la asignación del caso y en la similitud sintagmática con nombres, verbos y adjetivos (las otras tres categorías léxicas).

La aparición de la noción de funcionalidad en la caracterización de las categorías lingüísticas supuso la posibilidad de flexibilizar y ajustar los rasgos delimitativos de las clases de palabras. De este modo, se abordaba no solo elementos o rasgos tradicionales de la gramática descriptiva como los determinantes, la concordancia, el aspecto o el tiempo, sino también otras categorías como el complementizador (COMP). En este ajuste se irá desdibujando la visión homogénea y unitarista sobre qué constituye un ítem lingüístico y cómo debe caracterizarse. Simultáneamente, la gramática cognitiva participa en este debate que discute la rigidez de los límites entre categorías lingüísticas y propone una visión no discreta de los ítems lingüísticos. La historia y las bases epistemológicas de esta nueva perspectiva ha sido extensamente presentada (Langacker, 2008).

Además, la colección de investigaciones con aportaciones empíricas en la década de los noventa del siglo XX, de carácter interlingüístico y comparativo, ha ido mostrando que la caracterización negativa [-N, -V] de la preposición era insuficiente (Muysken, 2008). Se fueron acumulando datos y evidencias sobre hechos semánticos en lenguas de las familias anglogermánica y románica, junto a otras lenguas donde las relaciones preposicionales se expresaban con posposiciones o caso, como el ruso y el polaco —con caso y preposiciones—; el japonés —con preposiciones y postposiciones—; o el igbo, de la familia nigercongolesa —con una sola preposición—, (Gehrke, 2008; Zeller & Ngoboda, 2018).

En ese nuevo escenario investigador sobre cómo se caracterizan las categorías lingüísticas, condiciones, requisitos y restricciones, se empiezan a encontrar evidencias y analogías entre las categorías funcionales y las preposiciones. Así, por ejemplo, se

determina que la preposición es parte de un sistema diferente, puesto que tiene más en común con determinantes, pronombres o complementizadores (Stewart, 2015).

La crisis categorial se ha ido resolviendo con las aportaciones que el significado preposicional incorpora en su conducta. Esta vía de acceso, que sea y cómo se interprete semánticamente la preposición, fue uno de los primeros territorios productivos en el ámbito de la lingüística computacional y el procesamiento del lenguaje natural (PLN), donde se sitúa nuestra aproximación. De hecho, ese interés por el significado preposicional procede de su ambigüedad en su procesamiento a causa de su carácter polisémico y de cómo conseguir explicarlo (Akkaya et al., 2012).

Uno de los modelos más efectivos en la identificación de significado ha sido el de la semántica neodistribucional (Baroni & Lenci, 2010; Boleda & Erk, 2015; Boleda & Herbelot, 2017; Nazar & Renau, 2012). El acceso, el análisis y la exploración de millones de contextos y distribuciones de las preposiciones han permitido ahondar en la compleja expresión de la semántica preposicional, tanto desde la perspectiva cuantitativa como cualitativa (Baldwin et al., 2009).

Por otra parte, la posibilidad de manejar con rapidez y eficacia observacional corpus con cientos de millones de palabras y examinarlas con herramientas estadísticas (Hastie et al., 2009) y algorítmicas del ámbito del aprendizaje automático, de los *word embedding*, o de la Teoría de la Información, están permitiendo que la cuantificación y la medición aumenten la confianza predictiva de las hipótesis de investigación sobre la disposición del significado (Baroni et al., 2014; Bentz & Alikaniotis, 2016).

Nuestra aportación parte de la insuficiente caracterización semántica de la preposición en el debate analítico de la tradición teórico-sintáctica, y se aproxima al significado desde los hechos distribucionales en corpus lingüísticos y con herramientas de predicción estadística y la medición del orden y/o azar (entropía) (Ben-Naim, 2007) de ciertos contextos de las preposiciones del español.

1. Hipótesis y objetivos

Nuestra investigación sobre la naturaleza semántica de la preposición en español, se enmarca dentro lo que denominamos como Hipótesis de la Gradualidad de Similitud Semántica (HGSS) (Reina et al., 2019; Reina, 2020) y que reproducimos en (A).

(A) Hipótesis de la Gradualidad de la Similitud Semántica

Dado un conjunto de valores semánticos descriptivos asignados a elementos lingüísticos, denominados preposiciones, predecimos que la identificación de estos valores ocurre gradualmente desde los valores más funcionales (con poco o nulo contenido semántico) a los léxicos (con un peso semántico relevante). En esta

progresión, encontramos etapas intermedias que se llaman semifuncionales y/o semiléxicas.

Así, en el experimento que presentamos a continuación, se desarrolla esta concepción de gradualidad en el contexto de los sintagmas preposicionales seleccionados por una serie de verbos y procedentes de corpus lingüísticos que detallamos a continuación.

Proponemos como magnitud de medición de la densidad semántica la entropía estadística, proveniente de la Teoría de la Información de Claude Shannon (Holik, 2016). De este modo, y como consideramos en (B), la aportación preposicional a sus contextos o distribuciones se medirá a partir de esa noción.

(B) La contribución del significado preposicional se mide gracias a la entropía estadística de las categorías semánticas a las que pertenecen los nombres seleccionados por los verbos de los cuales depende la preposición. Esta contribución sigue la siguiente correlación: a partir de la entropía estadística a mayor restricción de las categorías semánticas y mayor concentración en algunos, más baja es la entropía y más fácil es su predicción; y viceversa, a medida que se distribuyen las categorías y aumenta la entropía, también aumenta el peso semántico de las preposiciones.

Los objetivos de nuestra aportación son tres. El primero, es la verificación de la gradualidad para tres de las cuatro agrupaciones preposicionales que proponemos: las funcionales, las semifuncionales y las léxicas. En segundo lugar, examinaremos la organización semántica de los nombres que acompañan a esas clases preposicionales, según su distribución entrópica. Por último, concluiremos si es posible extender esta magnitud de medida a otras estructuras gramaticales de la preposición.

A continuación, presentaremos las clases de sintagmas preposicionales que vamos a clasificar, así como el proceso metodológico y la definición de entropía que manejamos para los cálculos que pretenden probar nuestras hipótesis.

