

CHAPTER XII

Relacionando los análisis cualitativo y cuantitativo. Una propuesta de modelo estadístico predictivo para completar la descripción compleja de los verbos cognitivos^{1 2}

Relating qualitative and quantitative analysis. A predictive statistical model proposal to complete the complex description of cognitive verbs

M. Amparo Soler Bonafont
Universidad Complutense de Madrid – España

Resumen: El objetivo del presente capítulo es realizar una propuesta de descripción de los usos semántico-pragmáticos de unas formas verbales complejas como son las formas performativas de los verbos cognitivos, concretamente, de su forma paradigmática *creo*, en la interacción oral. Para ello se lleva a cabo una aproximación cognitiva con base en una propuesta estadística predictiva, creada a partir de un sistema de regresiones multinomiales (con la herramienta STATA). Se persigue que el modelo diseñado permita reconocer con un elevado grado de explicatividad ante qué

¹ Este capítulo profundiza en algunos de los resultados parciales que son fruto de la tesis doctoral de la autora (Soler, 2019), así como de la ampliación que de ellos se realiza en Soler (2021b).

² La investigación se enmarca en el proyecto *Los procesos de gestión de la imagen y la descortesía: perspectivas históricas, lingüísticas y discursivas*, concretamente, en la subdivisión de análisis de procesos discursivos (ref. PID2019-107668GB-I00, Ministerio de Ciencia e Innovación, Gobierno de España).

significados y funciones pragmáticas de unidades polisémicas y polifuncionales como *creo* nos encontramos, una vez sistematizadas las principales circunstancias de aparición cualitativas que las rodean. El estudio de formas se da en un corpus compilado de conversaciones coloquiales y de discursos de debate parlamentario. Como resultado, se obtiene un modelo de análisis predictivo replicable en otros textos y géneros en los que pueden aparecer unidades epistémicas similares.

Abstract: The goal of this chapter is to bring a description proposal for the semantic and pragmatic uses of some complex verbal forms like the performative forms of cognitive verbs, specifically *creo*, in oral interaction. For this purpose, it is carried out a cognitive approach based on a predictive statistical pattern created with a multinomial regressions system (through STATA tool). It is intended that the designed model allows the researcher to recognize what senses and pragmatic functions is dealing with in so polysemic and polyfunctional units as *creo*, with a high degree of explanation, once the main circumstances of its qualitative appearances are systematized. The study of forms is done in a compiled corpus of colloquial conversations and parliamentary debates. As a result, it is obtained a predictive model of analysis which is replicable in other texts and genres in which some similar epistemic units can appear.

1. Introducción

1.1. Dificultades descriptivas en el grupo de las formas performativas de los verbos cognitivos

La explicación funcional del uso semántico-pragmático de algunas unidades epistémicas en los textos resulta aún hoy, y tras siglos de disquisiciones, compleja. Tal es el caso, reveladamente frecuente en la oralidad, de algunas formas verbales de primera persona del singular del presente de indicativo: *creo, pienso...*, también conocidas como formas performativas de los verbos cognitivos o de opinión (Fetzer y Johansson, 2010; Fetzer, 2014; González Ruiz, 2015; Soler, 2018). Estas formas verbales son subjetivas y, en algunas ocasiones, pueden manifestarse de manera integrada (*creo que* + verbo) o parentética (uso de *creo* con movilidad posicional), desde el punto de vista morfosintáctico. No obstante, estas características que las identifican no son tan llamativas como otros de sus rasgos definitorios, los cuales dificultan su reconocimiento: estos son su polisemia y su polifuncionalidad anunciadas (Hartwell *et al.*, 2017; Jansegers, 2017; Soler, 2019).

Los diferentes significados y funciones que pueden manifestar unidades como *creo*, la forma paradigmática de este conjunto por tratarse de la más compleja y la más polisémica y polifuncional de su clase (Soler, 2019), han sido estudiadas en diferentes géneros (tanto en español como en otras lenguas), entre los que destacan los de interacción oral, especialmente, la conversación y el debate parlamentario. Así bien, incluso en este tipo de géneros, *creo* y formas verbales semejantes a esta manifiestan desde funciones atenuantes hasta intensificadoras (Cutting, 2007; Fuentes Rodríguez, 2010, 2016; De Hoop *et al.*, 2018), a la vez que despliegan una gran variedad de valores semánticos, desde la creencia hasta el juicio (Soler, 2019, 2021). Distinguir la multiplicidad de sus posibilidades semántico-pragmáticas no es tarea sencilla para el lingüista, que se encuentra, desde hace más de un siglo con un escollo adicional en estos verbos: la limitación de las herramientas lingüísticas tradicionales para el estudio de fenómenos complejos como el citado. Los pragmatistas se preguntan cómo definir los significados y significados en uso de unidades subjetivas como las que son objeto de este trabajo, para los que no son suficientemente explicativas las pruebas veritativo-condicionales ni las de la pragmática clásica. Por estos motivos, son cada vez más numerosos los estudios que realizan una aproximación cognitiva a estas formas, gracias a su concepción de la semántica y de la pragmática como un mero *continuum* (Achard, 1998; Buceta, 2014; Jansegers, 2017; Jansegers y Gries, 2017; Boas y Ziem, 2018), lo que ayuda a superar algunos obstáculos definitorios.

No obstante, y de acuerdo con diferentes estudios pragmáticos y sociolingüísticos recientes (Díaz-Campos y Gradoville, 2011; González *et al.*, 2014), la explicación cualitativa cognitiva queda incompleta si no se realiza un análisis riguroso de corpus, de tipo cuantitativo (Roldán 2005; Abdulrahim, 2014; Milin *et al.*, 2016). Dicha incompletud se observa en la falta de diseños metodológicos cuantitativos capaces de dar una respuesta procedimental adecuada a la hora de operar ante estos casos, así como en la falta de homogeneidad ante la descripción tanto de unos valores semánticos cerrados de *creo*, como de las funciones concretas que puede desarrollar.

