

Lingüística de corpus y análisis cuantitativo en la gramática (histórica) del español: el caso de las oraciones completivas de sustantivo

Anton Granvik

Universidad de Gotemburgo/Universidad de Cádiz

Resumen

El uso de (grandes) corpus textuales como base empírica para el análisis de fenómenos gramaticales ocupa un lugar central dentro de la lingüística contemporánea. La gramática histórica del español no es ninguna excepción, y desde principios del presente milenio los historiadores de la lengua disponen de dos grandes corpus diacrónicos ampliamente usados en el mundo entero, como son el CORDE de la Real Academia Española y el *Corpus del español* de Mark Davies (2002-). Al lado de los grandes corpus muestras de textos de menor extensión, pero con características relevantes para la investigación en cuestión, también se utilizan como base de análisis empíricos.

En conexión con el desarrollo tecnológico que ha permitido la creación de corpus cada vez más extensos, los métodos de análisis cuantitativos también han entrado en el campo de la lingüística descriptiva y funcional, donde, a día de hoy, constituye un área en fuerte expansión y con un potencial importante debido tanto a la coyuntura de los estudios lingüísticos (descripción gramatical basada en el uso) como al desarrollo y accesibilidad a métodos de análisis lo suficientemente sofisticados como para dar cuenta de fenómenos tan complejos como son los datos lingüísticos. Como ejemplo puede mencionarse el hecho de que muchos análisis cuantitativos se basan en asunciones como que los datos están distribuidos según la función gaussiana y que son arbitrarios, algo que muchas veces no se puede establecer para los datos lingüísticos (Kilgariff 2005).

La idea de mi conferencia ha sido presentar algunos análisis básicos que, por un lado, demuestran la importancia de introducir conceptos cuantitativos más allá de la mera frecuencia (relativa) de uso, y por otro, pueden resultar útiles para la descripción y la explicación de un gran número de fenómenos gramaticales. El primer ejemplo de ello ha sido el llamado Análisis colostrucciona (ACol, *Collostructional analysis*) desarrollado por Stefanowitsch & Gries (2003). Este análisis combina las nociones de **colocación** y **construcción**, enfocando la relación entre una construcción (gramatical) y un elemento constituyente, por ejemplo, las oraciones completivas de sustantivo del tipo N (*de que* o N *de* + infinitivo, donde el elemento investigado son los diferentes sustantivos que pueden ocupar el lugar del elemento N).

1. La **idea de ir** a Las Vegas me espanta.
2. La **posibilidad de que** vengas me encanta.

El ACol lo que hace es partir de cuatro valores de frecuencia recuperables de un corpus y sobre la base de estos producir un valor que describe la fuerza de asociación que existe entre los diferentes sustantivos que ocurren en la posición de N. La comparación de los valores para diferentes sustantivos permite ordenarlos según la fuerza de asociación, con lo cual se identifican los sustantivos más típicos de la construcción.

El Análisis distintivo de colexemas (ADC, *Distinctive collexeme analysis*), desarrollado por los mismo profesores alemanes Gries & Stefanowitsch (2004), supone una ampliación del ACol, pues permite contrastar dos construcciones entre sí sobre la base de sus colexemas, o

colocados. En el caso de las oraciones completivas de sustantivo, por ejemplo, usando la frecuencia de uso de un sustantivo dado, N_1 , con complementos infinitivos (N *de* + infinitivo) y finitos (N *de que*), y comparándola con la frecuencia de los demás usos de las dos construcciones, respectivamente, se adquiere un valor de probabilidades que muestra una preferencia por uno u otro tipo de completiva. Así, puede determinarse qué sustantivos, entre los más frecuentes de ambas construcciones, prefieren las completivas con *de que* y cuáles prefieren el infinitivo.

Evidentemente, aplicando estos análisis sobre la diacronía, pueden realizarse sobre muestras de diferentes periodos de tiempo, y los resultados pueden compararse para hacerse una idea de los cambios y tendencias que experimentan las diferentes variantes de las oraciones completivas de sustantivo.

Referencias bibliográficas

CORDE = Real Academia Española: Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español*. <<http://www.rae.es>> [enero de 2016]

Davies, Mark. 2002-. Corpus del Español: 100 million words, 1200s-1900s. En línea: <http://www.corpusdelespanol.org>.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9:1, 97-129.

Kilgarriff, Adam. 2005. Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory* 1-2, 263-275,

Stefanowitsch, Anatol & Stefan Th. Gries 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209-243.