

EL PROYECTO *AGLE* (ARCHIVO GRAMATICAL DE LA LENGUA ESPAÑOLA): DESARROLLO Y PERSPECTIVAS

MANUEL LEONETTI
SILVIA GUMIEL
PILAR PÉREZ
JUAN ROMERO¹
Instituto Cervantes

El objetivo central de esta charla es presentar el *Archivo Gramatical de la Lengua Española (AGLE)*, que puede ser visitado en la página web del Instituto Cervantes², y mostrar su funcionamiento. En segundo lugar, trazamos la historia del proyecto haciendo hincapié en su estado actual, pero sin olvidar las perspectivas futuras y sus posibles aplicaciones.

1. CARACTERÍSTICAS GENERALES Y BREVE HISTORIA DEL *AGLE*

1.1. *¿Qué es el AGLE?*

El *AGLE* es un archivo compuesto aproximadamente por 115.000 fichas (o cédulas) procedentes la mayoría de textos escritos que D.

¹ El orden de los autores atiende a los siguientes criterios: en primer lugar aparece el director del proyecto y el resto de los autores aparecen por orden alfabético.

² URL: <http://cvc.cervantes.es/obref/agle/>

Salvador Fernández Ramírez (1896-1983) fue recopilando a lo largo de 50 años de manera ordenada en 75 cajas de cartón con el propósito de utilizarlo para escribir una gramática, de la que sólo llegó a publicar un volumen³. El proyecto *AGLE* consiste en la edición digital íntegra del Archivo de Salvador Fernández Ramírez con el propósito de que cualquier usuario pueda consultarlo en la página web del Instituto Cervantes.

La característica fundamental del *AGLE* (tanto del Archivo como de la edición digital del mismo) es que la información que contiene está ordenada siguiendo criterios gramaticales y no lexicográficos; en este sentido, el *AGLE* se parece más a una gramática que a un diccionario. El acceso a las cédulas o papeletas se realiza por medio de su categoría gramatical, lo que nos permite encontrar datos sobre las construcciones sintácticas del español de todas las épocas. Aunque actualmente no es posible, se espera que en el futuro el acceso a las cédulas también pueda realizarse por cualquiera de las palabras que contiene el Archivo, como en un archivo lexicográfico.

El *AGLE* tiene otra característica que lo distingue de otros corpus y es que, en sentido estricto, no es un corpus, sino un *archivo*. Esto es, las miles de cédulas no están tomadas al azar de un conjunto representativo de textos escritos, sino que son el resultado de una selección cuidadosa de datos gramaticales escogidos por Fernández Ramírez. Así pues, si comparamos el *AGLE* con otros corpus encontramos que sólo el primero está clasificado atendiendo a criterios gramaticales y sólo éste contiene observaciones gramaticales hechas por el autor (Salvador Fernández Ramírez) y los editores.

1.2. *Breve Historia del Proyecto*

Los cincuenta años que Salvador Fernández Ramírez empleó en la confección de su Archivo tienen su propia historia y merecen mucho más tiempo y espacio del que disponemos aquí. José Polo y otros gramáticos han plasmado con precisión en varios estudios (recogidos en la bibliografía) el recorrido histórico y científico de la obra de Salvador Fernández Ramírez. Nosotros nos limitaremos en esta presentación a trazar la historia del Archivo a partir de mayo de

³ Salvador Fernández Ramírez (1951): *Los sonidos, el nombre y el pronombre*, Madrid: Revista de Occidente.

1987, cuando Ignacio Bosque y José Antonio Millán, miembros del anterior equipo editorial, concibieron el proyecto de la edición íntegra del fichero en soporte digital y surgió, a partir de ese momento, el *AGLE*, proyecto que acoge y financia desde 1995 el Área Académica del Instituto Cervantes.

El primer equipo editorial, compuesto por Ignacio Bosque, José Antonio Millán y Maite Rivero, culminó la edición y publicación de la parte del Archivo que corresponde a “Las Partículas”, formada por algo más de 18.000 fichas o cédulas. En 2002 toma el relevo un segundo equipo editorial⁴, formado por Manuel Leonetti, como director del proyecto, Juan Romero, como coordinador, y Silvia Gumiel y Pilar Pérez, como técnicos. El nuevo equipo ha editado la parte correspondiente a “El Verbo” (casi 60.000 cédulas) que se publicará próximamente.

