EL VERBO HACER EN EL ESPAÑOL DE CHILE: TIPOS Y COMBINACIONES FRECUENTES EN EL GÉNERO ENSAYO

MARÍA NATALIA CASTILLO FADIĆ
Pontificia Universidad Católica de Chile

RESUMEN

Este artículo estudia el verbo <hacer> en un corpus etiquetado de 100.033 palabras en contexto, correspondiente al mundo Ensayo *del* Corpus Básico del Español de Chile *(Castillo Fadić, 2012a).*

Los datos fueron analizados cuantitativamente con ayuda de AntConc 3.2.4w. Se estudiaron los tipos más frecuentes, organizados por persona, número, tiempo y modo y se contrastaron los resultados relativos a los tiempos frecuentes con los obtenidos del análisis de otros mundos, también extraídos de Castillo Fadić (2012a). Se realizó luego un análisis por clusters *y colocaciones, para determinar las combinaciones más frecuentes del verbo y dar cuenta, tanto de posibles restricciones léxicas, como de las estructuras sintácticas predominantes en las que se presenta <hacer> en los ensayos chilenos contemporáneos.*

Se observó, en la conjugación de <hacer> en Ensayo, una preferencia marcada por la 3ª persona singular, el tiempo presente y el modo indicativo. Se confirmó la escasa permeabilidad del verbo a restricciones léxicas y se abstrajeron siete estructuras sintácticas frecuentes que se combinan con este verbo.

Palabras clave: estadística léxica, lingüística de corpus, español de Chile, verbo <hacer>, léxico-sintaxis.

ABSTRACT

The present article studies the verb "hacer" in context in a tagged corpus of 100,033 words, belonging to the Ensayo (Essay) world of Corpus Básico del Español de Chile (Castillo Fadić, 2012a).

The data was analyzed quantitatively using AntConc 3.2.4w. The most frequent types in Ensayo were studied, organized by grammatical person, number, time and mode. The most frequent verbal times were compared with results from previous analyses of other worlds, also extracted from Castillo Fadić (2012a). In addition, clusters and collocations were analyzed, in order to determine this verb's most frequent combinations and identify both the lexical restrictions and the dominant syntactic structures in which "hacer" participates in contemporary Chilean essays. Regarding the conjugation of "hacer" in Ensayo, a definite preference was observed for the third person singular, present time and indicative mode. The limited permeability of "hacer" to lexical restrictions was confirmed, and seven syntactic structures were identified in which this verb participates.

Key words: Chilean Spanish, verb "hacer", corpus linguistics, lexical statistics, lexical-syntax.

1. INTRODUCCIÓN

Este estudio lexicoestadístico, cuyo principal antecedente es Castillo Fadić (2012b), se centra en las realizaciones y combinaciones frecuentes del verbo <hacer> en ensayos chilenos contemporáneos. Emplea herramientas de lingüística de corpus y lingüística computacional y se centra en los niveles léxico y sintáctico. Sus principales objetivos son: a) describir estadísticamente los tipos de <hacer>, organizados por persona, número, tiempo y modo; b) describir estadísticamente las combinaciones léxicas frecuentes del verbo; c) analizar las combinaciones frecuentes para determinar la

presencia de unidades fraseológicas, colocaciones o restricciones léxicas; d) abstraer las estructuras sintácticas más frecuentes que involucran al vocablo, a partir del análisis de la combinatoria léxica.

La elección de este verbo se debe a que Castillo Fadić (2012b) lo ubica en el rango 4 de los verbos de mayor dispersión y en el 5 de los de mayor uso: f_i=1777, D=0,81 (escala 0-1) y U=1439,37. Sus índices estadísticos suponen una covariación escasa respecto de variables externas.

La investigación pretende contribuir a la descripción del verbo <hacer> en el español de Chile y a la del género ensayo, brindando herramientas que puedan ser aprovechadas para la enseñanza-aprendizaje de verbos, así como para la creación de aplicaciones de lingüística computacional.