2. Las clases de sintagmas preposicionales para el español

De la HGSS se desprende que las agrupaciones preposicionales en español son cuatro: funcionales, semifuncionales, semiléxicas y léxicas, aunque nosotros solo investigamos en este experimento tres de esas clases. Los rasgos sintáctico-semánticos predominantes, así como las estructuras gramaticales mayoritarias son los que se muestran y se comentan en los siguientes ejemplos.

Así, entre (1) y (5) se recogen casos de funcionalidad, representados en estructuras o funciones sintácticas como los complementos directos de persona (3), las perífrasis verbales (4) o las locuciones prepositivas (5).

- (1) Envía a María a comprar.
- (2) Echaron a correr.
- (3) Se recauda en beneficio de una ONG.

En estos casos la aportación semántica descrita en las referencias tradicionales es muy baja, nula o residual (Aijón 2008; Moliner, 2012; RAE, 2009). La preposición es un mero índice gramatical o un orientador que apenas contribuye a la interpretación del predicado o de la estructura.

A continuación, tenemos los que reflejan una situación intermedia desde la perspectiva de la aportación semántica, y que denominamos semifuncionales. Para ello, tomamos como punto de partida los complementos preposicionales de los verbos de régimen, es decir, aquellos verbos que exigen una preposición y cuya interpretación semántica parece estar más cohesionada entre ambos términos, el verbo y el núcleo del sintagma nominal (Gallego, 2010; Candaliya, 2016). Son los ejemplos entre (4) y (7).

- (4) El libro trata de los problemas actuales.
- (5) La comida consistió en dos primeros y dos segundos.
- (6) Ha accedido a la cátedra sin dificultad.
- (7) Se preguntaban por la desaparición de Laura.

La tercera agrupación, se produce en contextos donde el núcleo sintagmático del que depende la preposición es un nombre. En este caso, se amplía parcialmente el significado y se va acercando a una interpretación léxica. En los casos de (8) a (12) se observan casos con esa progresiva apertura.

- (8) El abatimiento por la enfermedad.
- (9) Aquel abanico de bailaora.
- (10) Un boicot sin sanciones.
- (11) El capítulo sobre adiestramiento.
- (12) El complemento para dispositivo.

Por último, los ejemplos donde la preposición aporta un significado pleno, es decir, desde la perspectiva composicional del significado son una parte indispensable de toda la secuencia y se suelen denominar valores léxicos. Algunos ejemplos serían los recogidos bajo (13) y (19).

- (13) Depositó el libro sobre la mesa.
- (14) Iremos a Mataró de vacaciones.
- (15) Quedamos a las seis.
- (16) Vivimos en Sevilla.
- (17) Murieron por el bombardeo.
- (18) Escribe con cuidado.
- (19) Corrieron tras el ladrón.

Nuestro experimento se ocupa de la observación y el análisis de tres de estas agrupaciones: la funcional, la semifuncional y la léxica, dejando las semiléxicas para otros análisis.

3. Metodología y corpus

Hemos seleccionado de corpus lingüísticos —en las referencias se recogen los enlaces digitales a los repositorios— un total 3.898 nombres insertos en los sintagmas preposicionales de una serie de 133 verbos (véase la lista del apéndice 1). Estos verbos representan los universos semánticos mayoritarios (Albertuz, 2004); tanto desde el punto de vista del aspecto verbal (estado, eventos, realizaciones o logros) como desde el punto de vista de su contenido léxico-semántico (psicológicas, de habla, de movimiento, de acción, etc.).

Esa selección se ha realizado a través de una búsqueda arbitraria —sin criterio temporal ni espacial de su procedencia— del CREA de la RAE, del Wikicorpus y del Corpus del español (también conocido como Corpus Davies).

Una vez elegidos los nombres de los sintagmas preposicionales, se han clasificado manualmente en funcionales, semifuncionales o léxicos de acuerdo con las distinciones y los rasgos descriptivos establecidos en el apartado II. Sintácticamente, nos encontramos con los tipos genéricos de complementos verbales que se recogen en (C) - (E).

(C) Complemento argumental con función de complemento directo/complemento indirecto (preposiciones funcionales).

(D) Complemento argumental con función de régimen verbal (preposiciones semifuncionales).

(E) Complemento adjunto (preposiciones léxicas).

La segunda fase metodológica consiste en la clasificación de los nombres en seis categorías semánticas, procedentes de una síntesis de las estipulaciones semánticas del TCO (*Top Concept Ontology*, Álvez et al., 2008) y que usamos con anterioridad en una investigación previa (Reina, 2020). Se trata de las categorías semánticas de (F).

(F) Animado (AN), Inanimado (IN), Entidad Abstracta (EA), Evento (EV), Locativo (LO) y Temporal (TE).

Esos nombres ocupan la posición de complemento y término en los sintagmas preposicionales. Así, en los ejemplos de (20) - (22), son los nombres como *Pedro*, *palabras*, *ceremonia*, los que serán etiquetados.

(20) Invitamos a Pedro.

(21) Confía en sus palabras.

(22) Se fueron tras la ceremonia.

Una vez clasificados manualmente los nombres y contrastada la clasificación con otro anotador humano con una coincidencia del 73%, hemos aplicado a los resultados la fórmula de cálculo sobre la entropía de tales grupos. Ahora se procede a indicar algunas aclaraciones sobre nuestra aproximación a esta noción tan fértil en la investigación científica.

El concepto de entropía ha resultado altamente polisémico en el ámbito de la ciencia y la tecnología.

En términos generales existe cierto acuerdo en caracterizar la entropía como una medida de diversidad, de desorden o de ignorancia. Desde el punto de vista estadístico, la entropía medirá cuál es el grado de aleatoriedad de un determinado conjunto. Intuitivamente, una medida de diversidad permite cuantificar la variabilidad de una población, sean datos para un mensaje cifrado o, como en nuestro trabajo, la naturaleza semántica de un conjunto de sintagmas nominales.

La ecuación que propuso Claude Shannon para la entropía (H) fue el recogido en la fórmula (I).

$$(I) H = -\sum_{x \in X} p(x) \log_2 p(x)$$

Donde X representa el conjunto de valores posibles de una variable aleatoria discreta, x varía sobre cada uno de dichos valores, y $p(x)$ es la probabilidad de aparición del valor x . La entropía H es la media ponderada del opuesto del logaritmo en base 2 de la probabilidad de cada valor posible de la variable X .