1.2. Planteamiento de este trabajo

El objetivo de esta investigación, una vez enunciadas algunas de las fallas metodológicas en el estudio de fenómenos lingüísticos semántica y pragmáticamente complejos, es tratar de llevar a cabo una descripción sistemática del funcionamiento de las formas performativas de verbos cognitivos como *creo* en la interacción oral. Para ello, este trabajo combina el análisis cualitativo de corte cognitivo y el análisis cuantitativo, en lo que se propone como una propuesta predictiva de reconocimiento de valores de *creo*. Se expone un modelo diseñado

mediante distintas regresiones multinomiales de variables cualitativas de análisis (elaboradas con una base cognitiva), las cuales se realizan a través de herramientas como STATA. Con este modelo se consigue reconocer con un elevado grado de explicatividad ante qué significados y funciones pragmáticas de la unidad objeto de estudio podemos encontrarlos, una vez sistematizadas las principales circunstancias de aparición que las rodean.

Las formas analizadas como *tokens* se han extraído de un corpus constituido por dos géneros discursivos de tipo interactivo, en el español de España, de los últimos 20 años: conversaciones coloquiales (de los corpus disponibles *COGILA*, *COJEM*, *Val.Es.Co. 2002* y *Val.Es.Co. 2.0*) y discursos de debate parlamentario (pertenecientes al archivo del *Congreso de los Diputados* del Gobierno de España y de *Les Corts Valencianes* y accesible en línea). Sobre los datos obtenidos, se han aplicado las bases de la estadística descriptiva y predictiva, como también se ha realizado en otros trabajos de corte lingüístico previos (Abbhul y Mackey 2013, James *et al.*, 2013). En definitiva, se obtiene un diseño predictivo propio, el cual es replicable en otro tipo de textos y géneros textuales susceptibles de contener unidades epistémicas de funcionamiento similar al de las formas performativas de los verbos cognitivos.

2. Acercamiento teórico a *creo* y otras formas performativas de los verbos cognitivos

Una de las grandes preocupaciones de los filósofos del lenguaje desde el siglo XIX (desde Frege o Russell, hasta Kripke o Richard), y que ha perdurado en la lingüística aún hasta nuestros días, es la de la descripción de aquellas unidades subjetivas cuyo valor de significado no puede ser suficientemente explicado desde la semántica, pero para las que la pragmática tampoco puede ofrecer una solución aislada. Tenemos un claro ejemplo en las formas performativas, esto es, aquellos verbos en primera persona del singular del presente de indicativo, y que son de carácter cognitivos. Son casos como *considero*, *creo*, *opino*, *pienso*, *supongo*..., con los que no solo se hace evidente el *origo*, la presencia del hablante en la escena en que se produce lo dicho, sino que se observa que el significado del referente viene enriquecido con aspectos intencionales que le superpone el hablante gracias a una doble posibilidad de lectura: proposicional y también extraproposicional. Esta naturaleza hace ver, pues, que unidades como las comentadas se encuentran en el límite mismo entre la semántica y la pragmática.

Disquisiciones aparte, en este trabajo abordamos la problemática concreta que ofrece una forma verbal paradigmática: *creo*, por ser considerada esta la más compleja del conjunto de las formas performativas de los verbos cognitivos. Con su estudio pueden verse resumidas cuestiones que atañen al resto de unidades de primera persona del singular de

estos verbos, que funcionan de modo semejante, y cuyas dificultades definitorias (si no todas, sí muchas de ellas) pueden verse subsumidas en las que aquí planteamos para *creo*.

2.1. Polisemia de *creo*

Creo es considerada una forma verbal con un valor altamente subjetivizador de lo dicho (Soler, 2019). Es la forma performativa del verbo *creer*, el cual se caracteriza por ser polisémico, si bien esta polisemia no había sido aclarada hasta los últimos años. El reciente interés investigador por esta polisemia ha cristalizado en el reconocimiento de una alta complejidad cifrada en el conjunto de varios aspectos: su polimorfismo construccional (*creer en, creer que, no creer...*) (Buceta, 2014; Soler, 2019), la multiplicidad de contextos de aparición (conversación coloquial, entrevistas políticas, debates, etc.) (Fetzer, 2014; Fetzer & Johansson, 2010; González Ruiz, 2015; Soler, 2018), la frecuencia de un fuerte componente argumentativo en su contexto próximo (Fuentes Rodríguez, 2010, 2016), y la diferente variedad funcional, incluso complementaria (desde la atenuación a la intensificación, pasando por la neutralidad), que puede manifestar (González Ruiz, 2015; Soler, 2019).

De todo ello se desprende que *creo*, la forma más peculiar de su paradigma morfológico, supone un escollo para la investigación, que si bien ha observado las causas de su complejidad, no había conseguido dar hasta la fecha con una descripción consistente de sus usos. Y es que la bibliografía se ha tratado de acercar repetidamente a sus significados, los cuales fluctuaban entre dos y seis valores, sin que pudiera haber acuerdo, sino solo un resumen tradicional de los valores primordialmente en dos: el epistémico o débil y el de opinión o fuerte (Fetzer, 2014; Fetzer & Johansson, 2010; González Ruiz, 2015). Estos dos significados polares se resumen en los siguientes ejemplos:

B: pero ¿qué es/¿que ya lo has dejado oo?

A: **creo que** ya lo he dejado un poco por imposible (*valor epistémico o débil*)

B: yo **creo que** tienes que insistir (*valor de opinión o fuerte*)

Puede observarse que el valor débil de *creo* presenta a modo de duda y no de una convicción lo dicho por A, y expresa que el hablante puede no disponer de pruebas para manifestar lo dicho con mayor grado de seguridad. Por su lado, el valor de opinión se corresponde con la expresión de un juicio personal, independientemente de las pruebas de las que se disponga sobre lo aseverado. En ambos casos está presente la subjetividad, pero esta pone su foco en diferentes aspectos (bien en las pruebas de las que se dispone sobre ello, bien en la confianza de que lo dicho sea de tal o cual modo), incluso con el uso de una misma construcción formal.