2. ESTADÍSTICA Y COMBINATORIA LÉXICA

La estadística léxica, inserta dentro de la lingüística cuantitativa, exige pasar de la noción de *palabra* –unidad de texto–, a la de *vocablo* –unidad de léxico– (Müller, 1973, págs. 225-226); el vocablo "representa todo el paradigma de flexiones, derivaciones o conjugaciones que se forman a partir de una raíz o de un núcleo morfemático y que se reconocen como palabras" (Lara, 2006, pág. 138). En un nivel de abstracción intermedio se ubica la noción de *tipo*, que permite agrupar las ocurrencias de una palabra bajo una sola etiqueta, posibilitando el conteo estadístico.

En este conteo, se mide la *frecuencia*. La *absoluta* (f_i) corresponde al número de veces que aparece un elemento en un corpus, mientras la *relativa* se observa en relación con el tamaño de la muestra (f_i/n). Aquí trabajamos con f_i y consideramos la medida de *dispersión compleja* (D), que da cuenta del equilibrio en la distribución de las frecuencias en distintos *mundos* (*Worlds,* en Juilland & Chang-Rodríguez, 1964, pág.XXVI). Para su cálculo, utilizamos la fórmula de Juilland, Traversa, Beltramo, & Di Blasi (1973), que considera el número de mundos (n=5), la suma de las frecuencias reales de cada palabra o vocablo en cada mundo (x_i) y la suma de las frecuencias totales de cada palabra o vocablo (T):

$$D = 1 - \frac{\sqrt{nx_i^2 - T^2}}{2T}$$

El producto de la f_i por la D es el *uso* (U); los vocablos de mayor uso en una comunidad constituyen su léxico básico y se caracterizan por su distribución estable: según el rango en que se ubiquen, sus posibilidades de aparición presentan una dependencia de variables externas de nula a mínima.

Si es posible medir la f_i de una unidad léxica, también lo es medir la f_i de sus combinatorias. Las unidades que se combinan de modo frecuente pueden corresponder a colocaciones o a unidades fraseológicas diversas, según su grado de libertad o fijación, o pueden indicar una restricción de selección de la unidad predicativa. Para Bosque (2001, pág. 17), las colocaciones "constituyen "combinaciones frecuentes" y preferentes de unidades léxicas, pero la frecuencia y la preferencia son propiedades que pueden responder a factores lingüísticos y extralingüísticos sumamente diversos […]"; consecuentemente, afirma que "la mayor parte de las llamadas "colocaciones" no son combinaciones binarias establecidas estadísticamente […], sino asociaciones entre un predicado y un conjunto restringido de argumentos que determinan una subclase léxica de mayor o menor extensión". Las restricciones de selección corresponden básicamente a "las distribuciones semánticamente aceptables que admite una determinada palabra"

(López Martínez, 2004, pág. 269). Un verbo de alto uso debería presentar escasas restricciones.

3. METODOLOGÍA

3.1. Obtención del corpus

Operamos sobre un corpus de referencia etiquetado de 100.033 palabras en contexto, extraídas de 83 ensayos producidos por 76 autores chilenos y publicados por primera vez entre 1981 y 2006. El corpus *Ensayo* forma parte del *Corpus Básico del Español de Chile* (Castillo Fadić, 2012a), de más de 500.000 palabras pertenecientes a cinco mundos –Drama, Narrativa, Ensayo, Técnico-Científico y Prensa–, base a partir de la cual se obtuvo el *Léxico Básico del Español de Chile* (Castillo Fadić, 2012b).

3.2. Procesamiento del corpus

El corpus fue convertido a *.TXT, codificación ANSI, y se procesó con AntConc 3.2.4w.

Se buscaron las conjugaciones de <hacer> presentes en *Ensayo* (Castillo Fadić (2012b), lo que permitió obtener listados de concordancias para cada tipo (cf. Figura 1). Luego, las conjugaciones de <hacer> se lematizaron en un segundo archivo *.TXT, reemplazando las variantes por el infinitivo; al procesar el corpus lematizado, se obtuvieron las estadísticas de combinatorias del vocablo.