La entropía de una variable aleatoria se puede interpretar como el número de bits necesarios en promedio para codificar de forma óptima el resultado de dicha variable, suponiendo que se asignan códigos más cortos a valores más frecuentes. También es frecuente interpretarla como el grado de aleatoriedad, desorden, o impredecibilidad de la variable, ya que cuanto más impredecible es una variable, más bits son necesarios para transmitir su resultado. Es esta última interpretación la que nos interesa en este trabajo, tal como se describe en el siguiente apartado.

La entropía se calculó a partir de los nombres que se encontraban en la posición de complemento de la preposición. Esos nombres fueron clasificados manualmente para los tres tipos de clases preposicionales: funcionales, semifuncionales y léxicas.

4. Comparación y análisis de datos

En la Tabla 1, se presentan los resultados globales de la entropía en la distribución de los nombres para cada una de las agrupaciones o clases preposicionales del experimento. Cada columna contiene la distribución de las categorías semánticas de

los nombres que aparecen vinculados a cada categoría preposicional en posición de complemento. Considerando esta distribución como una variable aleatoria y aplicando la fórmula (1), obtenemos el valor de la entropía de dicha variable.

Para cada columna en la Tabla 1, se aplica la fórmula (1) como sigue: X es el conjunto de categorías semánticas nominales, y cada una de ellas es un valor posible $x \in X$. La probabilidad $p(x)$ de cada valor x se estima por máxima verosimilitud dividiendo el número de ocurrencias de la clase semántica x con una preposición del tipo correspondiente a la columna, por el número total de apariciones de preposiciones de ese tipo (casilla «total de casos»).

De esta forma, obtenemos el grado de desorden o heterogeneidad de los nombres vinculados a cada tipo de preposición (casilla «Entropía total (H)»).

Tabla 1. *Detalles de los resultados de la entropía para las tres clases preposicionales*

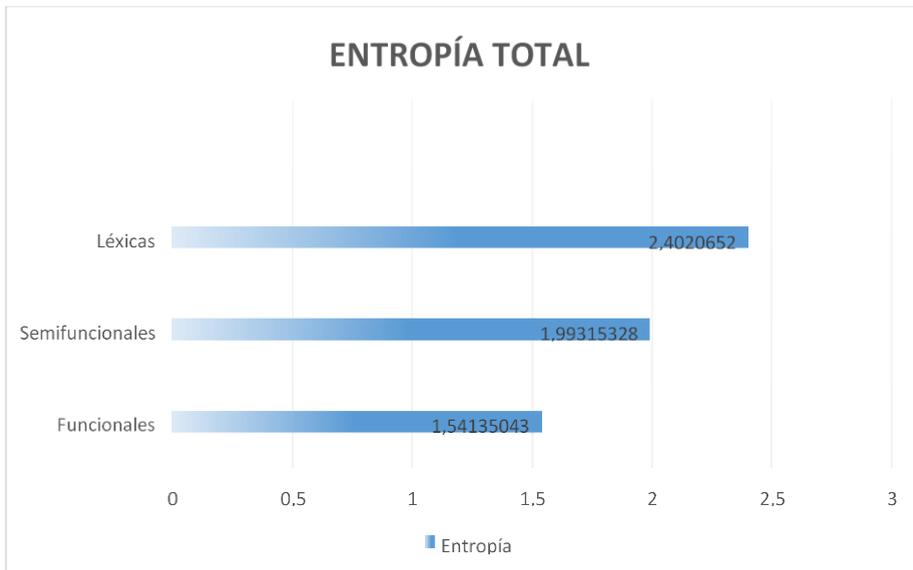
CATEGORÍAS SEMÁNTICAS (3898 nombres)	CLASES PREPOSICIONALES		
	FUNCIONALES (F)	SEMIFUNCIONALES (SF)	LÉXICAS (L)
Animado	871	203	120
Inanimado	55	160	152
Entidad Abstracta	146	696	374
Evento	73	168	181
Locativo	90	74	376
Temporal	23	26	110
Total de casos	1258	1327	1313
Entropía total (H) =	1,54	1,99	2,40

El análisis global de los resultados obtenidos, que se presentan en la Tabla 1, pasa por tener en cuenta tres aspectos. En primer lugar, comprobar qué cifras de entropías globales se obtienen para cada clase y confirmar el valor de gradualidad o de progreso entre los tres grupos: funcionales, semifuncionales y léxicos. A continuación, y de manera pormenorizada, estudiaremos la composición de cada una de las categorías semánticas de nombres de las tres clases o grupos (secciones 4.1, 4.2 y 4.3). Por último, precisaremos la confirmación empírica de la formulación de la Hipótesis Gradual de la Similitud Semántica, expresada en (1) y (2), en las conclusiones.

El total de nombres clasificados es de 3.898 sin repetición alguna. Se trata de un resultado satisfactorio y coincidente con la hipótesis gradual porque confirma la relación entre el valor entrópico y el significado preposicional en las tres clases. Así, la ausencia casi completa de significado en el caso de las preposiciones funcionales, es decir, las que participan en los sintagmas nominales de la primera columna. A su vez, las preposiciones del régimen verbal, o semifuncionales, están orientando o dirigiendo el significado de los verbos que acompañan y de ahí el aumento entrópico de los nombres. Finalmente, el grupo con mayor entropía coincide con las preposiciones con

mayor significado: las léxicas. El Gráfico 1, recoge la información de los totales detallados en la tabla anterior.

Gráfico 1. Resultados totales de la entropía para las tres clases preposicionales



La medida entrópica de las tres clases preposicionales resultantes indica que cuanto más baja es menor la variabilidad que encontramos en el número de casos (nombres por categorías semánticas) y, por tanto, mayor predictibilidad. Ese hecho se debe a que la preposición contribuye con más significado.

Un análisis pormenorizado de los resultados nos va a permitir observar que la distancia entrópica entre los tres grupos está relativamente equilibrada. De hecho, entre funcionales y semifuncionales la diferencia es de 0,45; entre semifuncionales y léxicos es de 0,41; y entre funcionales y léxicos de 0,86, tal como se muestra en la Tabla 2.

