

LA LINGÜÍSTICA DEL CORPUS COMO HERRAMIENTA DE INVESTIGACIÓN EN ESCRITURA ACADÉMICA EN CAMPOS DISCIPLINARES ESPECÍFICOS AL ÁREA DE FORMACIÓN DOCENTE.¹

Prof. Elisa Cadario
Universidad Nacional de Río Cuarto.
Facultad de Ciencias Humanas. Dpto. de Lenguas.
ecadario@hum.unrc.edu.ar

Años de experiencia como formadores de Profesores de Inglés en la Universidad Nacional de Río Cuarto nos muestran que muchas de las dificultades con las que deben enfrentarse nuestros alumnos son muy complejas y en esencia, muy diferentes a las que deben sortear alumnos de otros profesorados en los que los contenidos disciplinares se imparten en lengua materna. A partir del momento en que comienzan la carrera, los alumnos de profesorados en lenguas extranjeras deben hacer propia la idea de que no solo deberán formarse incorporando contenidos disciplinares diversos que les son propios a su formación como docentes, sino que deberán además, lograr una excelente competencia en la lengua en cuestión para poder luego enseñarla. El desafío es grande, y mucho más aún si tenemos en cuenta que muchos de ellos llegan a la instancia de comenzar la carrera contando sólo con los conocimientos y habilidades desarrollados durante su paso por la escuela secundaria.

Este desafío se hace mucho mayor cuando comienzan a transitar las materias comúnmente llamadas “de contenido” pertenecientes, además, a áreas disciplinares muy diversas como lo son la lingüística, historia o literatura. Si tenemos en cuenta, y si dudas debemos hacerlo, que áreas disciplinares específicas muestran perfiles lingüísticos también específicos (Ej. ítems léxicos de alta frecuencia utilizados en combinaciones típicas), y que es a través del uso apropiado de la lengua que define a un campo que nuestros alumnos demuestran el conocimiento de la disciplina, entonces, el desafío para ellos, es mayúsculo.

Avanzando en la problemática no debemos dejar de mencionar que el medio, por elección o por necesidad, a través del cual nuestro alumno es evaluado es el medio escrito, y todos los educadores sabemos que esto plantea una dificultad extra. El docente en formación no solo debe conocer y profundizar contenidos conceptuales específicos a las disciplinas que aborda, sino que para comunicar sus mensajes efectivamente en las mismas debe además, manejar los discursos, y por ende, los recursos formales que los tipifican o definen. Es implica que deberá, por ejemplo, poder identificar aquellos ítems léxicos de alta frecuencia que representan a las unidades de sentido más importantes para la conceptualización del conocimiento en un dicha área y que definen el “acerca de que” (“aboutness”) se está escribiendo. (Lewis 2000 p 190) y utilizarlos en combinaciones adecuadas.

¹ El presente trabajo que se enmarca dentro de un proyecto de investigación desarrollado en el Profesorado de Inglés de la Universidad Nacional de Río Cuarto que se centra sobre el notar (“noticing”) y la habilidad de escritura en Inglés como lengua extranjera.

Sabemos que el texto escrito se caracteriza por poseer una densidad léxica muy superior al texto oral, o sea un número mayor de palabras de contenido (sustantivos, adjetivos, verbos y adverbios) que de palabras funcionales o gramaticales (auxiliares, preposiciones, pronombres, etc.) y que son precisamente las palabras de contenido, y muy especialmente dentro de este grupo, los sustantivos, los que definen el contenido epistemológico de las diferentes disciplinas o campos de conocimiento. Son, entonces, sin dudas, éstas, las que el alumno que construye un escrito académico en un área específica debe incorporar y utilizar. Estos ítems léxicos centrales y de alta frecuencia son la materia prima que le da forma al lenguaje académico y técnico en una disciplina ya que delimitan los marcos de referencia y realizan un área particular del conocimiento (Nation, 2001 p.18)

Obviamente, la habilidad de producir textos académicos en áreas disciplinares específicas no es una habilidad que pueda desarrollarse naturalmente; un correcto andamiaje, la toma de conciencia (“noticing”) y la práctica sistemática son, sin dudas, elementos esenciales en el desarrollo de esta habilidad.

Nuestros alumnos del Profesorado de Inglés en la UNRC tienen dificultades en la práctica de la escritura académica, y si bien dentro de los objetivos de la asignatura a partir de la que estamos realizando el estudio, -Didáctica General de una Segunda Lengua- no se encuentra el de desarrollar las habilidades necesarias para producir textos académicos, los mismos, son los medios a través de los cuales los alumnos deben mostrar su conocimiento de los contenidos ya sea en ensayos, monografías, planes de clases o exámenes parciales y finales. Conocer más a fondo el problema, creemos, nos ayudará a apoyarlos más sistemáticamente en esta empresa.