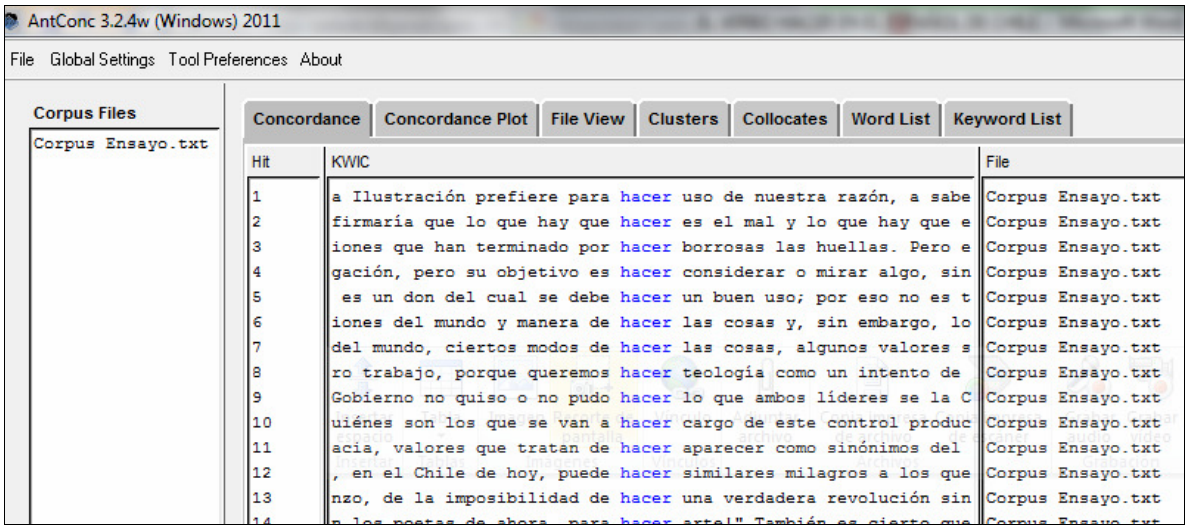

Figura 1. AntConc 3.2.4w, recorte de pantalla: concordancias <hacer>

3.3. Descripción y análisis de los datos

Los tipos de <hacer> se organizaron por persona, número, tiempo y modo y se describieron estadísticamente. Se realizó luego un análisis por *clusters* en AntConc 3.2.4w, para determinar las combinaciones más frecuentes. La búsqueda se centró en combinaciones de $f_i \geq 2$, de mínimo dos y máximo cinco unidades; la consulta más productiva para nuestros efectos fue, indudablemente, la que sitúa a <hacer> a la izquierda del *cluster*.

A partir de los archivos *output* generados en AntConc 3.2.4w, se observaron las combinaciones léxicas más frecuentes del verbo <hacer>. Del análisis de los índices estadísticos y de las concordancias de las combinaciones léxicas frecuentes, se

abstrajeron las estructuras sintácticas frecuentes y se estudiaron las posibles unidades fraseológicas o restricciones.

4. RESULTADOS

4.1. Tipos frecuentes

Las variantes de <hacer> están representadas predominantemente por el infinitivo (29%) y la tercera persona singular de Presente Indicativo (28%) (cf. Figura 2).