La distinción básica revisada puede resultar viable en un primer momento, pero no lo es si nos encontramos ante casos como los que siguen: *creo que tu papi va a jugar con el barquito más que tú* (en que además del grado de seguridad, también podríamos hablar de opinión); *hospital de la Vega Baja, hospital –creo recordar– de Elda...* (en que tenemos construcciones de doble acusativo, muy características, en las que tampoco es fácil discernir ante qué valor nos encontramos); *eso es lo que creo* (en que una nueva construcción encapsulada en función de atributo parece estar acercándose más al valor de certeza que al de duda o al de opinión), etc. Con ello, vemos que la polisemia debe abordarse desde un criterio efectivo, que ordene los semas de cada valor de manera rigurosa para poder reconocer límites entre ellos, que supere la diversidad de descripciones bibliográficas y que, de acuerdo con lo visto, evidencie los puntos de conexión con las diferentes construcciones formales del verbo. Asimismo, y como persigue este capítulo, se espera que la categorización obtenida se acompañe de un criterio de reconocimiento sencillo y viable para el analista.

2.2. Polifuncionalidad de *creo*

La polifuncionalidad, no solo de *creo*, sino también de otras unidades de su mismo conjunto de formas performativas, subjetivas y cognitivas, viene de la mano de su reconocida polisemia. Como hemos avanzado, en usos como los de *creo* se han reconocido tradicionalmente funciones de atenuación (*creo que ya lo he dejado un poco por imposible*, Val.Es.Co. 2002), neutralidad (*hospital de la Vega Baja, hospital –creo recordar– de Elda*, Les Corts Valencianes), e incluso intensificación (A: es que los mayores↑ además a mí seguro que se me comen (RISAS)/ tienes que tener un SEXTO= // B: NO↓ **yo creo que** exige más↑, Valesco 2.0, C. 1, 68-69).

Diversos estudios monográficos previos que han versado sobre el objeto de estudio de este capítulo se han preguntado si existe una correlación entre los significados reconocidos y las funciones pragmáticas de *creo*. La bibliografía ha llegado a establecer una correlación casi directa entre el valor débil y la atenuación, por un lado, y el valor de opinión, y la intensificación (Fuentes Rodríguez, 2016; González Ruiz, 2015), por otro lado. Estas correlaciones establecidas de forma automática y asumidas por la comunidad científica llevan, no obstante, a arrastrar varios errores conceptuales básicos. Así, por ejemplo, cabe destacar que los estudios de corpus realizados hasta la fecha no aportan una amplitud suficiente de datos basados en corpus de lengua real ni cotejan las observaciones con pruebas objetivas y replicables a partir de las que puedan ofrecerse resultados concluyentes, con lo cuales pudiera confirmarse dicha automaticidad de relaciones semántico-pragmáticas de *creo*. Asimismo, en los estudios se observa una ausencia de criterio para la detección de otros posibles significados, o funciones, distintos a los básicos, ya comentados. Prueba de ello es

que la neutralidad suele quedar fuera de los análisis, pese a que algunos investigadores han llegado a reconocer esta función en casos aislados, o incluso que la atención a las diferentes manifestaciones formales de *creo* y sus repercusiones a nivel semántico y pragmático suelen estar ausentes en las investigaciones. Serán estos aspectos los que tratará de solventar este capítulo con la propuesta de un modelo de análisis concreto, que se presenta como replicable también para otros análisis de unidades doxásticas complejas.

3. Exploración de un análisis cognitivo experimental

La búsqueda de metodologías de análisis, si no alternativas, sí complementarias a las explicaciones cualitativas de la semántica tradicional, ha llevado a la comunidad científica a explorar enfoques integradores, como es el caso de la lingüística cognitiva. Abdulrahim 2014; Fetzer y Johansson 2010; Jansegers 2017; Jansegers y Gries 2017; Milin *et al.*, 2016; o Roldán 2005 son algunos de los casos de análisis semántico-pragmáticos de tipo cognitivo combinados con estadística. Este marco teórico entiende la semántica y la pragmática como un continuo, lo cual ha facilitado la comprensión de formas como *creo* desde este paradigma, como prueban dichos estudios. Gracias a este enfoque, la observación cualitativa del analista no se ve anulada, sino que es, además de reconocida, apoyada en datos reales y comprobables. Se trata, por tanto, del motivo por el que el enfoque cognitivo está tomando cada vez más auge en los últimos años. Asimismo, el acercamiento estadístico predictivo y experimental también se ha visto incrementado recientemente en distintos trabajos lingüísticos, cognitivos, e incluso funcionales y sociolingüísticos (Boas & Ziem, 2018; Díaz-Campos & Gradoville, 2011), en los cuales, como planteamos en este trabajo, un sistema de análisis cuantitativo riguroso completa adecuada y necesariamente la aproximación cualitativa.

4. Metodología del estudio

De acuerdo con lo expuesto, la hipótesis de partida que planteamos es que debe de existir la posibilidad de realizar un cálculo aproximado de los valores semánticos y pragmáticos que manifiestan formas performativas como *creo*, si el inventario de categorías (significados y funciones) que se les reconoce es cerrado³. Por esta razón, la pregunta de investigación a la

3 Desde el punto de vista de la Semántica Cognoscitiva, los significados de *creo*, así como los de otras palabras polisémicas, pueden concebirse como continuos y ordenables a partir de la ganancia o pérdida de algunos semas. Ahora bien, lo que esta concepción del significado conlleva es que existan valores básicos, prototípicos, desde los que derivan extensiones significativas. Luego los significados nucleares sí que componen inventarios cerrados y, por consiguiente, pueden ser estudiados de una manera más sistemática que si el investigador se enfrentara a toda la polisemia de elaboraciones y extensiones semánticas en su conjunto que puede generar una palabra.

que se pretende responder es qué método, complementario al análisis cualitativo, puede permitir una descripción más amplia y certera de los usos de unidades complejas como *creo*.

