

Procesos de gramaticalización y lexicalización a la luz de los corpus académicos*

Cristina BUENAFUENTES DE LA MATA
Carlos SÁNCHEZ LANCIS
Universitat Autònoma de Barcelona

1. LOS CORPUS DE LA ACADEMIA

Como es bien sabido, la lingüística de corpus es una disciplina muy reciente. Según Rojo (2008), tiene su origen en la confección del *Brown Corpus* en 1964, por lo que cuenta con casi medio siglo de existencia. Sin embargo, en tan pocos años de vida se ha convertido en una herramienta indispensable en los análisis lingüísticos, tanto sincrónicos como diacrónicos. Pese a que en la lingüística española el desarrollo de corpus se inició con cierto retraso con respecto a otras lenguas, la evolución experimentada en estos años ha sido importantísima, tal y como afirma Rojo (2010: 12). Sin duda, la Real Academia Española ha contribuido al avance en este ámbito. Como afirman Pascual & Domínguez (2009: 79), la Institución “lleva casi diez años dedicada a la preparación de materiales que permitan a los filólogos disponer de un número razonable de ejemplos del léxico español empleado a lo largo de la historia”. Ese espíritu, además del objetivo fundamental de que esos materiales fueran la base documental para la elaboración de un diccionario histórico para el español, constituyó la génesis del *Corpus diacrónico del español (CORDE)* “que cuenta en la actualidad con más de 300 millones de registros léxicos desde los orígenes hasta 1975, complementados con los más de 150 millones que, procedentes de los últimos treinta años, integran el *Corpus de referencia del español actual (CREA)*” (Pascual & Domínguez 2009: 79)¹.

Así pues, los corpus académicos no han tenido como objetivo principal en su confección su aprovechamiento desde el punto de vista investigador, ya que su propósito inicial fue constituir la fuente de datos primaria para la labor lexicográfica de la Academia. De todas maneras, a medida que se elaboraban los dos bancos de datos, se observaban las posibilidades que podían tener en otros terrenos. Por ello, Madroñal (2003: 178) comentaba que “aparte de esta utilidad lexicográfica, la consulta de corpus es de utilidad para los editores de textos clásicos por la facilidad de encontrar ejemplos paralelos; pero también servirá a los estudiosos de la historia del español, que podrán encontrar en él ejemplos de la evolución de determinada palabra o expresión”. Así, desde el año 1998, la Real Academia ha puesto a disposición de la comunidad científica sus dos grandes Bancos de Datos, lo que constituye, en palabras de Sánchez Sánchez & Domínguez (2007: 138), “el recurso más importante del que se haya podido disponer jamás para el estudio de esta lengua”.

* La presente investigación ha sido parcialmente financiada con una ayuda del MEC y FEDER (n.º de ref. FFI2011-29440-C03-01 y FFI2011-24183) y de la CIRIT del Comissionat per Universitats i Recerca de la Generalitat de Catalunya (n.º de ref. 2009SGR 1067).

¹ Para una descripción detallada de ambos corpus, *vid.* Sánchez Sánchez & Domínguez (2007).

Sin duda, ambos corpus son sólo el inicio de una larga lista de bancos de datos que se han elaborado para el español, como son el *Corpus del español*, *Biblia Medieval* o el *Corpus de Documentos Españoles Anteriores a 1700 (CODEA)*, por sólo citar tres de ellos y dejando de lado algunos excelentes corpus desarrollados para otras lenguas iberorrománicas². Por tanto, la lingüística de corpus en español demuestra gozar de una excelente salud no sólo por la productiva creación de distintos bancos de datos, sino por el auge en su aplicación a los estudios lingüísticos, en general, y a los de carácter diacrónico, en particular. De hecho, la propia Academia, a propósito del *CORDE*, señala que “Hoy es fuente obligada para cualquier estudio diacrónico relacionado con la lengua española. La Academia utiliza sistemáticamente el *CORDE* para documentar palabras, para calificarlas de anticuadas o en desuso, para saber el origen de algunos términos, su tradición en la lengua, las primeras apariciones de las palabras [...]”. Esta afirmación no hace más que constatar que, actualmente, los corpus se han convertido en una herramienta imprescindible en la descripción lingüística, ya que son incontables los estudios que se basan en ellos para establecer o validar hipótesis.

La ventaja del uso de corpus en la investigación no sólo reside en poder disponer de un numeroso conjunto de textos que se puede consultar simultáneamente, sino en las posibilidades que ofrecen para recuperar la información obtenida, que, posteriormente, será sometida a la interpretación del investigador³. Es evidente que tanto el *CORDE* como el *CREA* reúnen un número muy elevado de palabras (450 millones entre ambos, según señalábamos antes), por lo que, desde el punto de vista cuantitativo, se trata de dos de los bancos de datos más importantes del español. Pero su valía también se refleja cualitativamente, por lo que el investigador puede sacar mucho más provecho de lo que a primera vista en ellos se contiene. La mayor virtud de *CORDE* y *CREA*, según Rojo (2010: 25), es su sistema de codificación, que permite “consultas tan abiertas como dicten las necesidades de los usuarios”. Esta codificación posibilita al investigador moldear el corpus a su medida, ya que puede acotar según los distintos parámetros que se han tenido en cuenta en la codificación, es decir, por año, autor, obra, tipología textual, etc. Pero, además, el usuario puede también adaptar los datos obtenidos a sus necesidades gracias a las posibilidades que ambos corpus ofrecen a la hora de ordenar tales resultados (por pivote, por la palabra que aparece a la derecha o a la izquierda, etc.). Ello se debe, esencialmente, a la exhaustiva labor de codificación que subyace tras ellos. Gracias a esto, la explotación que se puede realizar permite en muchos casos poder observar el proceso de gramaticalización y lexicalización que experimenta un elemento en la lengua y también establecer la periodización del fenómeno. En definitiva, estos corpus facilitan la recopilación y ordenación de la información que contienen, pero es el investigador en última instancia quien debe regir todo el proceso para realizar finalmente su interpretación.