Tabla 2. Diferencias entre las entropías de las clases preposicionales

	SF	L
F	0,45	0,86
SF		0,41

Asimismo, hemos realizado una prueba de azar para comparar los resultados con los obtenidos en el experimento. Se han construido tres grupos de nombres y cada uno de ellos contiene, de manera proporcional, una tercera parte de los nombres de las tres clases, funcionales, semifuncionales y léxicos. Se han ordenado alfabéticamente los verbos del experimento hasta que obtengan una cifra arbitraria de cada tipo y una variación aleatoria de las tres clases, sin atender a ningún criterio posterior.

A partir de aquí, se vuelven a introducir los datos de clasificación de los nombres en la hoja de cálculo para verificar los nuevos resultados (2,26 para las funcionales; 2,29 para las semifuncionales; y 2,30 para las léxicas) y se comprueba que no se aprecian resultados significativos, tal y como se recoge en la Tabla 3.

Tabla 3. Resultados de la prueba de azar

Categorías semánticas	Segundo experimento Prueba de azar		
	Grupo 1	Grupo 2	Grupo 3
Animado	441	371	370
Inanimado	123	113	130
Entidad abstracta	341	424	451
Evento	123	155	143
Locativo	155	182	200
Temporal	49	50	55
Total de nombres al azar	1232	1295	1349
Entropía (H) =	2,26	2,29	2,30

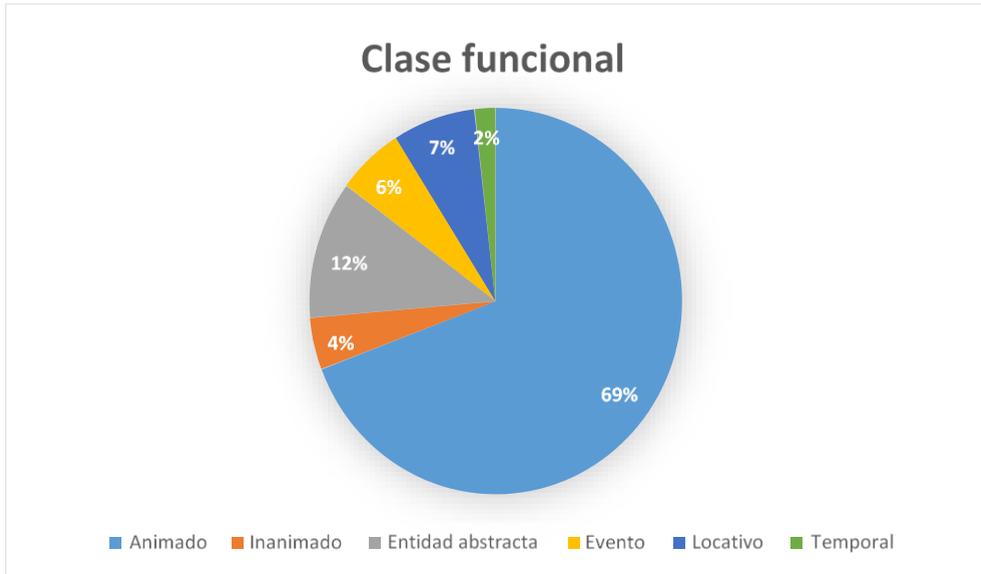
Los índices de entropía obtenidos en los tres grupos son muy parecidos, de manera que el azar no daría cuenta en modo alguno de nuestros resultados, ni tan siquiera de manera aproximada. La irrelevancia de los resultados no permite introducir factores aleatorios en nuestra justificación.

Pasamos a detallar algunos aspectos relevantes en la organización semántica de los tres grupos, en consonancia con las predicciones de la gramática descriptiva para las clases preposicionales.

4.1 La clase funcional

En el Gráfico 2, podemos ver la distribución de categorías semánticas para la clase funcional, es decir, aquellos casos donde la preposición es muy débil semánticamente.

Gráfico 2. Distribución de las categorías semánticas en la clase funcional



La preposición que domina de manera absoluta la funcionalidad en nuestro corpus es la *a*. La primera evidencia empírica es la concentración: 871 de 1.258 —un 69,23% del total de sintagmas nominales analizados— pertenecen a la categoría semántica ‘Animado’. El rasgo más frecuente es el vinculado a la condición humana, incluyendo desde las denominaciones más genéricas: hombre, mujer, niño, joven, viejo, adolescente, hasta una diversidad muy extensa de condiciones, oficios, estados, propiedades, ocupaciones o actividades encarnadas por los seres humanos: vecino, madre, funcionario, investigador, profesional, esposo, soldado, agente, víctima, escritor, trabajador. En este grupo también se incluyen los cargos públicos, electos o no, y los colectivos u organizaciones, tales como público, población, ciudadanía, autoridad, política, presidente, ministro, sindicato, universidad o grupo.

Por otra parte, todos los nombres que se refieren a seres vivos o a partes u órganos de seres vivos se incluyen también aquí como avestruz, paloma, perro, árbol, cuello o nervio.

Una síntesis de esta pluralidad expresiva se recoge en la Tabla 4.

Tabla 4. Distribución de subdivisiones en la clase funcional para la categoría semántica *Animado*

Subdivisiones de la categoría <i>Animado</i>	Ejemplos
Nombres genéricos	<i>señora, varón, persona, chaval, churumbel, etc.</i>
Gentilicios	<i>cretenses, croatas, madrileño, bávaros, etc.</i>
Nombres propios	<i>Marta, Nina, etc.</i>
Oficios y cargos	<i>empresario, tejedores, profesores, redactora, retratista, emperador, ingeniero, soldado, chef, etc.</i>
Nombres sobre relaciones personales	<i>sobrino, esposa, hijo, abuela, etc.</i>
Pronombres	<i>usted, quién, cualquiera, etc.</i>
Colectivos	<i>avanzadilla, colectividad, contingente, sección, etc.</i>
Numéricos	<i>treintena, etc.</i>
Partes de un ser animado	<i>cuello, hombro, nervio, muslo, ventrículo, córnea, etc.</i>
Condiciones (adjetivos o participios sustantivados)	<i>encadenados, allegados, concurrente, maligno, extraño, avaros, acusado, etc.</i>

El segundo grupo que predomina, a una distancia remarcable, son los sintagmas nominales que representan la categoría ‘Entidad Abstracta’, un total de 146 respecto de 1.258, con un 11,60%. Aquí nos encontramos nombres como ley, idiotez, capacitación, argumento, aprobación, abstinencia, aquiescencia, industria o fuero. La distancia cuantitativa respecto del primer grupo es considerable en términos relativos.