Este capítulo se propone, por consiguiente, aplicar un análisis de *creo*, como forma paradigmática del conjunto de unidades performativas complejas de los verbos cognitivos, desde el paradigma del cognitivismo, el cual ha resultado eficaz para la descripción de otras formas lingüísticas (adverbiales y verbales) de funcionamiento semejante a la que es objeto de estudio (Abdulahim, 2014; Fetzer y Johansson, 2010; Jansegers, 2017; Jansegers y Gries, 2017; Milin *et al.*, 2016; Roldán, 2005), y probar su operatividad. Asimismo, se quiere determinar qué parámetros afectan en el proceso de detección de la semántica y la pragmática de la forma verbal para establecer un protocolo jerárquico de las características observables y que, a partir de estas, pueda certificarse un alto grado de reconocimiento del significado y de la función pragmática de *creo*.

Con este fin, planteamos una metodología de análisis de corpus. Se compila un conjunto de textos disponibles de interacción oral de diferentes géneros discursivos: conversación coloquial y debate parlamentario. Son estos dos los formatos en los que más se ha estudiado hasta la fecha el comportamiento de los verbos cognitivos, tanto en el caso del español como en otras lenguas. Asimismo, se trata de géneros que suponen puntos opuestos de un continuo tanto de formalidad como de otros rasgos como dialogicidad, grado de planificación y determinación en el reparto de los turnos de los participantes, lo que permite obtener un espectro ancho de circunstancias de la oralidad adecuadas para realizar un estudio general de tendencias de uso de *creo* lo más amplio posible. La compilación la conforman textos de conversaciones coloquiales de los corpus *COGILA*, *COJEM*, *Val.Es.Co. 2002* y *Val.Es.Co. 2.0*; y sesiones de debate parlamentario del *Congreso de los Diputados* (del Gobierno de España) y de *Les Corts Valencianes* (del gobierno autonómico de la Comunitat Valenciana), en una proporción equitativa. En el caso de los corpus conversacionales, se analizan en su totalidad el *COGILA* (36 000 palabras); el *COJEM* (100 000 palabras); *Val.Es.Co. 2002* (91 366 palabras); y *Val.Es.Co. 2.0* (128 394 palabras). De los corpus parlamentarios se obtiene, de manera aleatoria, una muestra de una cantidad similar de palabras, repartida esta entre las dos fuentes: *Congreso de los Diputados*, 177 522 palabras; *Les Corts Valencianes*, 174 366. La siguiente tabla resume esta base de la muestra:

Realizado el análisis cualitativo con la observación de los aspectos cifrados en las variables previas sobre el total de los 730 casos obtenidos, pasamos a realizar el análisis cuantitativo principal que este trabajo presenta. Este consiste en la aplicación de una estadística exploratoria (mediante tablas de contingencia comunes) que permite discriminar algunos datos básicos (ej. la determinación de algunos resultados semánticos, a partir de algunos aspectos formales de las manifestaciones del verbo). Tras ello, se propone un modelo de análisis de estadística descriptivo-predictiva basado en un protocolo de tres pasos: 1. regresiones logísticas, 2. obtención de valores de verosimilitud de cruces de las variables en la determinación del grado de explicación sobre la semántica y sobre la pragmática de *creo*, y 3. cálculo de errores. Todos estos cálculos se realizan en una programación experimental de 1 000 iteraciones, mediante el programa STATA. Ahora bien, para poder aplicar las pruebas estadísticas, se crea un corpus ampliado en el que se aumentan los datos hasta llegar a un mínimo de 5 casos por cada variante de las contenidas por variable aplicada (ya que se trata del número mínimo de casos para que los que las pruebas estadísticas pueden arrojar resultados significativos). Estos ejemplos se obtienen de los corpus *COLAm* y *CORPES XXI*, para el caso de la conversación coloquial, y de otras sesiones no consultadas de las mismas fuentes parlamentarias, para el caso del debate. La Tabla 2 resume los datos de *creo* extraídos del corpus ampliado (un total de 865 casos), sobre los que se aplica el protocolo de análisis, frente a los del corpus base (730 ejemplos).

Tabla 2. Datos de los corpus *base* y *ampliado*, para el análisis estadístico predictivo significativo

	Corpus base	Corpus ampliado
Ocurrencias de <i>creo</i>	730	865

Las regresiones logísticas que planteamos para este análisis son de tipo multinomial. Las regresiones son un cálculo predictor sobre la incidencia de una variable dependiente (Y) sobre una independiente (X). El valor de la regresión ($Y \approx \beta_0 + \beta_1 X$) permite obtener un coeficiente que cifra la estimación de los valores, el cual se denomina R^2 . Ahora bien, cabe destacar que esta prueba estadística presupone linealidad entre las variables. Dado que esta no se da entre aspectos cualitativos de análisis lingüísticos como el que presentamos y, por consiguiente, el cálculo obtenido en el primer paso no es exacto, en un segundo paso o instancia calculamos complementariamente un número de verosimilitud de la relación entre las variables cotejadas. Lo hacemos a partir del modelo de McFadden, el cual permite obtener, frente a la estimación de valor de R^2 , un valor probabilístico de pseudo- R^2 . Este valor permitirá ordenar jerárquicamente las variables preestablecidas de mayor a menor grado de expli-

cación sobre el valor semántico de *creo*, por un lado, y sobre el valor pragmático, por otro. Finalmente, como este cálculo no es exacto y se realiza sobre 1 000 repeticiones del experimento, se calculan posibles errores a partir de la creación de dos variables: la máxima probabilidad de acierto del resultado y la mínima probabilidad de esta. Estas también se entrecruzan con las previas para obtener las diferencias y el margen de error.

En lo que sigue, se verán los resultados obtenidos de la aplicación de este modelo de análisis. Asimismo, se comprobará su viabilidad como metodología replicable.

5. Análisis y discusión de los resultados

El análisis efectuado sobre la semántica y la pragmática de *creo* ha ofrecido resultados en diferentes planos. En lo que sigue, presentamos los obtenidos en cada fase del estudio, y un resumen del modelo metodológico aplicado, el cual puede considerarse también como un resultado de la investigación.