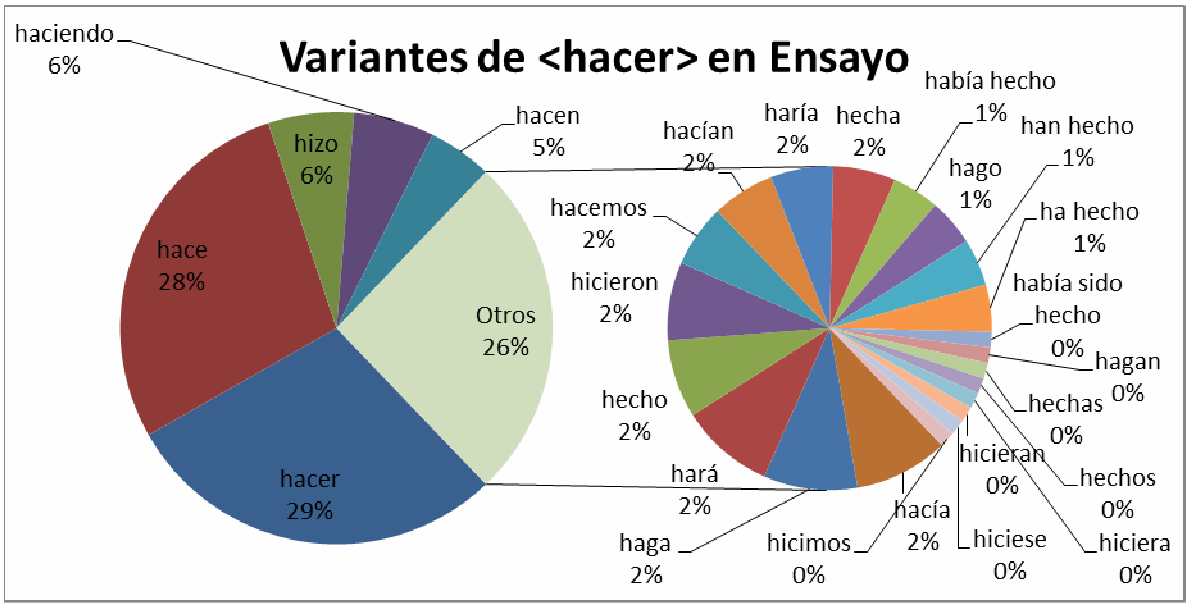

Figura 2. Distribución porcentual de variantes de <hacer> en *Ensayo*.

Al agrupar estas realizaciones en categorías mayores, se observa una preponderancia de las formas finitas (61%) por sobre las no finitas (39%). Dentro de las no finitas, el 74% corresponde al infinitivo; el 15%, al gerundio y el 11%, al participio. Dentro de las finitas, el 99% corresponde al Modo Indicativo y sólo el 1% al Subjuntivo, que sería completamente marginal.

En lo concerniente a los tiempos (cf. Figura 3), predomina ampliamente el Presente, seguido por el Pretérito Perfecto Simple y el Imperfecto. Los cuatro pretéritos sumados alcanzan el 30% del total de realizaciones.