2. ORGANIZACIÓN Y CONTENIDOS DEL ARCHIVO

Archivo Gramatical de la Lengua Española de Salvador Fernández Ramírez

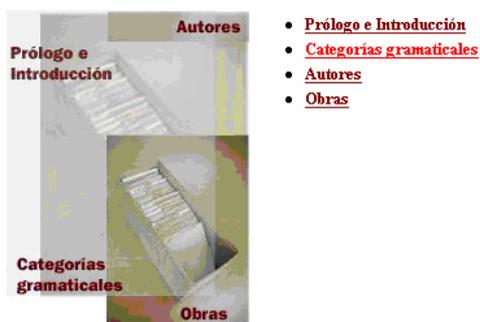


Figura 1. *Página principal de consulta del AGLE.*

Pasamos a la parte central de esta presentación, que tiene como objeto esbozar el diseño y manejo del *AGLE*, que se encuentra en la página web antes indicada (*vid.* nota 2), y que recogemos en la Figura 1 (*vid. supra*). En esta misma página se encontrará

⁴ Cabe destacar la colaboración y aportación de Edita Gutiérrez, Norberto Moreno e Isabel Pérez Jiménez en diferentes etapas del proyecto.

información sobre estos y otros aspectos del Archivo. Además, el *AGLE* cuenta con una opción de *Ayuda* que muestra las diferentes posibilidades de consulta.

2.1. *Contenidos*

Siguiendo el orden con el que Salvador Fernández Ramírez diseñó su *Gramática Española*, las 115.000 cédulas del Archivo se organizan en cinco grandes grupos que, a su vez, se organizan en grupos más pequeños:

| |
|--|
| Las partículas (18.000 cédulas) |
| El verbo (58.000 cédulas) |
| Formas no personales (14.500 cédulas) |
| Tiempo, modo y aspecto (12.000 cédulas) |
| Verbos copulativos (11.500 cédulas) |
| Clases de verbos (8.000 cédulas) |
| Complementos verbales (6.000 cédulas) |
| Número y persona (3.000 cédulas) |
| Valores de ‘se’ (3.000 cédulas) |
| El nombre (5.000 cédulas) |
| El nombre sustantivo (3.000 cédulas) |
| El nombre adjetivo (2.000 cédulas) |
| Determinantes y pronombres (26.000 cédulas) |
| El artículo (4.000 cédulas) |
| Los posesivos (2.000 cédulas) |
| Los demostrativos (2.000 cédulas) |
| Los pronombres personales (8.000 cédulas) |
| Relativos, interrogativos y exclamativos (6.000 cédulas) |
| Indefinidos y numerales (4.000 cédulas) |
| La oración (9.000 cédulas) |

Figura 2. *Organización del Archivo.*

Actualmente pueden ser consultadas “Las Partículas”, en el próximo año podrá ser consultado “El Verbo” y las próximas entregas se harán teniendo en cuenta el orden indicado.

2.2. *Organización*

Pasamos a continuación a repasar los aspectos fundamentales que han de tenerse presentes cuando se realicen búsquedas en el Archivo tal y como aparece en la página del Instituto Cervantes. El *AGLE* en

su versión actual permite tres tipos de búsqueda: “Por categoría”, “Por autor” y “Por obra”.

El criterio fundamental de clasificación que articula el Archivo es el *categorial*. Los apartados matrices de los que surge el árbol son las tradicionales clases de palabras: Preposición, Adverbio, Conjunción, Interjección y Partículas. Los apartados matrices no son ni construcciones ni funciones gramaticales, sino las partes de la oración. Como se muestra en la Figura 3, en una estructura arborescente, los subapartados van presentando las particularidades sintácticas y semánticas que estas unidades léxicas adquieren en diferentes construcciones. Así, la información relevante para las oraciones subordinadas la encontraremos en diferentes partes del Archivo, como en “Conjunciones subordinantes completivas”. Por esta razón es fundamental un sistema de referencias cruzadas del que hablaremos más adelante.