Evidentemente, a medida que va avanzando la Lingüística de corpus como disciplina, se pretende sacar más partido de los bancos de datos. Y de esta exigencia no han es-

² Vid. a este respecto el trabajo de Briz & Albelda (2009).

³ Para un análisis de las ventajas e inconvenientes del uso de corpus en el análisis lingüístico, vid. Rojo (en prensa).

tado exentos ni el *CORDE* ni el *CREA*. Las carencias más evidentes y recurrentemente señaladas en la bibliografía han sido dos. En primer lugar, que el motor de búsqueda se ha quedado obsoleto, por lo que no permite sacar todo el provecho a la detallada codificación antes mencionada. En segundo lugar, la ausencia de una lematización que facilite la realización de estudios morfosintácticos sin tener que limitar necesariamente las búsquedas a verbos más frecuentes o a estructuras sintácticas concretas, por ejemplo. Ya en el año 2003, Madroñal señalaba para el *CORDE* que la Academia quería que este corpus presentase “una lematización y una etiquetación gramatical, que en estos momentos se está empezando a desarrollar”. Sin embargo, tal objetivo no ha visto todavía la luz, pese al salto cualitativo que supondría su implementación, tal y como ha señalado Rojo (2010: 32).

De todos modos, estas carencias no restan valor al *CORDE* y el *CREA* como herramientas para la investigación lingüística. Y este hecho queda ratificado por la cantidad de trabajos que se han basado en ambos bancos de datos para poder demostrar, refutar, describir o analizar distintos fenómenos lingüísticos, entre los que se encuentran los procesos de gramaticalización y lexicalización. Este trabajo quiere servir también para dar una pequeña muestra de ello.

2. LOS PROCESOS DE GRAMATICALIZACIÓN Y LEXICALIZACIÓN

Actualmente, la teoría de la gramaticalización presenta un desarrollo teórico importante desde que Meillet en 1912 definiera por primera vez el término. Según señalan varios trabajos (Kuryłowicz 1965, Lehmann 1985, Heine, Claudi & Hünnemeyer 1991, Company Company 2003, entre muchos otros), todas las aportaciones que se han realizado durante este tiempo coinciden en señalar, a grandes rasgos, que un proceso de gramaticalización transforma un elemento léxico en un elemento gramatical o convierte un elemento gramatical en otro más gramatical. Por su parte, la lexicalización presenta un menor desarrollo teórico, por lo que los rasgos que se le atribuyen suelen extraerse a partir de la gramaticalización. Recientemente, trabajos como los de Brinton & Traugott (2005) o Elvira (2006) han profundizado en este último y han señalado que, de manera general, puede considerarse que mediante un proceso de lexicalización se transforma un elemento gramatical en uno léxico o un elemento léxico se hace más léxico.

Si ahondamos más en la caracterización de ambos procesos, se pueden establecer más concretamente qué manifestaciones del cambio lingüístico puede considerarse que responden a un proceso de gramaticalización o de lexicalización. Así, la gramaticalización consistiría en:

— La conversión de un elemento léxico en un elemento gramatical, como ocurre, por ejemplo, en la transformación de *mente* en el sufijo adverbializador *-mente*.

— El aumento de la funcionalidad de una unidad ya gramatical. Por ejemplo, cuando un participio de presente se convierte en una preposición (*vid.* el caso de *durante*) o en un adverbio (como ocurre en *bastante*).

— La creación de un nuevo elemento gramatical en el sistema, como en el caso de la formación del artículo en romance a partir del demostrativo latino *ILLE*.

En cuanto a la lexicalización, ésta puede basarse en:

— La transformación de un elemento gramatical en uno léxico, como sucede con el compuesto *nomeolvides*, lexicalización de la oración *no me olvides*.

— El incremento de los valores léxicos de una unidad que ya es léxica, como ocurre con el vocablo *hidalgo*, lexicalización del compuesto sintagmático *hijo de algo*⁴.

Aunque por las definiciones presentadas pudiera pensarse que ambos procesos son contrarios, su observación en la evolución lingüística demuestra que esto no es así, tal y como señalan trabajos como los de Moreno Cabrera (1998) o Lehmann (2003). En primer lugar, porque en determinadas circunstancias pueden llegar a interactuar y, en segundo lugar, porque son manifestaciones del cambio lingüístico y, como tales, presentan muchas características en común: son procesos graduales que, por tanto, se desarrollan en un *continuum*⁵; unidireccionales, es decir, irreversibles; pueden producir modificaciones de carácter morfofonológico⁶, y ambos implican un cambio semántico.