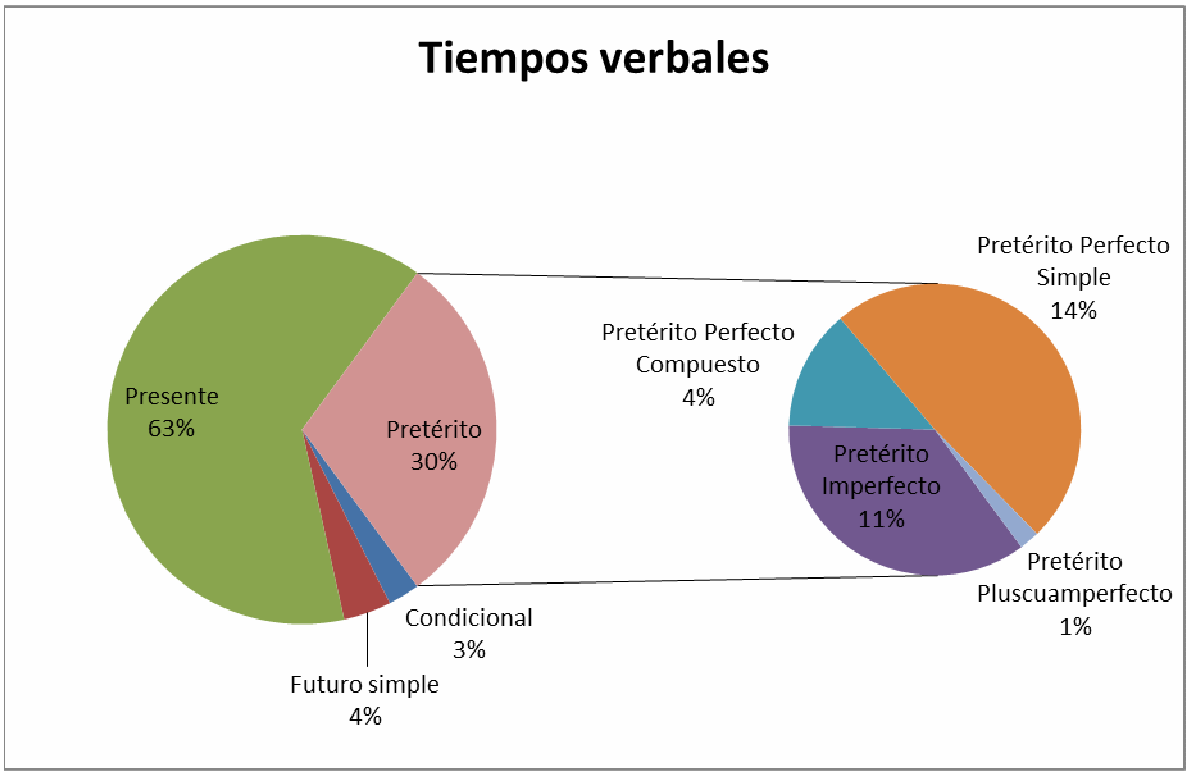

Figura 3. Tiempos verbales presentes en Ensayo

En el corpus, sólo figuran la primera y la tercera personas. La tercera predomina con un 95%, distribuido en un 81% para el singular y un 14% para el plural. La primera alcanza sólo a un 5% (cf. Figura 4).

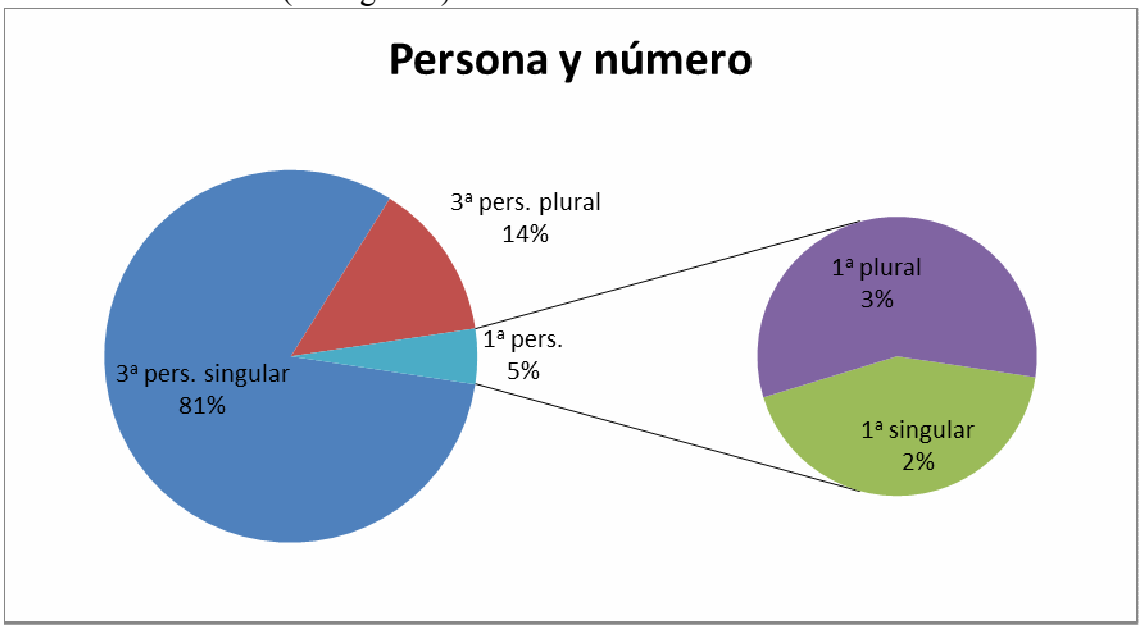

Figura 4. Persona y número en Ensayo

Procesados los datos, nos pareció de interés precisar si la preponderancia de algunas realizaciones era general en los distintos mundos o si correspondía a una especificidad de Ensayo, que pudiera contribuir a la clasificación automática de textos. Consideramos únicamente el infinitivo y el presente, por su mayor frecuencia. Las realizaciones menos frecuentes en Ensayo fueron agrupadas bajo la etiqueta “Otros”.

Observamos (cf. Figura 5) que la proporción de <hacer> en infinitivo es similar en Drama (28%), Ensayo (29%) y Técnico-Científico (31%), pero marcadamente superior

a la de Narrativa (22%) y Prensa (20%). La proporción de las realizaciones del Presente es claramente mayor en Ensayo (39% en total), seguido por Prensa (35%); más abajo se sitúan Técnico-Científico (31%) y Drama (30%); Narrativa (20%) arroja el valor más bajo.