El “material disperso”, como Fernández Ramírez llamaba al Archivo, era la materia prima con la que el autor diseñaba su gramática y, por esta razón, algunas partes del Archivo presentan cierto grado de provisionalidad. En este sentido, la labor de los editores consiste en articular el árbol de categorías respetando al máximo los criterios de don Salvador y creando nuevas categorías –que aparecen entre corchetes– sólo para agrupar aquellas cédulas que están poco organizadas; para ello seguimos los criterios de clasificación del autor que veremos a continuación.

| | | |
|-------|---|---------------------------------|
| 1 | <u>La preposición</u> | 7468 |
| 11 | <u>Preposición 'a'</u> | 1124 (<u>Ver Comentarios</u>) |
| 12 | <u>Preposición 'ante'</u> | 43 (<u>Ver Comentarios</u>) |
| 13 | <u>Preposición 'bajo'</u> | 66 (<u>Ver Comentarios</u>) |
| 14 | <u>Preposición 'cabe'</u> | 3 (<u>Ver Comentarios</u>) |
| 15 | <u>Preposición 'con'</u> | 461 (<u>Ver Comentarios</u>) |
| 16 | <u>Preposición 'contra'</u> | 59 (<u>Ver Comentarios</u>) |
| 17 | <u>Preposición 'de'</u> | 1491 |
| 171 | <u>[Dependiente de un verbo]</u> | 538 (<u>Ver Comentarios</u>) |
| 172 | <u>[Dependiente de un nombre]</u> | 555 (<u>Ver Comentarios</u>) |
| 1721 | <u>Complemento regido</u> | 136 (<u>Ver Comentarios</u>) |
| 17211 | <u>Complemento objetivo</u> | 42 (<u>Ver Comentarios</u>) |
| 17212 | <u>[Complemento de origen]</u> | 2 (<u>Ver Comentarios</u>) |
| 17213 | <u>[Complemento de relación o jerarquía]</u> | 7 (<u>Ver Comentarios</u>) |
| 17214 | <u>Complemento subjetivo</u> | 32 (<u>Ver Comentarios</u>) |
| 17215 | <u>Complemento interno</u> | 41 (<u>Ver Comentarios</u>) |
| 17216 | <u>[Complemento identificativo o ejemplificativo]</u> | 8 (<u>Ver Comentarios</u>) |
| 17217 | <u>[Complemento de representación]</u> | 4 (<u>Ver Comentarios</u>) |
| 1722 | <u>Complemento atributivo o de identidad</u> | 98 (<u>Ver Comentarios</u>) |

Figura 3. Ejemplo de árbol de categorías.

2.3. Criterios de clasificación

Cuatro son las características principales que el propio Salvador Fernández Ramírez destaca de su *Gramática Española*, y que, por extensión, se ven reflejadas en la organización del Archivo: perspectiva sincrónica, concepción unitaria de la gramática, criterio formal y método inductivo. Aunque el conocimiento de estas características no es imprescindible para el manejo del AGLE, les dedicaremos unas breves líneas.

El propósito de Salvador Fernández Ramírez era el de describir un estado de lengua; pensaba en componer una *sintaxis del español común actual*. Sin embargo, a pesar del carácter sincrónico de su *Gramática*, no se llegó a desentender completamente de la visión diacrónica. El AGLE contiene miles de cédulas relacionadas con la diacronía del español; una colección de muestras de lengua española desde sus orígenes hasta la actualidad.

El autor tiene una concepción unitaria y orgánica de la gramática. Destaca la dificultad de deslindar los fenómenos de rección, subordinación, agrupamiento y orden de palabras en el campo sintáctico, etc. de lo puramente morfemático y fonológico. Por ejemplo, en la parte dedicada a las oraciones interrogativas Fernández Ramírez destaca el valor que tienen los hechos tonales en la estructura gramatical. Prueba de la importancia que nuestro autor concedía a los fenómenos morfofonológicos son los cuatro ficheros del AGLE dedicados a la fonología y a la morfología⁵.