Por último, el tercer grupo, numéricamente soslayable, está formado por los que indican lugares o localizaciones con 90 de 1.258, es decir, un 7,15% del total. Sirvan como ejemplos casos como: *derecha, izquierda, aquí, abí*; hasta lugares genéricos tales como: *muralla, oficina, calle, fuente, avenida, fábrica, tienda, puerta, templo, pantano, canal, país, ciudad, pueblo, casa o vivienda*.

Los datos anteriores están en consonancia con la previsión descriptiva de la gramática. De hecho, se asume que los complementos directos de personas, así como los complementos indirectos, van introducidos por una *a*. Los resultados estadísticos confirman una predicción conocida y regular en el uso y en el conocimiento gramatical de la lengua. El hecho de que la mayoría holgada de nombres (un 70% del total) de los verbos que tienen estructuras sintácticas con preposiciones argumentales sea ‘Animado’ confirma una predicción establecida en la investigación descriptiva.

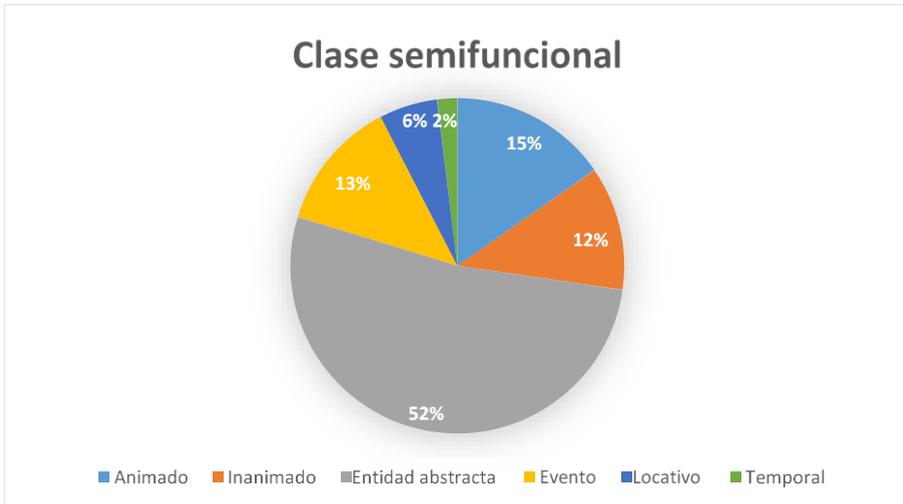
Por lo que respecta a los tres últimos grupos, representan un 12% del total (186 casos). La cifra inferior es la de temporalidad con nombres como *año, hoy, lunes, primavera* o *década*, y apenas un 1,76% (23 casos). Las otras dos, ‘Evento’ (73) e ‘Inanimado’ (55), poseen un 5,12% y 4,5% respectivamente.

El hecho cuantitativo que mejor se adapta a nuestro argumento sobre la gradualidad es la alta concentración de una categoría semántica, *Animado*, indicio de la escasa relevancia interpretativa de la preposición *a*.

4.2 La clase semifuncional

Si observamos el Gráfico 3, comprobaremos que la distribución de categorías semánticas varía respecto del primer tipo. Esta alteración se produce con un aumento parcial o relativo de los tipos y, por tanto, un reparto de casos superior y una mayor complejidad en el seno de la clase.

Gráfico 3. Distribución porcentual de categorías semánticas en la clase semifuncional



El segundo grupo de nombres lo conforman los que aparecen con los verbos de régimen verbal. Los resultados nos indican que 695 de ellos, un 52,41%, pertenecen a la categoría semántica ‘Entidad Abstracta’, tales como: *seguridad, fe, amor, concepto, verdad, igualdad, odio, bondad, alma, política, factorización, empeñamiento, conciencia, traición, democracia, idea, ilusión o deseo.*

De nuevo, como en el caso de los funcionales, presentamos la Tabla 5, ejemplificadora del espectro de la categoría semántica mayoritaria en la clase.

Tabla 5. Distribución de subdivisiones para la categoría semántica de Entidad Abstracta

Subdivisiones de la categoría semántica	Ejemplos
Entidad Abstracta	
Nombres sobre ideas y conceptos (políticas, religiosas, morales...)	<i>ideales, anticlericalismo, autonomía, pragmatismo, democracia, etc.</i>
Nombres sobre actividades y operaciones humanas culturales, científicas...	<i>música, arte, cultura, zoología, agricultura, filatelia, metáfora, etc.</i>
Nombres sobre la cognición, la conducta y la mente	<i>conocimiento, ignorancia, voluntad, cordura, capacidad, memoria, etc.</i>
Nombres sobre emociones y actitudes	<i>sectarismo, intransigencia, victimismo, astucia, coraje, etc.</i>

La presencia, el uso o la coocurrencia de nombres con valor abstracto se relaciona con la clase de significados incompletos o condicionados que coocurren con los

verbos de régimen en español o en otras lenguas como el inglés, en los denominados *phrasal verbs*. Así, los verbos preposicionales —la denominación que reciben en el español— como *calificar de, influir en, constar de, contar con, creer en, consistir en, tratar de, abogar por, incidir en, incitar a, andar en, dudar de, luchar por* o *pecar de*, por ejemplo, esperan que su caracterización semántica sea consumada con un significado externo que posibilite la determinación de todo el significado. A menudo se utiliza el concepto de ‘tema’ para aludir a ese complemento de régimen, es decir, aquello que predica el verbo. Otra forma de presentarlo sería indicar que el significado entre el verbo y el nombre está compartido, o expresan una particular e inexcusable solidaridad semántica, como una especie de trabazón irreductible.