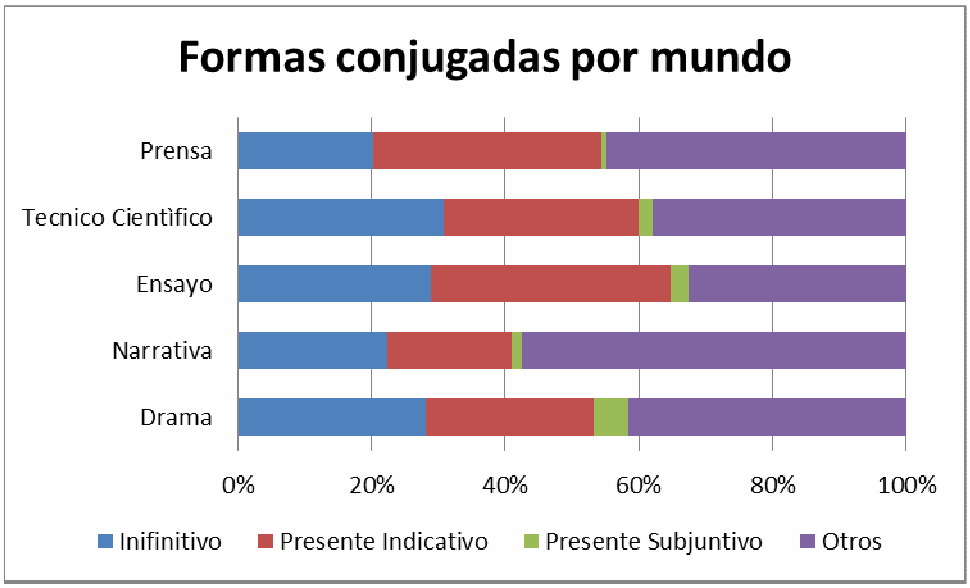

Figura 5. Formas conjugadas por mundo

4.2. Combinaciones frecuentes

El análisis por *clusters* (cf. Figura 6), ordena las combinaciones por frecuencia. Cuando el verbo <hacer> aparece en mayúsculas, indica que en su lugar había un tipo diferente, que fue lematizado.