Su intento fue manejar criterios rigurosamente formales, pero realizó un esfuerzo para buscar las conexiones que existen entre la forma del lenguaje y su aspecto interno. Así, es fácil encontrar categorías que hacen referencia a “categorías psicológicas”, “estilísticas”, “informativas”, “semánticas”, “pragmáticas”, etc. A modo de ejemplo, la preposición *por* en los siguientes ejemplos [*Por la salud de mis hijos, le digo*], [*Más despacio, por favor*] se clasifica como “De súplica o juramento”.

Emplea un método inductivo. A partir de los materiales tomados de textos escritos intenta descubrir y describir la norma común del español; no se basa en materiales de invención propia. La

⁵ Conviene no olvidar que Salvador Fernández Ramírez redactó las partes dedicadas a la fonología y a la morfología del *Esbozo* de la RAE.

procedencia de los materiales escritos es muy variada: desde la poesía hasta las disposiciones legales, desde el anuncio comercial hasta el tratado filosófico. Los autores de las obras impresas pertenecen a varias generaciones y a regiones diversas, aunque un gran número pertenece a obras publicadas en el siglo XX. No obstante, hay miles de cédulas que contienen datos del español de América y de otras épocas del español. En la parte de los verbos copulativos existen cientos de cédulas con construcciones del español antiguo hoy ya en desuso, pero que Fernández Ramírez recogía para estudiar la lengua en su evolución [*Si ellos aquí fuesen*], [*¿Somos cerca?*].

2.4. La Cédula

La ficha-tipo consta de las siguientes entradas: Número de Referencia, Cita, Autor, Obra y Árbol de categoría. Opcionalmente pueden aparecer Comentarios del autor o de los editores.

Cédulas seleccionadas

Nº de registros encontrados: (1 a 10 de 42)

| | |
|----------------------------|--|
| Referencia: | 2296 |
| Cita: | Jamás en las calles he presenciado el acoso de una mujer |
| Autor: | Azorín (José Martínez Ruiz) |
| Obra: | Valencia, Obras selectas VI, p. 374. |
| Comentario de los editores | [Parece que se refiere a la interpretación objetiva -la mujer es la acosada- no a la subjetiva.] |
| Árbol de la categoría: | <u>17211</u> La preposición Preposición 'de' [Dependiente de un nombre] Complemento regido Complemento objetivo |

Figura 4. Ejemplo de ficha-tipo.

El Número de Referencia es el número que los editores del *AGLE* han dado a cada cédula para facilitar su identificación. La Cita contiene la cita propiamente dicha. En la cita se respeta siempre la forma en que aparece la ficha, es decir, no se han acortado los textos. Dentro del texto aparecen tipos de letra distintos que indican la relevancia gramatical de la palabra marcada. Así, la porción de texto marcada con negrita y cursiva es el fenómeno que se está

clasificando. El texto marcado sólo con negrita es un criterio subsidiario del principal. Como se muestra en la Figura 4, en el ejemplo *el acoso de un mujer*, clasificado bajo las categorías “Preposición *de*” [Dependiente de un nombre], “Complemento regido”, “Complemento objetivo”, la preposición “*de*” aparecerá marcada con negrita y cursiva, y el nombre del que depende la preposición aparecerá marcado sólo con negrita.

Los campos Autor y Obra corresponden respectivamente al autor y la obra a la que pertenece la cita. Sólo en ocasiones la ficha contiene un ejemplo construido por el autor, lo que se hace notar expresamente. Estas informaciones no han sido cotejadas por el equipo editorial pero, dada la conocida meticulosidad de nuestro autor, creemos que la fiabilidad de estos datos es máxima.

Los Comentarios se aplican a una ficha o a un grupo de fichas; en este último caso, el comentario se asigna al nombre de la categoría correspondiente. Los comentarios pueden estar hechos o bien por el autor (Salvador Fernández Ramírez) o por los editores (Ignacio Bosque y Maite Rivero en “Las Partículas” y el actual equipo editorial en las demás partes del fichero). Este es otro de los inestimables valores del AGLE, ya que la búsqueda de datos se ve enriquecida con la presencia de comentarios hechos por gramáticos de primera línea.

Además, el apartado de Comentarios es aprovechado por el autor y por los editores para anotar las *Referencias Cruzadas*. Cuando el autor o los editores encuentran que el fenómeno gramatical tratado en un ficha o en una categoría se relaciona con otro fenómeno gramatical recogido en el AGLE se remite a él en el Comentario. Así, por ejemplo, cuando se trata la preposición *a*, los editores nos remiten a la sección del Archivo que recoge el complemento directo⁶. Gracias a este sistema de referencias, el AGLE permite comparar construcciones sintácticas.