Como se puede apreciar, son muchos los puntos en común de ambos fenómenos, pero también resultan muy relevantes los rasgos que los diferencian. En primer lugar, ya se ha comentado que parten y desembocan en ámbitos lingüísticos distintos (a grandes rasgos, la gramaticalización se inicia en el léxico y finaliza en la sintaxis, mientras que la lexicalización comienza en la sintaxis para acabar en el léxico). En segundo lugar, la gramaticalización conlleva pérdida de autonomía y adquisición de una naturaleza ligada, por tanto, dependiente. La lexicalización, contrariamente, implica adquisición de independencia funcional por parte del elemento lexicalizado.

Finalmente, tal y como se ha comentado, ambos procesos provocan un cambio semántico, aunque éste es de distinta naturaleza en cada caso. En este sentido, la gramaticalización produce una abstracción del significado que, según los enfoques, ha sido entendida como pérdida de contenido léxico o como generalización de la significación. Por el contrario, la lexicalización conlleva una concreción del significado, de modo que la unidad pasa a tener un sentido más específico. En relación a este cambio semántico, autores como Moreno Cabrera (1998) han asociado la gramaticalización con la metáfora y la lexicalización con la metonimia. Sin embargo, cabe precisar esta identificación. En primer lugar, ni la metonimia se restringe a la lexicalización, ni la metáfora sólo se produce en los procesos de gramaticalización, por lo que tanto metáfora como metonimia se pueden dar tanto en la gramaticalización como en la lexicalización. Así pues, pese a que la metonimia se asocia con la concreción y la metáfora con la abstracción, no hay que olvidar que la metáfora es, esencialmente, cualquier proceso de analogía por el que concebimos un concepto en términos de otro, mientras que la metonimia es un proceso referencial por contigüidad a través del cual efectuamos una remisión a un elemento mediante otro de mayor relevancia. En segundo lugar, ni todos los procesos de gramaticalización y lexicalización implican un cambio semántico explicable a partir de una metáfora o una metonimia, ni siempre una metáfora o una metonimia dan pie a gramaticalizaciones o lexicalizaciones.

⁴ De hecho, esto se observa también en el paso del latín al español en palabras como *pezuña* (del latín PEDIS UNGULA), *condestable* (del lat. COMES STABULI), *melocotón* (del lat. MALUM COTONIUM), *musaraña* (del lat. MUS ARANEUS), etc.

⁵ “En el paso de una categoría a otra existe siempre un estadio intermedio donde coexisten las dos estructuras, solapándose conceptual y morfosintácticamente sus respectivas propiedades en determinados contextos” (Cifuentes Honrubia 2003: 32).

⁶ Por ejemplo, pérdida de sustancia fónica, aglutinaciones, etc.

En definitiva, la ventaja de los procesos de gramaticalización y de lexicalización reside en sus posibilidades a la hora de adaptarse a las peculiaridades del cambio lingüístico.

3. DEL LÉXICO A LA SINTAXIS: LA APARICIÓN DE NUEVAS LOCUCIONES PREPOSICIONALES POR GRAMATICALIZACIÓN

Tal y como señalan en su nueva gramática la Real Academia Española y la Asociación de Academias de la Lengua Española (2009: § 29.1g), “las preposiciones forman una clase gramatical CERRADA”. Sin embargo, el número de elementos lingüísticos que desempeñan tal función ha sufrido cambios a lo largo de la historia del español, ya que se han perdido algunas preposiciones al ser sustituidas por otras de creación más reciente (el caso de *so* ‘debajo’ reemplazada por *bajo* o *cabe* por la locución *junto a*) o se han incorporado a este inventario nuevas unidades producto de un proceso de gramaticalización, como ha sucedido con los otrora participios de presente latinos *durante* y *mediante*, estudiados ya en diferentes trabajos. A pesar de ello, el número de elementos que se puede considerar estrictamente preposiciones en español sigue siendo muy limitado, por lo que, siempre desde una perspectiva gramatical, no agotan ni mucho menos la gran variedad de valores semánticos que la lengua utiliza para expresar diferentes tipos de relaciones. En estos casos, uno de los recursos más empleados es el de la creación de locuciones preposicionales mediante procesos de gramaticalización, que vienen así a complementar y subsanar las limitaciones señaladas, como ha estudiado Cifuentes Honrubia (2003). De entre todas, merece nuestra atención la locución *esquina a*, la cual se emplea para expresar ubicación o dirección. Esta estructura, que ha recibido muy poca atención, ha sufrido un proceso de gramaticalización al reanalizarse un sustantivo con sentido locativo (*esquina*) como un elemento con valor preposicional.

El sustantivo *esquina* aparece documentado en el *CORDE* a partir del siglo XIV y se mantiene como tal hasta la actualidad:

- (1) Et en cada uara que den vna pulgada al traues. & que mjdan el panno por la *esquina* del (*CORDE*: 1348, Anónimo, *Ordenamiento de Alcalá de Henares*).
- (2) E los almogauares que cauauan sacaron los cantos de la *esquina* de la torre (*CORDE*: c 1348-1379, Anónimo, *Gran crónica de Alfonso XI*).