5.1. Fases del análisis

El primer resultado que ofrece el acercamiento cognitivo a los usos discursivos de *creo* ha permitido reconocer cinco valores semánticos básicos: *creencia*, *certeza*, *conjetura*, *predicción* y *juicio*, de acuerdo con lo apuntado en estudios previos (Soler, 2018; 2019)⁴. Estos valores se ordenan en un continuo de subjetividad, según el grado de implicación del hablante en la escena que proyecta. Nuestro estudio estadístico descriptivo del corpus base, así como del corpus ampliado, permite ver que la construcción formal de *creo* determina en el 100 % de los casos alguno de estos cinco valores. Véanse las tablas de contingencia 3 y 4:

4 Si bien la descripción de los significados de *creo* excede los objetivos de este capítulo (véase, para ello, Soler 2021), describimos mínimamente los semas básicos de cada uno de ellos para aclarar su lectura. El valor de *creencia* describe la adhesión completa del hablante a lo dicho, con independencia de las pruebas que se tengan para ello (ej. *creo en dios*). El valor de *certeza* describe verdades que son absolutas únicamente para el propio hablante, el cual también las presenta como independientes de su comprobación (ej. *me lo creo*). El valor de *conjetura* hace referencia a un cálculo realizado por el hablante cuando este dispone de algunas pruebas sobre lo dicho (ej. *creo que fue ayer*). La *predicción*, como la *conjetura*, se basa en algunas pruebas, pero se proyecta sobre hechos futuros (ej. *creo que viene mañana*). Por último, el *juicio* manifiesta una opinión personal, basada en la comprobación de lo dicho, que ahora no es factual, sino que se basa en la única escala de valores que son los personales del hablante (ej. *creo que eso no está bien*).

Tabla 3. Cruce de datos obtenidos entre la construcción y el valor semántico de *creo*, con prevalencia del valor semántico (Soler, 2019).

Variable formal (1): Construcción						
24. SIGNIFICADO	creencia	certeza	conoci- miento	posibili- dad	juicio	intr. im- preciso
no creo en	100 %	0 %	0 %	0 %	0 %	0 %
(no) me (lo) creo (X)	0 %	38,46%	0 %	0 %	0 %	0 %
ya lo creo (X)	0 %	38,46%	0 %	0 %	0 %	0 %
(no) lo creo	0 %	21,15%	0 %	7,89%	0 %	0 %
creo	0 %	0 %	11,67%	0 %	4,79%	0 %
no creo	0 %	0 %	0 %	17,54%	0 %	0 %
creo que	0 %	0 %	59,17%	34,21%	75,80%	100 %
no/tampoco creo que	0 %	0 %	0,00%	35,09%	0,23%	0 %
sí/también creo que	0 %	0 %	1,25%	0 %	3,88%	0 %
creo que no/Æ o verbo	0 %	0 %	7,92%	5,26%	4,57%	0 %
creo que sí/también + Æ o verbo	0 %	1,92%	9,58%	0 %	2,74%	0 %
X + creo + PVO del OD/ pron. + creo + CC	0 %	0 %	0 %	0 %	4,57%	0 %
creo + infinitivo	0 %	0 %	8,33%	0 %	0 %	0 %
lo que creo + Æ o verbo	0 %	0 %	2,08%	0 %	3,42%	0 %

Tabla 4. Cruce de datos obtenidos entre la construcción y el valor semántico de *creo*, con prevalencia de la construcción (Soler, 2019).

(1) CONSTRUCCIÓN						
24. SIGNIFICADO	creencia	certeza	conoci- miento	posibili- dad	juicio	intr. im- preciso
(no) creo en	100 %	0,00%	0 %	0 %	0 %	0 %
(no) me (lo) creo (X)	0 %	100 %	0 %	0 %	0 %	0 %
ya lo creo (X)	0 %	100 %	0 %	0 %	0 %	0 %
(no) lo creo	0 %	55 %	0 %	45 %	0 %	0 %
creo	0 %	0 %	57,14%	0 %	42,86%	0 %
no creo	0 %	0 %	0 %	100 %	0 %	0 %
creo que	0 %	0 %	27,63%	7,59%	64,59%	0,19%
no/tampoco creo que	0 %	0 %	0 %	97,56%	2,44%	0 %
sí/también creo que	0 %	0 %	15 %	0 %	85 %	0 %
creo que no/Æ o verbo	0 %	0 %	42,22%	13,33%	44,44%	0 %
creo que sí/también + Æ o verbo	0 %	2,78%	63,89%	0 %	33,33%	0 %
X + creo + PVO del OD/ pron. + creo + CC	0 %	0 %	0 %	0 %	100 %	0 %
creo + infinitivo	0 %	0 %	100 %	0 %	0 %	0 %
lo que creo + Æ o verbo	0 %	0 %	25 %	0 %	75 %	0 %

Como puede observarse, tanto el valor de *creencia* ((*no*) *creo en*) como el de *certeza* ((*no*) *me (lo) creo (X)* o *ya lo creo (X)*) vienen determinados en el 100 % de los casos por una construcción concreta de *creo*. Asimismo, en la totalidad de los casos analizados en los que aparece una construcción concreta de *creo*, el valor semántico reconocido es el mismo, si bien esta relación no se da ahora siempre en el sentido inverso. Se trata de *creo* + *infinitivo*, que conlleva el valor de *conjetura*; y *no creo*, que expresa predicción; y *X* + *creo* + *PVO del OD/ pron.* + *creo* + *CC*, asociada al *juicio*. De ello se desprende que la aproximación cognitiva es eficaz, y que la estadística descriptiva ofrece una prueba patente de ello, pues certifica la viabilidad de las pruebas para discernir algunas de las relaciones de variables determinantes en el reconocimiento, en este caso, del valor semántico de *creo*. No obstante, no es determinante para el reconocimiento de su pragmática, ni explica todos los valores semánticos que ha distinguido el enfoque cognitivo aplicado. Por consiguiente, en una segunda fase del estudio, se aplica la estadística predictiva al corpus ampliado, con el fin de alcanzar resultados más concretos.