Nuestra hipótesis confirma que la mayor parte de nombres que se combinan con esta clase de verbos consigue un 52% para un cierto tipo de expresión semántica: las ‘Entidades Abstractas’. La fuerza del dato orienta nuestra confirmación más allá de los matices introducidos por la preposición.

El segundo grupo de categorías, con cifras semejantes entre ellas, lo componen las categorías de ‘Animado’ (203), ‘Evento’ (168) e ‘Inanimado’ (159), en ese orden.

Resulta pertinente para nuestra predicción que la primera categoría del segundo grupo sea ‘Animado’, en tanto que una buena parte de los verbos de régimen se caracterizan como procesos mentales o verbales, exclusivos de la condición y la naturaleza humana, con rasgos de valoración, comunicación o sensación. Algunos de esos verbos son *advertir, consistir, tratar, decidir, advertir, abogar, acusar, apelar, ceder* o *complacer*. Del mismo modo, el desarrollo de la categoría Evento implica, en muchos casos, que los agentes, los actores o los participantes sean, necesariamente, Animado.

La temporalidad interna o tácita de los eventos se expresa en ejemplos como *celebración, reunión, encuentro, tribunal, conferencia, fiesta, actividad, concierto, boda, elecciones, cumpleaños, congreso* o *concurso*.

El tercer grupo, con porcentajes de 6% y 2%, respectivamente, incluye a Locativo y Temporal. Cualquiera de estas circunstancias expresivas matiza, de manera ulterior, el significado del núcleo verbal pero no son indispensables en la realización argumental de los verbos.

La gradualidad semántica se manifiesta con la reducción del grupo mayoritario categorías, desde el 65% de Animado en la clase funcional a un 52% en la categoría ‘Entidad Abstracta’ en la clase semifuncional. Son 13 puntos porcentuales de diferencia en el aumento entrópico y en la distribución. Luego, existiría un segundo grupo de categorías semánticas entre un 12% y un 15%. Finalmente, se encuentran las que están por debajo del 6%.

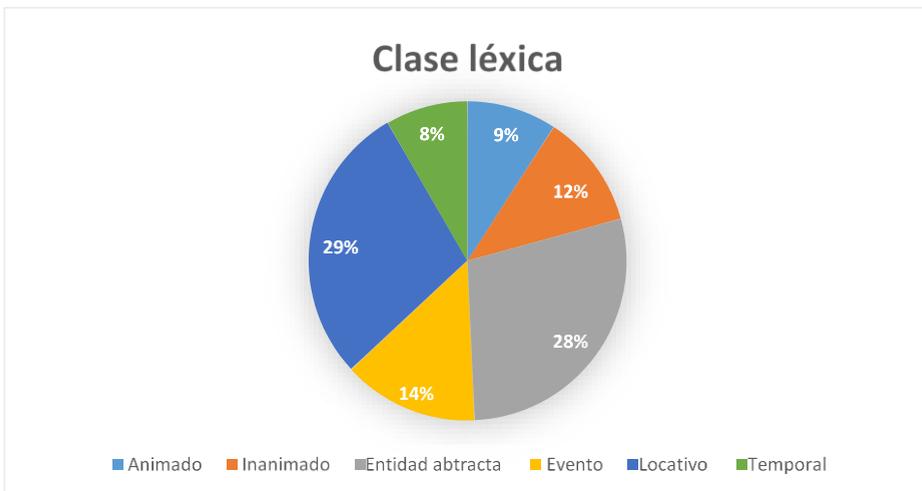
En la clase semifuncional aparecen siete preposiciones con el siguiente total de ejemplos entre paréntesis y ordenados: *contra* (5) *hacia* (10), *sobre* (61), *a* (77), *por* (125), *con* (177), *en* (270) y *de* (602).

Crece la entropía respecto de la funcionalidad y se dispersan las categorías semánticas de los contextos mientras van ganando protagonismo interpretativo las preposiciones, asociadas inexorablemente con el significado del verbo.

4.3 La clase léxica

La tercera clase, la léxica, muestra su distribución entrópica en el Gráfico 4 y confirma nuestra expectativa respecto de su distribución entrópica: es el más diverso y disperso en categorías semánticas.

Gráfico 4. Distribución porcentual de categorías semánticas en la clase léxica



El primer hecho reseñable es el relativo equilibrio en la cantidad de nombres correspondientes a cada una de las seis categorías. Estamos en la clase preposicional donde la entropía es mayor, es decir, ni la predicción, ni el orden, ni la certidumbre, tres propiedades de la medida instrumental del experimento, se cumplen. Entre las categorías semánticas con menos nombres, la ‘Temporal’ (con 110) y con más la ‘Locativo’ (con 376), solo existe una diferencia de 266 nombres. Pensemos que ese contraste se ha ido reduciendo desde los dos extremos de la clase funcional (de 21 casos a 808) y de la semifuncional (de 25 casos a 685).

Asimismo, es importante comprobar cómo ha cambiado el protagonismo de los grupos semánticos en cabeza. En la clase semifuncional, la posición central era para ‘Entidad Abstracta’. Ahora, en cambio, se comparte la primacía con el ‘Locativo’ (376).

La expresión espacial adquiere una presencia notable en el operativo semántico de las preposiciones. En este sentido, se ha presentado una categorización lingüística de las entidades espaciales que ahonda en su riqueza expresiva (Aurnague et al., 2007).

No cabe duda, además, de que la naturaleza de la localización espacial vuelve a ser muy diversa y heterogénea, tal y como se recogen en las subdivisiones de la Tabla 6.

Tabla 6. Subdivisiones de la categoría semántica de *Locativo*

Subdivisiones de la categoría semántica	Ejemplos
<i>Locativo</i>	
Nombres comunes de carácter genérico	<i>ciudad, aldea, municipio, lugar, sitio, etc.</i>
Topónimos	<i>Nepal, Neptuno, Cuba, Segovia, Cádiz, Abisina, etc.</i>
Indicaciones generales y orientaciones	<i>espacio, allá, este, norte, salida, afueras, etc.</i>
Espacios abiertos (naturales o no)	<i>cima, lomas, llanura, zoco, carretera, cauce, paso, cenagal, etc.</i>
Espacios cerrados	<i>ducha, garita, laberinto, etc.</i>
Edificios, construcciones o instalaciones	<i>comisaría, castillo, piso, etc.</i>

Por último, nos encontramos con que el grupo preposicional que menos nombres incluye es el de la categoría semántica *Animado*. Parecería que el significado verbal no espera —en términos generales— ni un receptor, ni un beneficiario, ni un destino de carácter personal, ni la culminación o término, es decir, ninguno de los referentes habituales incluidos en esa categoría semántica.