Como se puede constatar, este vocablo se comporta en su primera documentación como un sustantivo con el significado de ‘parte exterior de un ángulo’, precedido por un determinante y acompañado de complementos del nombre introducidos por la preposición *de*, como en *la esquina del* en (1) o *la esquina de la torre* en (2), en los que se alude a la ubicación de una parte concreta de una tela o una edificación, respectivamente. Sin embargo, a partir de principios del siglo XV también es posible hallarlo en construcciones prepositivas, en algunas ocasiones con un claro grado de fijación:

- (3) La çiuðat de costantinopla es muy bien çercada de alto muro & fuerte & de fuertes torres Et grandes que ha en[e]lla de tres *esquinas* & *de esquina a esquina* ha seys mjllas asy es que en derredor de toda la çiuðat ha diez & ocho mjllas que son seys leguas & las dos partes della çerca la mar & la otra tierra & al vn cabo *a la esquina* que non çerca el mar en vn alto estan los palaçios del enperador & [...] (CORDE: c 1406, Ruy González de Clavijo, *Historia del gran Tamorlán*).

Si bien en *a la esquina* el sustantivo conserva el artículo para indicar un lugar, no ocurre lo mismo con la estructura *de esquina a esquina*, en la que ya no es posible la presencia del determinante ni la variación de número (*cfr.* el sintagma *tres esquinas* inmediatamente anterior), características del comportamiento de todo nombre. En este caso, el valor de ubicación propio del sustantivo, categoría a la que sigue perteneciendo en este ejemplo, se emplea para señalar una dimensión espacial con un origen y un final mediante la presencia respectiva de las preposiciones *de* y *a*. De todos modos, no se trata de una construcción muy habitual en el CORDE, ya que sólo se hallan 42 casos en 29 documentos (1/1 en el siglo XV, 20/11 en el siglo XVI, 15/11 en el siglo XVII, 0/0 en el siglo XVIII, 2/2 en el siglo XIX y 4/4 en el siglo XX). Dada la poca frecuencia de esta estructura en español moderno, resulta muy interesante observar la aparición de la locución *esquina a* con valor preposicional, considerada como tal por la nueva gramática académica (NGLE: § 29.9j), a pesar de que los distintos diccionarios también académicos no la recogen, por lo que podría pensarse que se trata de una construcción muy reciente en español. Por ello, resulta fundamental disponer de un corpus para realizar el estudio de la evolución de este sustantivo a la categoría preposicional, fruto, por supuesto, de un proceso de gramaticalización, como se desprende del análisis de los siguientes ejemplos:

- (4) [...] de manera que el alquitrabe ha de hazer *esquina a* la parte vaja de dicho çimborio (CORDE: 1596, Anónimo, *Capitulación con Matías Gárate, obrero de villa* [...]).
- (5) [...] hipotecamos unas casas que tenemos nuestras propias en esta villa, en la parroquia de San Ginés, en la calle de las Fuentes que sale *esquina a* la avaxada de los caños del Peral, [...] (CORDE: 1624, Anónimo, *Censo fundado por D. Francisco de Montalvo* [*Documentos sobre la vida privada española*]).
- (6) Una casa en Guadalajara, calle de la Zapatería junto a la iglesia de San Andrés, *esquina a* la calle que va al convento de San Juan de Dios (CORDE: 1790, Anónimo, *Bienes convento de los Remedios* [*Documentos de los conventos de Guadalajara*]).
- (7) Yo vivía en una hermosa casa de la calle de la Inquisición, *esquina a* la Flor Baja, cerca del edificio de la Inquisición de Corte y a poca distancia de los Premonstratenses (CORDE: 1876, Benito Pérez Galdós, *La segunda casaca*).
- (8) Se me ocurrió la idea, y comencé a escribir la obra en París, en un pequeño hotel de la calle de Vaugirard, que estaba *esquina a* la calle de Tournon y enfrente del Senado (CORDE: 1944-1949, Pío Baroja, *Desde la última vuelta del camino. Memorias*).
- (9) [...] éstos entregaron a la autoridad judicial las llaves de la casa de aquél, en Jesús Peregrino *casi esquina a* Belascoáin (CORDE: 1943-1974, Ángel Augier, *Prosa varia*).