Implementamos la metodología diseñada a partir de sucesivas pruebas de regresiones logísticas previas al corpus base ampliado. El método de ensayo y error nos permite obtener un protocolo de actuación ordenado y aplicado, finalmente, para 1 000 iteraciones, mediante STATA. Este experimento lo realizamos dos veces ya que, al no tratarse de un cálculo exacto (porque las variables cotejadas son cualitativas) las pruebas son de realización extensa y apenas puede llegarse a un valor de verosimilitud, y no a un 100 % de exactitud, aunque sí lo más cerca posible de este porcentaje. Así, en una primera instancia, se aplica una regresión logística multinomial tomando como variable dependiente la relativa al valor semántico de *creo*, lo cual se lleva a cabo para 1 000 iteraciones o repeticiones. Tras ello, se repite el proceso, esta vez partiendo de la variable de la función pragmática como dependiente, con el mismo número de repeticiones. En el siguiente apartado aportamos los resultados obtenidos en ambas repeticiones del protocolo diseñado, siguiendo los pasos concretos y ordenados del diseño.

5.2. Resumen del diseño de un modelo predictivo de los valores de *creo* en tres fases

En el modelo diseñado para el análisis predictivo de los valores de significado de *creo*, en primer lugar, y de sus funciones pragmáticas, en segundo lugar, determinamos para comenzar (1) *la capacidad explicativa de las variables cotejadas*. Obtenemos una tabla como la que sigue con los valores de R^2 de McFadden por cada uno de los cruces de variables:

Tabla 5. R² de McFadden para las variables independientes en la determinación del significado de *creo* (Soler, 2019).

Modelos de regresión multinomial	Log Likelihood	pseudo-R ² de McFadden (1.ª INSTANCIA)
SIGNIFICADO (sin variables)	-753.04064	
SIGNIFICADO - CONSTRUCCIÓN	-570.74281	0,242082326
SIGNIFICADO - INTEGRACIÓN	-744,02822 (el modelo converge)	0,011968039
SIGNIFICADO - OD	-718,37835 (el modelo converge)	0,046029773
SIGNIFICADO - OI	-751,82063 (el modelo converge)	0,001620112
SIGNIFICADO - NEGACIÓN V.	-647,25325 (el modelo converge)	0,14048032
SIGNIFICADO - NEGACIÓN V. SUB.	-751,94636 (el modelo converge)	0,001453149
SIGNIFICADO - PERS. Y NÚM. V. SUB.	-683,25781 (el modelo converge)	0,09266808
SIGNIFICADO - TIEMPO V. SUB.	-639,50454 (el modelo converge)	0,150770216
SIGNIFICADO - MODO V. SUB.	-639,71697 (el modelo converge)	0,15048812
SIGNIFICADO - SUJETO	-730,20534 (el modelo converge)	0,030324127
SIGNIFICADO - POSICIÓN SINT.	-704,22015 (el modelo converge)	0,064831149
SIGNIFICADO - REPETICIONES	-706,214 (el modelo converge)	0,062183417
SIGNIFICADO - MMDD	-727,2134 (el modelo converge)	0,034297272
SIGNIFICADO - OTROS ELEMENTOS	-642,77243 (el modelo converge)	0,146430623
SIGNIFICADO - GÉNERO	-667,35646 (el modelo converge)	0,113784271
SIGNIFICADO - TIP. TEXTUAL	-598,01081 (el modelo converge)	0,205871797

Seguidamente, a partir de estos datos, se calcula (2) *la jerarquía de las variables en el aumento paulatino de explicación* que proporcionan sobre el significado de *creo*. En la ordenación de esta jerarquía, nos fijamos en el valor de verosimilitud proporcionado por R² de McFadden, si bien también se tienen en cuenta cuestiones cualitativas de aplicación de las variables al análisis. Así, por ejemplo, se observa cualitativamente que las características de tipo formal son más rápidamente reconocibles por parte del analista (las cuales subimos en la escala de jerarquía), y que otras de tipo semántico presentan una detección más compleja (razón por la que, en algunos casos, las relegamos a puestos inferiores de la jerarquía de aplicación). Así mostramos los resultados de la segunda instancia en dos tablas. Primeramente, observamos que en la Tabla 6 aparecen todos los resultados de verosimilitud obtenidos. Seguidamente, en la Tabla 7 reordenamos los parámetros de análisis de mayor a menor grado de explicación sobre el valor semántico del verbo y añadimos el porcentaje de error que este pueda estar generando.

Tabla 6. R² de McFadden ordenados por valor, en la determinación del significado de *creo* (Soler, 2019).

	VARIABLES JERARQUIZADAS	PSEUDO-R ² DE MCFADDEN PORCENTUAL (2.ª INSTANCIA)
1	TIPOLOGÍA TEXTUAL	25,44 %
2	CONSTRUCCIÓN	24,21 %
3	TIEMPO VERBO SUBORDINADO	16,43 %
4	OTROS ELEMENTOS	15,58 %
5	GÉNERO	12,43 %
6	PERSONA Y NÚMERO VERBO SUB.	6,44 %
7	POSICIÓN SINTÁCTICA	6,42 %
8	SUJETO	5,36 %
9	NEGACIÓN <i>CREO</i>	2,94 %

Tabla 7. R² de McFadden reordenados por jerarquía de aplicación, en la determinación del significado de *creo* (Soler, 2019).