El último aspecto destacado de la clase léxica es el crecimiento notable de las dos categorías hasta entonces minoritarias ('Locativo' y 'Temporal'), como se aprecia en la Tabla 7, indicándose los casos en términos absolutos y porcentualmente.

Tabla 7. Las categorías *locativa* y *temporal* en las tres clases preposicionales

Clase de preposición	Funcional	Semifuncional	Léxica
Categoría Semántica			
Locativo	(90 / 7,15%)	(74 / 5,5%)	(376 / 28,63%)
Temporal	(23 / 1,8%)	(26 / 1,9%)	(110 / 8,3%)

La ubicación espacial es fundamental, como ya hemos indicado, en el despliegue semántico de los predicados verbales. El crecimiento en esta categoría semántica de entre un 400% a un 500% de las semifuncionales a las léxicas guarda relación con la ampliación de los campos semánticos del lugar y el tiempo en ese grupo de verbos.

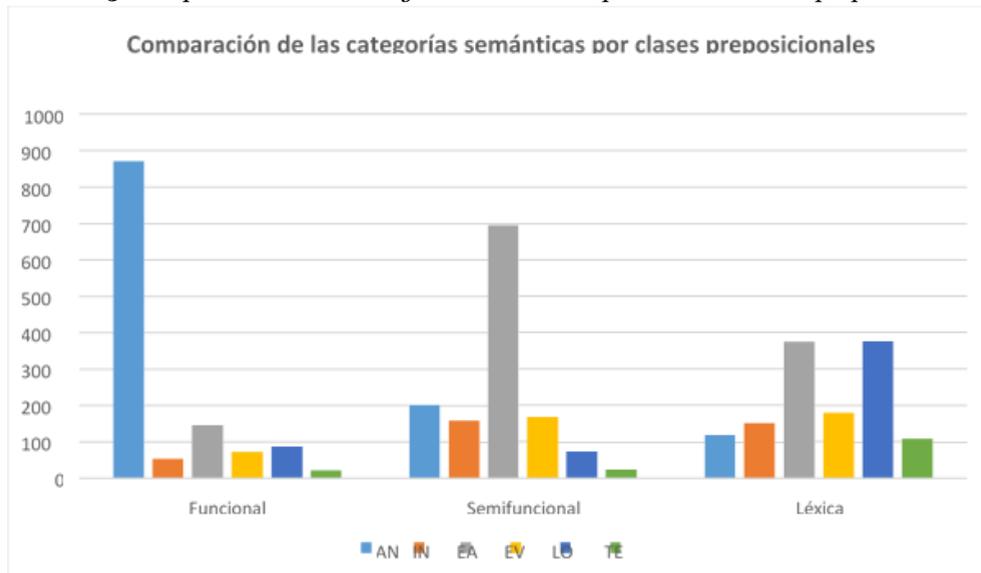
Además, el papel de la preposición es fundamental en estas tesituras porque orienta, dirige, delimita y sitúa el propósito del sentido o la acción verbal. En las lenguas con preposiciones, por tanto, su tarea es cardinal para la comprensión del significado y evitar una sobreinterpretación impropia de los requisitos de las lenguas humanas.

Otro hecho destacado en esta clase es la diversidad de preposiciones. Hemos encontrado ejemplos para dieciocho, con la siguiente relación de casos: *ante* (1), *sin* (1), *según* (5), *bajo* (6), *entre* (7), *tras* (21), *desde* (29), *durante* (32), *sobre* (36), *mediante* (39), *a* (40), *por* (44), *de* (49), *hasta* (58), *hacia* (108), *para* (119), *con* (218) y *en* (502).

En la serie anterior se observa, también, que a medida que crece la diversidad de valores léxico-semánticos atribuidos a las piezas se incrementa los ejemplos que hemos encontrado en los corpus. Las dos que se han mostrado más productivas han sido *con* y *en*, destacadas transmisoras de la información modal y espacial en los predicados de los ejemplos.

Desde el punto de vista de la comparación de los datos, en el Gráfico 5 recogemos los resultados que permiten verificar la hipótesis en relación a las dos conclusiones del experimento.

Gráfico 5. Comparación de las categorías semánticas para las tres clases preposicionales.



Nota. Leyenda por orden de aparición en el gráfico: AN: Animado, IN: Inanimado, EA: Entidad Abstracta, EV: Evento, LO: Locativo, TE: Temporal.

De la concentración inicial (Animado en funcional) se va transitando hacia una disgregación de posibilidades semánticas (‘Entidad Abstracta’ y ‘Animado’) en semifuncional, hasta la dispersión de las léxicas con dos mayoritarias muy igualadas: Locativo y Entidad Abstracta, y un reparto homogéneo de las otras (con unas cifras relativamente altas).

Recordemos cuál es la cifra de separación entre las tres clases preposicionales: 0,45 entre funcionales y semifuncionales; 0,41 entre semifuncionales y léxicas; y 0,86 entre funcionales y léxicas.

CONCLUSIONES

La primera conclusión del experimento, de la mano del análisis de los resultados en la comparación de datos, establece una correlación gradual entre la cantidad de entropía y las categorías semánticas de los nombres introducidos por los tres tipos de clases preposicionales (funcional, semifuncional y léxica).

Por otra parte, y como segunda conclusión, se confirma la tendencia del carácter progresivo de ese cambio de clases preposicionales a partir de que la distancia entrópica aumenta con una proporción bastante equilibrada.

El peso semántico de la preposición es inferior en la clase funcional en la medida en que la entropía de las categorías semánticas es baja, puesto que el motor semántico se sitúa en el verbo y el nombre del sintagma. A medida que se dispersan las categorías semánticas, la preposición irá ganando protagonismo en el significado. La prueba es que en el segundo grupo se inicia la propagación y se corrobora cuantitativamente en el tercero.