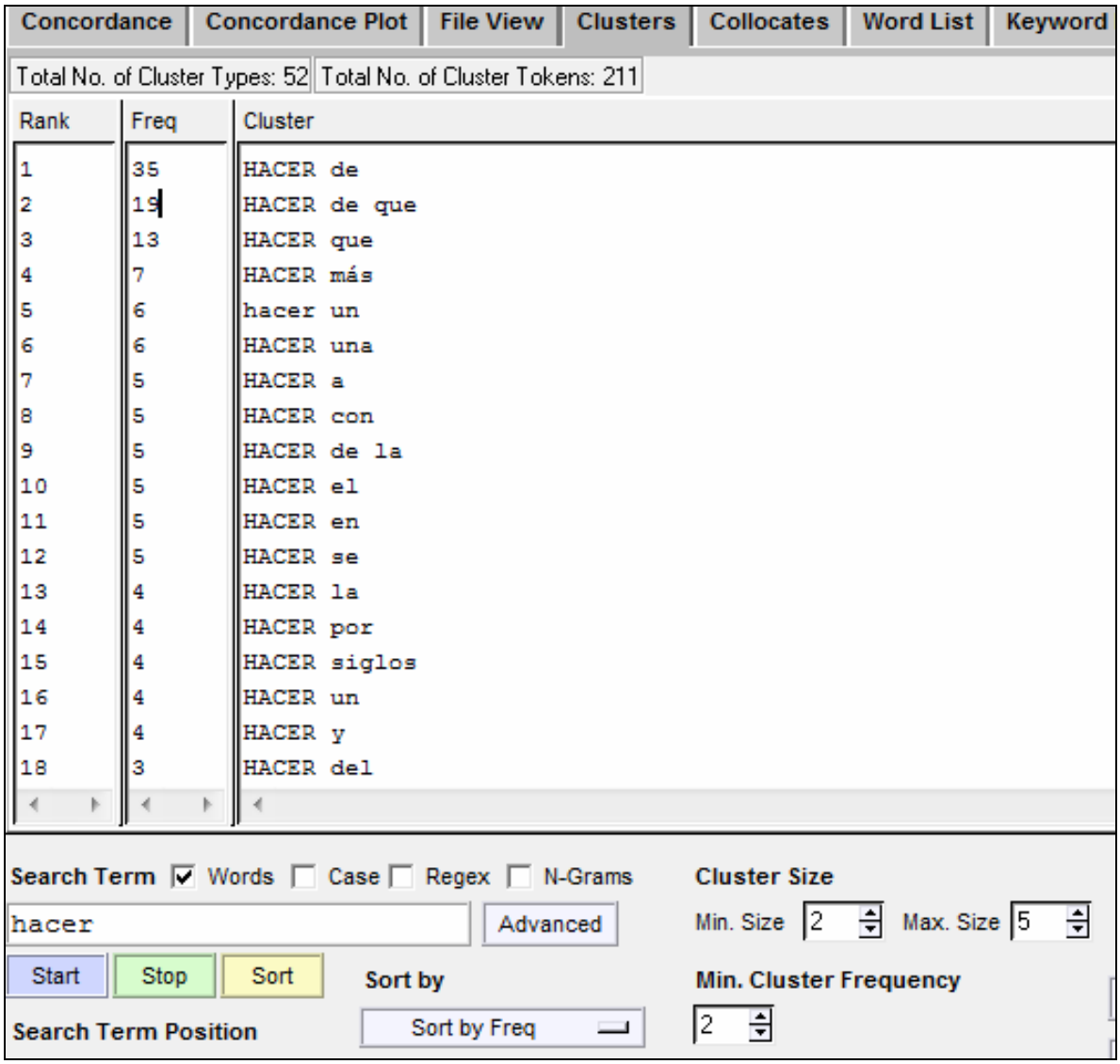

Figura 6. AntConc 3.2.4w, recorte de pantalla: análisis por *clusters*

Al analizar la combinatoria léxica, se observan construcciones sintácticas recurrentes, que presentamos en orden de frecuencia:

4.2.1 <hacer>+complemento (fi=56)

La combinación más frecuente es la de <hacer> + preposición <de> (f_i=35), a la que sumamos <hacer> + <del> (f_i=2), llegando a una f_i=37. Esta construcción puede incluir también otras preposiciones: <a> (f_i=5); <con> (f_i=5); <en> (f_i=5); <por> (f_i=4).

4.2.2 <hacer>+frase nominal (f_i=40)

La combinación <hacer> + artículo (f_i=35), da cuenta de esta construcción, puesto que en español los artículos anteceden a sustantivos. El que los sustantivos que suceden a los artículos no suelan alcanzar una fi ≥2 da cuenta de la ductilidad combinatoria de este verbo y de sus escasas restricciones. La estructura <hacer> + <las cosas> (f_i= 2) es probablemente la única locución identificada; aparece como término de un complemento encabezado por <de>, determinando a los sustantivos <manera> y <modo> respectivamente: <maneras de hacer las cosas> y <modos de hacer las cosas>.

El verbo puede combinarse directamente con un sustantivo: <hacer>+<sustantivo> (f_i=5). Sucede en <hacer>+sustantivo <nada> (f_i=3) y <hacer>+sustantivo <filosofía> (f_i=2). En un análisis manual, se observan numerosas combinaciones con f_i=1. La combinación <hacer>+<nada> requiere un marcador previo de negación: <*nunca* consiguen *hacer nada*>, <*no* se puede *hacer nada*>, <energía empleada en *no hacer nada*>; además, es sustituible por un pronombre, por lo que estimamos que puede funcionar como locución sustantiva: <nunca *lo* consiguen>, <no se puede *eso*>, <energía empleada en *eso*>.

4.2.3 <hacer> + oración subordinada sustantiva (f_i=13)

Esta construcción se desprende de la combinatoria <hacer>+<que> (f_i=13).

4.2.4 <hacer> + <adjetivo>

Aunque en el análisis por concordancias pueden detectarse abundantes casos de f_i=1, las únicas combinaciones con $f_i \geq 2$ son: <hacer>+adjetivo <patente> (f_i=4), <necesario> (f_i=2), <posible> (f_i=2) y <presente> (f_i=2).