VARIABLES ORDENADAS	ERROR COMÚN ESTÁNDAR
CONSTRUCCIÓN	43,82 %
SUJETO	32,80 %
NEGACIÓN <i>CREO</i>	44,45 %
TIEMPO VERBO SUBORDINADO	30,11 %
PERSONA Y NÚMERO VERBO SUB.	32,80 %
POSICIÓN SINTÁCTICA	45,70 %
OTROS ELEMENTOS	27,82 %
TIPOLOGÍA TEXTUAL	20,70 %
GÉNERO	31,72 %

En la tabla 7 vemos cómo, en el último paso de nuestro protocolo (3) **se obtiene un error ajustado de los cálculos realizados**. Este permite ver que no ha habido desfases entre la extracción de los valores de verosimilitud de las tablas previas y los de la probabilidad total de que se reconozcan los datos de cada variable. Dado que, en este caso, para la semántica de *creo*, todos los valores obtenidos son menores al 50 % y no presentan diferencias relevantes respecto a los datos de verosimilitud de las regresiones llevadas a cabo, no se plantea una nueva reorganización en la jerarquía de aplicación de las variables, respecto a la ya propuesta.

Para el caso de la determinación de las funciones pragmáticas de *creo*, que se han establecido en las tres categorías reconocidas por la bibliografía previa (a saber, *atenuación*, *neutralidad* e *intensificación*), dado que el análisis cualitativo cognitivo aplicado las reco-

noce, efectivamente, en los mismos términos, se repite el experimento de tres fases diseñado, pero ahora, sobre la base del significado, ya reconocido gracias a la aplicación de las fases de análisis explicadas. Véase la tabla final obtenida:

Tabla 8. R² de McFadden reordenados por jerarquía de aplicación, en la determinación de la función pragmática de *creo* (Soler, 2019).

Regresión multinomial	Log Likelihood	Pseudo-R2 de McFadden (1.ª instancia)
FUNCIÓN-SDO	-1227,0094	21,37 %
FUNCIÓN-SDO-INTERSUBJLOC.	-1117,7345	8,91%
FUNCIÓN-SDO-POSIC. SINT.	-1.125	8,34%
FUNCIÓN-SDO-GEN.	-1132,3951	7,71%
FUNCIÓN-SDO-IMAGEN	-1.134	7,62%
FUNCIÓN-SDO-POLIF.	-1142,6586	6,87%
FUNCIÓN-SDO-ASERTIVIDAD	-1.144	6,78%
FUNCIÓN-SDO-OTROSELS.	-1147,1233	6,51%
FUNCIÓN-SDO-TXT.	-1148,9373	6,36%
FUNCIÓN-SDO-CONVENC.	-1164,7245	5,08%

En este segundo experimento, se parte de que el significado de *creo* ya ha sido establecido con la primera aplicación del protocolo. De este modo, se reduce el número de variables en el cálculo de la función pragmática. En la obtención de errores, se estima que estos, de nuevo, no alteran los datos de verosimilitud de R² de McFadden y, por consiguiente, el orden y jerarquía de aplicación de las variables para la determinación de la función pragmática de *creo* se mantiene como muestra, más arriba, la Tabla 8.

El análisis demuestra, pues, que del total de variables cognitivamente descritas para el posible análisis semántico-pragmático de *creo*, solo algunas de ellas son eficaces con más de un 20 % de explicación y hasta más de un 50 %, mientras que otras, pueden descartarse, al menos, en un estudio genérico para detectar lo más automáticamente posible ante qué tipo de *creo* nos encontramos.

6. Conclusiones

El análisis de este capítulo confirma que es posible diseñar un modelo de análisis cuantitativo que, siempre como complemento del análisis cualitativo de fenómenos lingüísticos como el del funcionamiento de las formas performativas de los verbos cognitivos, permite determinar más del 60 % de sus valores semánticos, así como entre el 80 y el 100 % de sus funciones pragmáticas (si sumamos el valor de verosimilitud de la aplicación de las varia-

bles jerarquizadas en el protocolo). Asimismo, el modelo planteado para el caso de *creo* es replicable en otro tipo de textos y géneros discursivos. En contraposición con las carencias metodológicas de la bibliografía previa, el modelo de análisis creado mejora y perfecciona la aplicación de pruebas estadísticas que han resultado insuficientes en otros estudios. Este hecho confirma la hipótesis de partida de este trabajo, ya que es posible completar el análisis cualitativo de *creo* con el acercamiento cuantitativo riguroso no solo descriptivo, sino también predictivo. Este modelo de análisis puede describirse a partir de tres fases: (1) determinación de la capacidad explicativa de las variables seleccionadas con criterios cognitivos, (2) jerarquización de las variables para la descripción semántica y pragmática de *creo* (o la forma verbal considerada), y (3) cálculo de errores cometidos en el proceso, las cuales dan respuesta a la pregunta de investigación del trabajo, la cual se cuestionaba si era posible llegar a una sistematización de análisis para el reconocimiento semántico-funcional de unidades lingüísticas complejas como la que nos atañe.

En conclusión, cabe decir que este capítulo ha pretendido ofrecer un paradigma de estudio que es compatible con los ya conocidos, pero que viene a completar los puntos que no habían sido solventados hasta ahora por la investigación lingüística más tradicional. Queda para el futuro próximo replicar este patrón propuesto y perfilar el modelo de análisis y las fases de su consecución, así como también cotejar los resultados específicos que pueda dar su aplicación a otros formatos textuales y fenómenos lingüísticos.

— Referencias

- Abbhul, R. & Mackey, M. (2013). Experimental research design. In R. Abbuhl, S. Gass & M. Mackey, *Research Methods in Linguistics* (pp. 116-134). Cambridge University Press.
- Abdulrahim, D. (2014). Annotating corpus data for a quantitative, constructional analysis of motion verbs in Modern Standard Arabic. In N. Habash, & S. Vogel (Eds.), *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)* (pp. 28-38). Association for Computational Linguistics.
- Achard, M. (1998). Representation of cognitive structures. *Cognitive Linguistics*, 15(4), 588-594.
- Boas, H. & Ziem, A. (2018). Constructing a constructicon for German. Empirical, theoretical, and methodological issues. In B. Lyngfelt, L. Borin, K. Ohara, & T. Timponi (Eds.), *Constructicography: Constructicon development across languages* (pp. 183-228). John Benjamins. <https://doi.org/10.1075/cal.22.07boa>.
- Buceta, O. (2014). Construcciones del verbo 'creer'. *Factótum*, 12, 74-90.
- Cutting, J. (Ed.). (2007). *Vague Language Explored*. Palgrave MacMillan.
- De Hoop, H., Foolen, A., Mulder, G. & Van Mulken, V. (2018). *I think* and *I believe*: Evidential expressions in Dutch. In A. Foolen, H. de Hoop & G. Mulder (Eds.), *Evidence for Evidentiality* (pp. 77-97). John Benjamins. <https://doi.org/10.1075/hcp.61.04hoo>.