El corolario de las dos conclusiones anteriores tendría una forma como la de (G).

(G) A medida que crece la entropía en la distribución contextual de las categorías semánticas de los nombres, se enfatiza el papel o protagonismo semántico de las preposiciones en la interpretación del sintagma y viceversa.

Así, la *a* funcional de los sintagmas preposicionales contribuye con poco significado dado que ese valor semántico se puede inferir del conjunto o enunciado, o bien del núcleo (verbo), o bien del modificador o complemento (nombre). Una entropía baja determina, por tanto, que el conjunto de elementos —en este caso nombres— es más previsible porque se distribuye en menos categorías semánticas.

En los otros dos grupos, en cambio, no existe predominio exclusivo de ninguna preposición.

En este contexto casi el 50% de los casos se realizan con la preposición *de*, la mayoritaria en las estructuras de régimen verbal. La capacidad combinatoria de esta preposición está relacionada con su riqueza polisémica y su altísima frecuencia en variedad de lenguas (Rojo, 2011).

La entropía, por tanto, es una medida estable y segura en la inferencia del significado de la preposición. La aportación preposicional va en aumento siempre que el desorden, representado en la entropía, crece. Del mismo modo, la gradualidad semántica de las preposiciones se observa a través de la medida de la organización semántica de su contexto en la posición de complemento. Además, la organización interna de cada una de las clases preposicionales —funcionales, semifuncionales y léxicas— se caracteriza a partir de rasgos y condiciones semánticas predecibles en las

gramáticas normativa, descriptiva y teórica. La experimentación cuantitativa y computacional avala el reconocimiento de la descripción lingüística.

REFERENCIAS BIBLIOGRÁFICAS

- Aijón, M. A. (2008). *Complementos más bien indirectos: el régimen verbal como hecho de variación* [Ponencia]. VII Congreso de Lingüística General, UAM, Madrid.
- Akkaya, C., Wiebem, J., & Mihalcea, R. (2012). *Utilizing Semantic Composition in Distributional Semantic Models for Word Sense Discrimination and Word Sense Disambiguation* [Ponencia]. Sixth International Semantic Computing, Palermo.
- Albertuz, F. (2004). *Sintaxis, semántica y clases de verbos: clasificación verbal en el proyecto ADE SSE* [Ponencia VI Congreso de Lingüística General, Universidad de Santiago de Compostela, Santiago de Compostela.
- Álvarez, J., Atserias, J., Carrera, J., Climent, S., Oliver, A., & Rigau, G. (2008). *Consistent annotation of EuroWordNet with the Top Concept Ontology* [Ponencia]. International Conference on Language Resources and Evaluation, European Language Resources Association, Marrakech.
- Aurnague, M., Hickman, M., & Vieu, L. (2007). The Categorization of Spatial Entities. [Autoedición] *Language and Cognition: Human Cognitive Processing*. John Benjamins. <https://doi.org/10.1075/hcp.20>
- Baldwin, T., Kordoni, V., & Villavicencio, A. (2009). Prepositions in Applications: A Survey and Introduction to the Special Issue. *Computational Linguistics*, 35(2), 119-149.
- Baroni, M., & Lenci, A. (2010). Distributional Memory: A General Framework for Corpus-Based Semantics. *Computational Linguistics*, 36(4), 673-721.
- Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors [Ponencia]. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Baltimore.
- Ben-Naim, A. (2007). *La entropía desvelada. El mito de la segunda ley de la termodinámica y el sentido común*. Tusquets Editores.
- Bentz, C., & Alikaniotis, D. (2016). *The Word Entropy of Natural Languages*. arXiv:1606.06996v1.
- Boleda, G., & Erk, K. (2015). Distributional Semantic Features as Semantic Primitives –or not. *Association for the Advancement of Artificial Intelligence*, Primavera, 2-4.

- Boleda, G., & Herbelot, A. (2017). Formal Distributional Semantics: Introduction to Special Issue. *Computational Linguistics*, 42(4), 619-635.
- Candalija, J. A. (2006). *Preposiciones vacías del español: complemento de régimen preposicional* [Tesis de doctorado]. Universidad de Alicante.
- Gallego, A. J. (2010). El complemento de régimen verbal. *Lingüística Española Actual*, 32(2), 223-258.
- Gehrke, B. (2008). *Ps in Motion: On the Semantics and Syntax of P Elements and Motion Events*. LOT.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Holik, F. (2016). *Teoría de la información de Claude E. Shannon*. En C. E. Vanney, I. Silva, & J. F. Franck (Eds.), *Diccionario Interdisciplinar Austral*. <http://dia.austral.edu.ar/Teoría de la información de Claude E. Shannon>
- Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press.
- Moliner, M. (2012). *Uso de las preposiciones*. Gredos.
- Muysken, P. (2008). *Functional Categories*. Cambridge University Press.
- Nazar, R., & Renau, I. (2012). *Agrupación semántica de sustantivos basada en similitud distribucional. Implicaciones lexicográficas* [Ponencia]. V Congreso Internacional de Lexicografía Hispánica, Universidad Carlos III, Madrid.
- RAE (2009). *Nueva gramática de la lengua española*. Espasa Libros.
- Reina, F., Castellón, I., & Padró, Ll. (2019). Towards the Prepositional Meaning via Machine Learning: A Case Study of Spanish Grammar. *Glottometrics*, 44, 1-15.
- Reina, F. (2020). *Significado, distribución y frecuencia de la categoría preposicional en español. Una aproximación computacional* [Tesis de doctorado]. Universidad de Barcelona, Barcelona.
- Royo, G. (2011). Sobre la frecuencia de verbos y esquemas sintácticos. En J. J. Bustos Tovar (Coord.), *Homenaje a Antonio Narbona* (pp. 902-907). Universidad de Sevilla.
- Shannon, C. E. (1948). A Mathematical Theory of Information. *The Bell System Technical Journal*, 27, 379-423 y 623-656.
- Stewart, J. (2015). *Learning Functional Prepositions* [Tesis de doctorado]. CUNY, Nueva York.

Zeller, J., & Ngoboka, J. P. (2018). Agreement with locatives in Kinyarwanda: a comparative analysis. *Journal of African Languages and Linguistics*, 39(1), 65-106.