Por último, el Árbol de la categoría aparece junto al código de categoría (a través del cual se accede a la información sobre la categoría y sus comentarios) al que pertenece el fenómeno gramatical marcado.

⁶ Como es lógico, no podrán consultarse todavía los apartados a los que se remite cuando pertenecen a entregas futuras de la presente edición.

3. PRÓXIMAS ENTREGAS

Como se ha señalado, de las 115.000 cédulas que contiene el *AGLE*, en la actualidad solamente se ha publicado la parte correspondiente a “Las Partículas”, que supone, aproximadamente, el 15% del total del fichero. Durante el próximo año verán la luz todas aquellas cédulas incluidas en la parte de “El Verbo”, que supone la edición de más del 50% del Archivo. Las secciones de “El Nombre”, “El Pronombre” y “La Oración” se irán publicando sucesivamente. Existen además varios ficheros que contienen información sobre morfología y fonología y cuya publicación está pendiente de aceptación. Junto con la edición de las últimas partes del fichero, el equipo editorial publicará un libro con toda la información que aparece en el Archivo y que, por una u otra razón, no se recoge en la edición electrónica.

4. APLICACIONES

El principal interés del *AGLE* consiste en la posibilidad de encontrar un gran número de datos de un fenómeno tratado con detalle y amplitud, e incluso relacionarlo con otros fenómenos gramaticales conectados⁷. El sistema de referencias cruzadas permitirá al usuario establecer relación entre fenómenos aparentemente diferentes.

El *AGLE* resulta especialmente útil para ilustrar fenómenos gramaticales que no aparecen en diccionarios comunes. Por ejemplo, si un usuario necesita conocer las posibilidades de uso de un verbo como *romper* encontrará en un diccionario normal un ejemplo de uso transitivo y otro de uso reflexivo. El *AGLE* contiene gran cantidad de ejemplos reales del verbo *romper*, que además aparecen clasificados

⁷ A modo de ejemplo, un usuario al que le interese estudiar la posición y naturaleza del sujeto encontrará información sobre este en varios lugares dentro del *Archivo* perfectamente clasificados. Entre las categorías que hacen referencia a la posición del sujeto, podemos mencionar, por ejemplo, *Verbos intransitivos con sujetos animados*, *Verbos reflexivos con sujeto de cosa*, *Pasiva con se con sujeto de persona*, *Pasiva con se con sujeto postpuesto*, *Pronombres relativos en función de sujeto* y un largo etcétera. Todas estas categorías se encuentran aparentemente separadas pero, sin embargo, un sistema de referencias cruzadas permite no perder ninguna información.

gramaticalmente⁸. Por último, el *AGLE* nos permite hacernos una idea sobre la frecuencia de uso de las construcciones gramaticales⁹.

Los descubrimientos que la consulta del *AGLE* garantiza al investigador se originan en la atención prestada por Fernández Ramírez a los pequeños detalles y sobre todo en su curiosidad intelectual, que le llevaba a hacerse preguntas que en la mayor parte de las obras de referencia de nuestra tradición gramatical ni siquiera están planteadas. Conviene mencionar algunos ejemplos. Al consultar el archivo llama la atención el interés que el autor prestó a la clasificación de los verbos según el aspecto léxico, como factor importante para el estudio de los tiempos verbales. Si bien no encontraremos en el *AGLE* ninguna teoría elaborada acerca de la interacción entre clases aspectuales (estados, procesos, acciones) y valores de los tiempos, sí que es posible recopilar datos ya organizados de acuerdo con ciertas intuiciones acertadas. Hoy sabemos que dicha interacción entre contenidos es decisiva si queremos dar cuenta de por qué, y en qué casos, una forma temporal adquiere determinados valores.