- Díaz-Campos, M. & Gradoville, M. (2011). An Analysis of Frequency as a Factor Contributing to the Diffusion of Variable Phenomena: Evidence from Spanish Data. In L. Ortiz (Ed.), *Selected Proceedings of the 13th Hispanic Linguistics Symposium*, (pp. 224-238). Cascadilla Proceedings Project.
- Fetzer, A. (2014). *I think, I mean and I believe* in political discourse. Collocates, functions and distribution. *Functions of Language*, 21(1), 67-94.
- Fetzer, A. & Johansson, M. (2010). Cognitive verbs in context. A contrastive analysis of English and French argumentative discourse. *International Journal of Corpus Linguistics*, 15(2), 240-266.
- Fuentes Rodríguez, C. (2010). La aserción parlamentaria: de la modalidad al metadiscurso. *Oralia*, 13, 97-125.
- Fuentes Rodríguez, C. (2016). Atenuación e intensificación estratégicas. In C. Fuentes Rodríguez (Ed.), *Estrategias argumentativas y discurso político* (pp. 163-222). Arco/Libros.
- González Ruiz, R. (2015). Los verbos de opinión entre los verbos parentéticos y los verbos de recepción débil: aspectos sintácticos y semántico-pragmáticos. *Círculo de Lingüística Aplicada a la Comunicación*, 62, 148-173.
- González, J., Boeck, P. & Tuerlinchx, F. (2014). Linear mixed modelling for data from a double mixed factorial design with covariates: a case-study on semantic categorization response times. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(2), 289-302.
- Hartwell, L. M., Esperança-rodier, E. & Tutin, A. (2017). *I think we need...*: Verbal expressions of opinion in conference presentations in English and in French. *Romance Corpora and Linguistic Studies*, 4(1), 35-60.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). An Introduction to Statistical Learning: With Applications in R. Springer.
- Jansegers, M. (2017). *Hacia un enfoque múltiple de la polisemia. Un estudio empírico del verbo multimodal "sentir" desde una perspectiva sincrónica y diacrónica*. Mouton de Gruyter.
- Jansegers, M. & Gries, S. (2017). Towards a dynamic behavioral profile: a diachronic study of polysemous 'sentir' in Spanish". *Corpus Linguistics and Linguistic Theory*, 16(1), 145-187.
- Milin, P., Divjak, D., Dimitrijević, S. & Baayen, R. H. (2016). Towards cognitively plausible data science in language research. *Cognitive Linguistics*, 27(4), 507-526.
- Roldán, A. (2005). Applications of cognitive linguistics (CI) to languages for specific purposes (LSP). In M. L. Carrió (Coord.), *Perspectivas interdisciplinarias de la lingüística aplicada*, Vol. 2 (pp. 325-332). Universitat de València.
- Soler, M. A. (2018). Algunos apuntes bibliográficos en torno a los verbos de opinión. In C. J. Álvarez López & M. R. Martínez Navarro (Coords.), *En busca de nuevos horizontes. Algunas líneas actuales en los estudios hispánicos* (pp. 59-70). Edições Húmus,
- Soler, M. A. (2019). *Semántica y pragmática de los verbos doxásticos en la interacción oral en español. Un estudio monográfico sobre la forma verbal creo* [Tesis doctoral. Universitat de València]. RODERIC. <https://roderic.uv.es/handle/10550/71798>
- Soler, M. A. (2021a). Análisis cognitivo de la semántica de *creo* en el español occidental hablado. En L. E. Aguilera, E. de los Santos, M. E. Flores & J. Haidar (Eds.), *Enfoques alternativos en los estudios del discurso* (pp. 92-107). Universidad Autónoma de Nuevo León.
- Soler, M. A. (2021b). *Semántica de creo. Análisis cognitivo de la polisemia de una forma verbal doxástica en la interacción oral en español*. Peter Lang.

- Hevner, A. R. (2007). A three-cycle view of design science research. *Scandinavian journal of information systems*, 19(2), 4.
- Järvinen, P. (2007). Action research is similar to design science. *Quality & Quantity*, 41(1), 37-54.
- Kutela, B., and Teng, H. (2019). Prediction of drivers and pedestrians' behaviours at signalized mid-block Danish offset crosswalks using Bayesian networks. *Journal of Safety Research* 69, 75-83. <https://doi.org/10.1016/j.jsr.2019.02.008>
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K., & Teng, R. (1993). Five papers on WordNet (TM). *International Journal of Lexicography*, 3(4), 235-244.
- Naciones Unidas (2019). *Informe de los objetivos de desarrollo sostenible*. Naciones Unidas.
- O'Riain, S., Curry, E., & Buitelaar, P. (2012). Engaging Practitioners within Design Science Research: A Natural Language Processing Case Study. In M. Helfert, & B. Donnellan (Eds.), *Design Science: Perspectives from Europe. EDSS 2012. Communications in Computer and Information Science*, vol 388 (pp. 155-169). Springer, Cham. https://doi.org/10.1007/978-3-319-04090-5_14
- Pereira, A. R., Ferreira, J. J. P., & Lopes, A. (2020). A knowledge representation of the beginning of the innovation process: The Front End of Innovation Integrative Ontology (FEI2O). *Data & Knowledge Engineering*, 125, 101760. <https://doi.org/10.1016/j.datak.2019.101760>
- Sandri, M.; Berchiolla, P.; Baldi, I.; Gregori, D.; & De Blasi, R., A. (2014). Dynamic Bayesian Networks to predict sequences of organ failures in patients admitted to ICU. *Journal of biomedical informatics*, 48, 106-13.
- Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval*. Cambridge University Press.