La búsqueda de datos de la incidencia de la combinación del sustantivo *esquina* con la preposición *a* con valor locativo seguida de un complemento nominal en el *CORDE*, nos proporciona los siguientes resultados. En primer lugar, aparentemente no se observa la presencia de dicha estructura hasta el siglo XVI, como se puede constatar con *hazer esquina a la parte vaja de dicho çimborio* en (4). Esta construcción aparece precedida por el verbo *hacer* en 2 únicos casos en 2 documentos durante todo el siglo XVI, y se mantiene en el resto de las etapas con 5/3 en el siglo XVII, 1/1 en el siglo XVIII, 5/4 en el siglo XIX y 8/8 en el siglo XX, todos pertenecientes a España con excepción de 4 casos de Cuba y 1 de Puerto Rico en este último siglo, por lo que se trataría de una estructura de origen peninsular. El hecho de que el sustantivo aparezca sin ningún determinante y tras el verbo transitivo hace pensar que se trata en sus orígenes de una locución verbal formada por un verbo de apoyo (*hacer*) y un complemento locativo introducido por la preposición *a*. Como se puede apreciar por los datos recogidos, no resulta una estructura muy frecuente en todos los tiempos, ya que es casi dos siglos más tardía que la primera documentación del sustantivo solo, pero curiosamente se mantiene hasta fechas muy recientes (principios de la segunda mitad del siglo XX). En segundo lugar, se advierte en el siglo XVII la presencia de esta combinación con el verbo *salir* con valor semántico figurado en *sale esquina a la avaxada de los caños del Peral* en (5), en donde nuevamente aparece el sintagma preposicional introducido por la preposición *a* precedido por el sustantivo *esquina*, con la diferencia de que ya no precisa de la existencia de un verbo de apoyo desemantizado, aspecto que le otorgaría a dicha estructura una cierta independencia sintáctica. Esta evolución parece quedar demostrada en el momento en que, por primera vez, se halla esta construcción en el siglo XVIII totalmente independiente, sin depender explícitamente de ningún verbo, como en *esquina a la calle* en (6), funcionando como aposición de otro sintagma preposicional (*junto a la iglesia de San Andrés*). Es el único ejemplo en esta época que convive con sólo otro introducido por el verbo *hacer*. Sin embargo, a partir de esta fecha, se constata la presencia mayoritaria de esta combinación de forma independiente para expresar principalmente una ubicación relacionada con una calle específica (23/13 en el siglo XIX frente a 5/4 con el verbo *hacer*; y 68/47 en el siglo XX frente a 8/8 con *hacer*). Por todo esto se puede afirmar, a partir de los datos analizados, que nos encontramos entonces ante una auténtica locución preposicional a consecuencia de la gramaticalización del sustantivo, como se puede comprobar en *esquina a la Flor Baja* en (7), en donde ya simplemente se usa el nombre propio de la calle; en *estaba esquina a la calle de Tournon* en (8), en donde el verbo *estar*, único caso en todo el corpus, selecciona un complemento locativo introducido por esta locución; o en *casi esquina a Belascoáin* en (9), en que la presencia de un adverbio de cantidad que modifica el valor de la locución preposicional (5/5 en el siglo XX) sólo es posible si este elemento ha perdido sus rasgos nominales.

Si se realiza un estudio de las estructuras anteriores en el *CREA*, se constata su plena pervivencia en el español actual. Así, en el caso de la construcción *de esquina a esquina*, su presencia sigue sin ser muy significativa (sólo se hallan 13 ejemplos en 12 documentos), aunque posee una difusión general, ya que corresponde a diferentes países (España, Bolivia, Cuba, Ecuador, México y Venezuela) y distintas temáticas (Ficción, Artes, Ciencia y Tecnología, Ciencias sociales, creencias y pensamiento, Ocio, vida cotidiana y Oral), como se puede ver en (10):

- (10) Yo venía de un pueblo pobre y aunque la casa de mis abuelos quedaba en la calle Real no había más que una bombilla de pocas bujías en cada esquina que apenas alumbraba el área alrededor del poste, haciendo más espesa la oscuridad *de esquina a esquina* (CREA: 1986, Guillermo Cabrera Infante, *La Habana para un infante difunto*. Cuba).

Sin embargo, por lo que respecta a la locución preposicional *esquina a*, su proliferación es mucho mayor (92 casos en 38 documentos), como era de esperar a tenor de los resultados del *CORDE*, y aunque no se ciñe a una temática específica, sí que continúa apareciendo básicamente en textos de España (84 casos en 33 docs.), con algunos pocos ejemplos de Cuba (7 casos en 4 docs.) y Puerto Rico (1 caso en 1 doc.), por lo que habría que seguir considerándola como una construcción más propia del español peninsular. En relación a las posibles variaciones gramaticales del complemento que aparece pospuesto, no se observan grandes diferencias con los casos analizados anteriormente:

- (11) Cuando ocurre un devastador incendio en su fábrica —sita en Calzada del Monte *esquina a* Zulueta, en un céntrico barrio habanero—, la acción de los vecinos, aunque rápida, no logra salvar a varios aprendices que mantiene sometidos a castigo, en el cepo (CREA: 1998, Reynaldo González, *El bello habano. Biografía íntima del tabaco*. Cuba).
- (12) Situación: Calle Potosí, *esquina a* calle Bolivia (Príncipe de Vergara, 251) (CREA: 1986, Prensa, *ABC*, 1-11-1986: *La cesta de la compra*. España).
- (13) Específicamente la acción comienza la tarde del 16 de noviembre de 1780 en una casa de la calle Cuba *esquina a* la de Sol, [...] (CREA: 1977, Isabel Hernández de Norman, *La novela criolla en las Antillas*. Puerto Rico).
- (14) Alguna que otra vez me encontraba a Boyer que tomaba el autobús en Serrano, *casi esquina a* la plaza de la República Argentina (CREA: 1993, Julio Feo, *Aquellos años*. España).