⁸ Ilustraremos el citado ejemplo con cédulas del fichero; hemos seleccionado siete usos del verbo *romper* que encontramos en las siguientes categorías: *Verbos reflexivos con sujeto de cosa* y *Complemento directo, Dependiendo de la naturaleza del verbo*. A partir de las notas que aparecen en las correspondientes categorías, podemos obtener información adicional; así los *Verbos reflexivos con sujeto de cosa* incluyen una nota del autor que indica lo siguiente:

Muchas veces, los verbos transitivos se neutralizan con el pronombre para crear intransitivos con sujeto de persona; son los mismos que aparecen también en la construcción reflexiva con sujeto de cosa para significar, no la actividad más o menos deliberada y consciente de una persona, sino un movimiento, un cambio o un proceso operado en un objeto por causas o agentes desconocidos (el coche se detuvo). En ocasiones, el cambio o el proceso se presenta como una propiedad inherente de las cosas (las frutas acaban por corromperse).

Esta información nos servirá para intuir que los verbos que aparezcan bajo esta categoría podrán ser utilizados como reflexivos pero también como transitivos y, además, que en el caso de los primeros, el verbo seleccionará de forma general un sujeto de cosa, por lo que un sujeto de persona producirá una secuencia extraña a menos que se produzca un uso metafórico del verbo (*María se rompió*).

⁹ Vemos, por ejemplo, que son mucho más abundantes los usos de *ser* (277 cédulas) acompañado de frases nominales que aquellos casos en que acompañando a estas aparece *estar* (12 cédulas), lo que permite establecer generalizaciones como, por ejemplo, que los sintagmas nominales, en condiciones generales, aparecen con el verbo *ser* e, incluso, que los casos que aparecen con *estar* resultan extraños.

Otra muestra de la fertilidad del trabajo de Fernández Ramírez puede encontrarse en sus observaciones sobre los contextos y las intenciones comunicativas que se asocian al uso de las construcciones gramaticales. Un ejemplo, pequeño pero significativo, está en la sección dedicada al uso de la perífrasis *ir a + infinitivo* en enunciados exclamativos: las cédulas recogen expresiones como *¡Qué te va a conocer!* o *¡Qué los vas a molestar!*, y el autor las incluye en una sección titulada “[Contextos] De incredulidad, escepticismo y repulsa”, notando además un paralelismo con ciertos usos de *haber de + infinitivo*, caracterizados como “de repulsa indignada” (*¡Qué le ha de molestar!*). Estos usos constituyen un interesantísimo problema de análisis, ya que exigen la presencia obligada de las perífrasis mencionadas, pero no admiten fácilmente su sustitución por formas alternativas, y exhiben restricciones interpretativas realmente intrigantes (entre otras, el hecho de ser ecoicos). Las preguntas que surgen son evidentes: ¿qué relación existe entre la estructura exclamativa y las perífrasis? ¿por qué estos enunciados adquieren los valores discursivos que los caracterizan? ¿qué hace que sean ecoicos? De nuevo, no encontraremos en el *AGLE* una teoría desarrollada que responda a estas preguntas, pero sí los datos que nos permiten formularlas. Y esto representa ya un avance considerable. La consulta del archivo permitirá descubrir muchos de estos pequeños problemas, y será por lo tanto muy enriquecedora para cualquier gramático.

Esperamos poder ofrecer con el *AGLE* una herramienta de trabajo valiosa tanto para los investigadores como para profesores, estudiantes extranjeros de español o simplemente personas interesadas en la gramática de nuestra lengua. Y esperamos también ponerla a disposición de los usuarios lo antes posible.

BIBLIOGRAFÍA

- BOSQUE, I. (1997): “La búsqueda infinita. Sobre la visión de la gramática en Salvador Fernández Ramírez”, discurso de ingreso a la RAE.
- BOSQUE, I. (1995): “Introducción” y “Ayuda” de la versión web del *AGLE*.
Publicación electrónica en: <http://cvc.cervantes.es/obref/agle/>
- FERNÁNDEZ RAMÍREZ, S. (1985): *Gramática española*, 5 vol., Madrid: Arco/Libros.

POLO, J. (1998): *En torno a la obra científica de Salvador Fernández Ramírez*, Cáceres: Servicio de Publicaciones de la Universidad de Extremadura.

REAL ACADEMIA ESPAÑOLA (1973): *Esbozo para una nueva gramática de la lengua española*, Madrid: Espasa Calpe.