El reconocimiento de otras estructuras frecuentes exige una revisión complementaria de las concordancias:

4.2.5 <hacer>+<frase adjetiva> encabezada por adverbio intensificador (f_i=6)

La construcción <hacer>+<más> (f_i=10) puede develar diversas estructuras; mayoritariamente, da cuenta de <hacer>+<frase adjetiva> encabezada por adverbio intensificador (f_i=6), como en <hacer más fecundo>.

Construcción similar es <hacer>+<cada vez más>+<adjetivo>. (f_i=6).

4.2.6 Uso impersonal de <hacer> con valor temporal (fi =11)

La combinación frecuente <hacer> + <más> puede revelar un uso impersonal con valor temporal del verbo, determinado por las unidades con las que se combina. Sucede en la construcción preposición <desde> + <hacer> + <más de> + <numeral cardinal> + complemento encabezado por preposición <de>, donde el núcleo de la frase sustantiva que funciona como término del complemento es un sustantivo con significado de tiempo (f_i=2): <desde hace más de treinta millones de años>; <desde hace más de diez mil millones de años>.

Construcciones diversas incorporan un sustantivo o frase sustantiva con valor temporal: <hacer>+<siglos> (f_i=4) y <hacer>+ <una década> (f_i=2). Entre el verbo y el sustantivo o frase sustantiva puede incorporarse el adverbio <ya>: <hace ya un año>, <hace ya cinco años>, <hace ya una década> (f_i = 3).

4.2.7 <hacer>+<verbo en infinitivo> (fi =6).

Presenta dos opciones: a) <hacer>+<infinitivo> (f_i=2) y b) adverbio <no>+<hacer>+<más que/ sino>+<verbo en infinitivo> (f_i=4). El único ejemplo de la primera variante con $f_i \geq 2$ es <hacer>+<avanzar> (f_i=2); en el análisis de las concordancias se observan otros casos con f_i=1. La segunda alternativa puede construirse con <más que> (f_i=2): <no hace más que exhibir su incapacidad>, <no hace

más que tartamudear> o con <sino> (f_i=2): <no hace sino revitalizar>, <no hace sino reforzar>.

5. CONCLUSIONES

En Ensayo, <hacer> se realiza preferentemente en tercera persona, número singular, tiempo presente y modo indicativo.

Del análisis por *clusters* se desprende que el verbo se combina con complementos, frases y cláusulas sustantivas y con adjetivos; al incorporar el análisis por concordancias, se aprecia su combinatoria con frases adjetivas e infinitivos, así como sus usos impersonales con valor temporal.

Con vistas a nuevas aplicaciones en lingüística computacional, orientadas por ejemplo a clasificación automática de textos, sería relevante precisar si la distribución de las conjugaciones de <hacer> se reproduce en otros verbos de Ensayo, así como determinar si las construcciones observadas en Ensayo coinciden con las de otros mundos o con las de otras variedades dialectales.

REFERENCIAS BIBLIOGRÁFICAS

Bosque, I. 2001. "Bases para un diccionario de restricciones léxicas", *Moenia*, 7: 11-52.

Castillo Fadić, M. N. 2012a. *Corpus Básico del Español de Chile*.

Castillo Fadić, M. N. 2012b. *Léxico Básico del Español de Chile*. Tesis para optar al grado de doctora por la Universidad de Valladolid. Valladolid: Inédita.

Juilland, A., & Chang-Rodríguez, E. 1964. *Frequency Dictionary of Spanish Words, The Romance Languages and their Structures, First Series SI.* La Haya: Mouton.

Juilland, A., Traversa, V., Beltramo, A., & Di Blasi, S. 1973. *Frequency Dictionary of Italian Words.* The Hague- Paris: Mouton.

Lara, L. F. 2006. *Curso de lexicología.* México: El Colegio de México.

López Martínez, F. 2004. "Repercusiones del modelo cognitivo idealizado en las restricciones léxicas: carencias de la definición lexicográfica en los diccionarios de uso". En J. Varela Zapata, J. M. Oro, & J. Anderson, *Lengua y sociedad: lingüística aplicada en la era global y multicultural* (págs. 365-369). Santiago de Compostela, España: Servicio de Publicacións de Universidade de Santiago de Compostela.

Müller, C. 1973. *Estadística lingüística*. Madrid: Gredos.