En estos ejemplos el sintagma nominal introducido por la locución puede llevar determinante (*esquina a la de Sol* en (13) o *casi esquina a la plaza de la República Argentina* en (14), en donde se halla de nuevo un adverbio de cantidad como modificador de la construcción), o carecer de él (*esquina a calle Bolivia* en (12) o *esquina a Zulueta* en (11)), en donde la aposición nominal llega a ser sustituida por el simple nombre propio. Finalmente, se debe señalar la posible relación existente entre la locución estudiada y las estructuras en las que el sustantivo *esquina* va seguido de un sintagma nominal sin que exista una preposición de enlace:

- (15) En Seseña Nuevo, un antiguo realojamiento que nació tras la Guerra Civil, una promotora ofrece nueve chalés adosados en la calle Barataria *esquina calle Don Quijote* (CREA: 2001, Prensa, *El Mundo - Su Vivienda (Suplemento)*, n.º 198, 23-3-2001: *De paseo*. España).

Si se realiza una búsqueda de esta construcción a partir de la presencia del sustantivo *calle*⁷, no se encuentra ningún ejemplo de ella en todo el *CORDE*, mientras que sí se hallan 6 casos en 6 documentos en el *CREA* (3 de España y 3 de la República Dominicana). A falta de un estudio más pormenorizado que no es posible realizar aquí, no sería difícil establecer equivalencias sintácticas entre ambas estructuras a partir de los contextos en que aparecen y las características gramaticales del complemento que llevan, por lo que podría ser visto como el siguiente paso en el proceso de gramaticalización del sustantivo *esquina* hacia un uso preposicional independiente⁸.

4. DE LA SINTAXIS AL LÉXICO: LA LEXICALIZACIÓN DE SINTAGMAS EN LA FORMACIÓN DE PALABRAS POR COMPOSICIÓN

En español existe una serie de formaciones creadas mediante mecanismos compositivos que tienen la particularidad de reflejar claramente una estructura oracional. Así presentan un elemento verbal como parte del compuesto, que incluso puede ir acompañado de pronombres y adverbios (como ocurre, por ejemplo, en *bienteveo*, *nomeolvides*, *pésame*, *siguemepollo*), de sintagmas nominales formados por determinantes y sustantivos (como en *besalamano*), de sintagmas preposicionales (por ejemplo, *cenaoscuras*, *tentempié*, *tentenaire*) y de otro tipo de complementos verbales (como, por ejemplo, *sabelotodo*). Según Bustos Gisbert (1986: 263):

[...] todos estos compuestos reflejan un contexto verbal o situacional, en que las relaciones de los términos presentes en el compuesto y el referente se sitúan en un nivel en que los compuestos evocan una actividad o una entidad en mayor o menor grado en ellas relacionada, pero sin que esa relación —por su propia lejanía— sea de alguna manera formalizable, sin que llegue a formar un paradigma de regularidades.

No cabe duda, pues, de que en estos casos se ha producido una lexicalización, ya que a partir de una cláusula (plano sintáctico) obtenemos un nuevo vocablo (plano léxico). Por ello, Pena (1991: 335) afirma que estas formaciones nacen en la sintaxis mediante un proceso de yuxtaposición y, diacrónicamente, son reanalizadas por los hablantes como palabras simples. También destaca este hecho Alcoba (1988: 116), cuando señala que “no es difícil encontrar ejemplos de cómo algunos procesos morfológicos son el resultado diacrónico de lexicalizaciones de formaciones sintácticas”.

El desarrollo que sigue este tipo de compuestos parece partir de una construcción sintáctica en la que se lexicaliza el significado (por tanto, se produce la lexicalización de carácter semántico) y, además, se da la aglutinación de los componentes de la formación hasta presentar el máximo grado de cohesión.

Este proceso parece ser el que ha sufrido una palabra como *sepancuantos*. Es decir, en primer lugar existe la construcción sintáctica *sepan cuantos*, reducción de la oración

⁷ El total de ejemplos del sustantivo *esquina* en el *CREA* es de 5487 casos en 2097 documentos, por lo que dada tal cantidad de datos se ha limitado el análisis a su combinación con el nombre común *calle*, sin poder tener en consideración la posible aparición de nombres propios.

⁸ Razones de fonosintaxis favorecerían además esta evolución, al presentar la secuencia de dos vocales iguales en contacto: *esquina a*.

sepan, cuantos estas letras vieren, con su significado literal de ‘que sepan todos’. Esta expresión solía encabezar los edictos y documentos administrativos, sobre todo en español medieval y clásico, tal y como se puede apreciar en los siguientes ejemplos:

- (16) E por ende, *sepan quantos* esta carta de testamento vieren como yo doña Johana [...] fago e ordeno mi testamento a servicio de Dios [...] (CORDE: 1374, Anónimo, *Testamento de doña Juana en favor del priorato del Duero* [Cartulario de Siles]).
- (17) Yn Dei Nomine Amen, *sepan cuantos* esta carta de dote y arras vieren, como yo Andres Martínez de Heredia [...] digo que por quanto al tiempo que se trato que yo me oviese de casar por palabras de presente [...] (CORDE: 1621, Anónimo, *Carta de dote de Andrés Martínez de Heredia, familiar del Santo Oficio de la Inquisición*).

Gracias a su observación en el CORDE, es posible determinar la libertad que presentaba esta estructura, pues permitía en un principio la inserción de elementos en su interior, normalmente el cuantificador existencial *todos*⁹:

- (18) Item: *sepan todos quantos* esta carta uiren cómo yo inffante don Alfonso e cétera recebo en guarda e en comienda todos los vasallos e todas las cosas del obispo de León [...] (CORDE: 1270, Anónimo, *Confirmación de exención de impuestos* [Documentos de la catedral de León]).

La presencia de este encabezado en los documentos de carácter administrativo era tan común que incluso se hace referencia, en otro tipo de textos, como fórmula que identificaba el quehacer de los escribanos:

- (19) [...] anda el Escriven / siempre cabizbaxo, / y a decir no acierta / ni aun el “sepan cuantos” (CORDE: 1601, Gabriel Lobo Lasso de la Vega, *Manojuelo de romances*).

Su alta frecuencia de uso en los edictos provoca su reanálisis, que conlleva, a su vez, un proceso de concreción en su significado, ya que la construcción *sepan cuantos* pasa a hacer referencia, mediante un proceso de carácter metonímico, a distintas significaciones: desde el propio documento que presentaba dicha fórmula inicial, el verdugo que infligía los azotes a los condenados, hasta el propio castigo (significado con el que aparece en la mayoría de diccionarios). Estos sentidos que adquiere la estructura lexicalizada y que se desprenden de los ejemplos del corpus, empiezan a documentarse en textos de mediados del siglo XVII, época en la que comienza también a desaparecer *sepan cuantos* como fórmula que encabezaba los textos administrativos:

- (20) [...] y dando un grito el doliente, quisose levantar, por lo cual fue fuerza y mandamiento de apremio cruzarle no más de la mitad de la cara, que la otra mitad ya la tenía él cortada, [...]. Mas viéndolo en pie y con un *sepan cuantos* como mozo de golpe, y que por el rastro que dejaba podía caminar Montesinos, salime a la calle

⁹ La interpolación del cuantificador existencial *todos* se documenta en textos de finales del siglo XIII y principios del XIV (1341). Posteriormente ya no se hallan casos de esta estructura.

- [...] (CORDE: 1646, Anónimo, *La vida y hechos de Estebanillo González*). [*sepan cuantos* = ‘verdugo’].
- (21) Después que dejé la patria, / por no sé qué *sepan cuantos* / que daba yo, y el escriba / me los volvió duplicados [...]. (CORDE: 1650-1660, Conde de Rebolledo (Bernardino de Rebolledo), *Ocios*) [*sepan cuantos* = ‘edicto, documento’].
- (22) [...] mas al verle alzar repentinamente, el brazo en ademan de sacudir á aquel, “¿qué vas á hacer, Pelegrin? le dije. — Señor, me respondió, iba á arrimar un *sepancuantos* á este hombre; porque trae Chaleco azulado [...]” (CORDE: 1842, Modesto Lafuente, *Viajes de fray Gerundio por Francia, Bélgica, Holanda y orillas del Rhin*) [*sepancuantos* = ‘golpe o castigo’].

La fusión gráfica de los elementos que constituyen la formación es la prueba fehaciente de que ya es una unidad léxica plena, que se encuentra, pues, lejos de su origen oracional y, por tanto, que su proceso de lexicalización ha culminado. Resulta interesante observar cómo una fórmula legal y, por tanto, propia de registros formales, pasa a tener un significado que se asocia al registro coloquial o familiar, según indican los diccionarios académicos a partir de la edición de 1803.

En definitiva, es posible afirmar que una determinada estructura oracional puede sufrir a lo largo del tiempo una lexicalización semántica que implica una concreción de su sentido, de modo que pasamos de la designación de una acción o proceso expresado por la estructura oracional a aludir a algo más concreto, como es una persona o un objeto. Además, progresivamente, se produce también una mayor fijación de la estructura que acaba solidificándose y convirtiéndose en un compuesto léxico, que pierde la motivación sintáctica que fue su origen. De hecho, tal y como afirman Brinton & Traugott (2005: 48), “one of the commonest conceptions of lexicalization is the unification or univervation, of a syntactic phrase or construction into a single word”.

5. CONCLUSIONES

Los ejemplos anteriores de gramaticalización y lexicalización nos han permitido constatar la gran importancia de los corpus, en general, y del *CORDE* y el *CREA*, en particular, a la hora de afrontar un estudio de estos procesos en español. El avance en la observación y la explicación del cambio lingüístico en la lengua ha sido posible, desde fechas recientes, gracias al desarrollo de las teorías lingüísticas que tienen como base los fenómenos de la gramaticalización y la lexicalización. Sin embargo, un mayor conocimiento de su desarrollo sólo se ha conseguido con el análisis de un mayor número de datos históricos a partir de la existencia de nuevas herramientas informáticas. De este modo, la confección de bancos de datos textuales se ha convertido en pieza y complemento fundamental para el avance de la lingüística histórica. Así, el análisis de la transformación del sustantivo *esquina* a una locución preposicional como *esquina a*, así como la de la frase *sepan cuantos* a una nueva pieza léxica como *sepancuantos* son, simplemente, una nueva muestra de los frágiles límites entre sintaxis, morfología y léxico, y de cómo todos éstos se hallan a expensas de la evolución lingüística. En definitiva, la gramaticalización y la lexicalización, en este caso, pueden ayudarnos enormemente a desentrañar el proceso de cambio y los corpus a evidenciarlo.

REFERENCIAS BIBLIOGRÁFICAS

- ALCOBA, S. (1988): “Categoría léxica de las palabras compuestas”. *Verba* 15, 109-146.
- BRINTON, L. J. & E. C. TRAUGOTT (2005): *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- BRIZ, A. & M. ALBELDA (2009): “Estado actual de los corpus de lengua española hablada y escrita: I+D”. En *El español en el mundo. Anuario del Instituto Cervantes 2009*. Madrid: Instituto Cervantes.
- BUSTOS GIBERT, E. de (1986): *La composición nominal en español*. Salamanca: Universidad de Salamanca.
- CIFUENTES HONRUBIA, J. L. (2003): *Locuciones prepositivas. Sobre la gramaticalización preposicional en español*. Alicante: Universidad de Alicante.
- CODEA = SÁNCHEZ-PRIETO BORJA, P. (dir.): *Corpus de Documentos Españoles Anteriores a 1700*. <<http://demos.bitext.com/codea/>>.
- COMPANY COMPANY, C. (2003): “La gramaticalización en la historia del español”. *Medievalia* 35, 3-61.
- CORDE = REAL ACADEMIA ESPAÑOLA: Banco de datos [en línea]. *Corpus diacrónico del español*. <<http://www.rae.es>> [2012].
- CREA = REAL ACADEMIA ESPAÑOLA: Banco de datos [en línea]. *Corpus de referencia del español actual*. <<http://www.rae.es>> [2012].
- Davies = DAVIES, M. (2002-): *Corpus del Español: 100 million words, 1200s-1900s*. <<http://www.corpusdelespanol.org>>
- ELVIRA, J. (2006): “Aproximación al concepto de lexicalización”. En J. RODRÍGUEZ MOLINA & D. M. SÁEZ DE RIVERA (coords.): *Diacronía, lengua española y lingüística*. Madrid: Síntesis, 21-41.
- ENRIQUE-ARIAS, A. (dir.): *Biblia Medieval*. <<http://www.bibliamedieval.es>>.
- HEINE, B., U. CLAUDI & F. HÜNNEMEYER (1991): *Grammaticalization. A Conceptual Framework*. Chicago: The University of Chicago Press.
- KURYŁOWICZ, J. (1965): “The evolution of grammatical categories”. *Enquises Linguistiques* II, 38-45.
- LEHMANN, C. (1985): “Grammaticalization: Synchronic Variation and Diachronic Change”. *Lingua e Stile* 20/3, 303-318.
- LEHMANN, C. (2003): “New reflections on grammaticalization and lexicalization”. En I. WISCHER & G. DIEWALD (eds.): *New reflections on grammaticalization*. Amsterdam: John Benjamins, 1-18.
- MADROÑAL, A. (2003): “El Corpus Diacrónico del Español (CORDE), de la Real Academia Española”. *Letras de Deusto* 33/100, 163-181.
- MEILLET, A. (1912): “L'évolution des formes grammaticales”. *Linguistique Historique et Linguistique Générale*. Paris: Champion, 130-148.
- MORENO CABRERA, J. C. (1998): “On the relationships between grammaticalization and lexicalization”. En A. GIACALONE RAMAT & P. HOPPER (eds.): *The Limits of grammaticalization*. Amsterdam / Philadelphia: John Benjamins, 209-227.
- NGLE = REAL ACADEMIA ESPAÑOLA & ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2009): *Nueva gramática de la lengua española*. Madrid: Espasa Libros.
- PASCUAL, J. A. & C. DOMÍNGUEZ (2009): “Un corpus para un Nuevo diccionario histórico del español”. En A. ENRIQUE-ARIAS (ed.): *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*. Madrid: Iberoamericana / Frankfurt: Vervuert, 79-93.
- PENA, J. (1991): “La palabra: estructura y procesos morfológicos”. *Verba* 18, 69-128.

- ROJO, G. (2008): “Lingüística de corpus y lingüística del español”, ponencia plenaria en el *XV congreso de la Asociación de Lingüística y Filología de América Latina* (Montevideo, 18-21 de agosto de 2008). Montevideo, 2008. Edición en CD. [ISBN 978-9974-8002-6-7]
- ROJO, G. (2010): “Sobre codificación y explotación de corpus textuales: otra comparación del *Corpus del español* con el *CORDE* y el *CREA*”. *Lingüística* 24, 11-50.
- ROJO, G. (en prensa): “El papel de los corpus en el estudio de la historia del español”, intervención en la mesa redonda sobre *Los corpus diacrónicos en la historia de la lengua española* celebrada en el *VIII Congreso Internacional de Historia de la Lengua Española*, Universidad de Santiago de Compostela (14-18 de septiembre de 2009).
- SÁNCHEZ SÁNCHEZ, M. & C. DOMÍNGUEZ (2007): “El banco de datos de la Real Academia Española: *CREA* y *CORDE*”. *Per Abbat: boletín filológico de actualización académica y didáctica* 2, 137-148.