En referencia a las dificultades detectadas en los productos escritos de los alumnos es relevante mencionar que las mismas no se centran en la identificación de las palabras claves (lenguaje técnico o metalenguaje específico) que definen a nuestro campo disciplinar sino en su uso; en las combinaciones en que las utilizan o en la complejidad, por ejemplo, de los grupos nominales que producen a partir del sustantivo clave. A priori, en la simple exploración, vemos que sus elecciones, sin ser necesariamente erróneas a nivel gramatical o sintáctico, no se asemejan a las elecciones y /o combinaciones que realizaría el nativo experto en el área, y no reflejan, por lo tanto, el perfil léxico de la disciplina. Esto, sin duda, afecta y condiciona negativamente la manera en la que comunican sus significados.

¿Cómo aproximarnos al estudio de la problemática planteada? ¿Cómo llegar a profundizar nuestro conocimiento de este problema puntual en el desarrollo del intelenguaje de nuestro alumnos?

En la búsqueda de una metodología que nos permita responder a las preguntas planteadas, la Lingüística del Corpus se presenta como una herramienta de investigación, creemos, muy válida. El acceder a un corpus a través de un programa de concordancia nos permitirá identificar los patrones lingüísticos que caracterizan a diferentes áreas de discurso o

campos de conocimiento; palabras claves, frecuencia de uso de las mismas y combinaciones en las que diferentes actores las utilizan. Este es nuestro punto de partida.

De lo dicho se desprende que un recurso esencial para la utilización de esta herramienta de investigación es un corpus que debe ajustarse a las necesidades nuestro estudio en particular. Un corpus puede ser definido como un conjunto de textos representativos de la lengua en uso (la lengua utilizada con fines auténticos), “un cuerpo de evidencia que debe ser confiable y relevante a fines específicos” (Sinclair, en Ghadessy, 2002. p. xi).

Los corpus desarrollados para este estudio han sido creados con el fin de arrojar luz sobre algunos patrones lexico-gramaticales típicos de un área de conocimiento en particular. Existen corpus de grandes dimensiones mediante la utilización de los cuales el investigador puede obtener información relevante a la lengua en general, un corpus específico o especializado, en cambio, permitirá obtener importantes datos acerca de un género o área disciplinar en particular. El presente estudio se sustenta en la utilización de dos Corpus específicos diferentes: un corpus experto especializado y un corpus del alumno. La confección del corpus experto especializado requirió la digitalización (por escanéo óptico) de capítulos de los diferentes libros producidos por nativos expertos en el área de didáctica a los que los alumnos deben acceder como lectura obligatoria o de referencia durante el cursado de la asignatura. La confección del corpus del alumno que debía ser también digital, demandó tiempo y paciencia ya que el mismo se confeccionó a partir de la transferencia manual a documentos de texto de un gran número de producciones escritas de los alumnos a lo largo de más de tres años. Es importante mencionar que ambos corpus se actualizan en forma continua, tanto con nuevo material de lectura que agrega la cátedra, como con nuevas producciones de sucesivas cohortes.

Como ya se ha mencionado, la función específica del corpus experto que se elaboró y sobre el que se trabajará en el presente proyecto, es la de reflejar la lengua en uso producida por el nativo experto en el área disciplinar en la que se ha detectado el problema en las producciones escritas de los alumnos- la Didáctica de la enseñanza en lengua extranjera. A partir de este corpus experto de referencia, y por medio de la aplicación de software específico se identificarán las palabras claves al área mencionada –aquéllas que la definen como campo epistemológico-, y los patrones en los que las mismas emergen en el corpus experto. El acceso al corpus del alumno a través de las mismas herramientas en el programa de concordancia permitirá por medio de la comparación y el análisis, explorar similitudes y diferencias entre las elecciones lexico-gramaticales que realizan ambos- el alumno y el nativo experto- para expresar significados similares específicos al área mencionada.

De lo dicho anteriormente se desprende naturalmente que el otro elemento necesario para investigar bajo los fundamentos de la lingüística del corpus es un programa de concordancia. En nuestro estudio accederemos a la data a través del programa *Wordsmith Tools* desarrollado por Mike Scott (1999) para ser usado en el tipo de computadora personal a través de la cual

cualquier estudiante de lenguas o investigador pueda acceder a un corpus digital, ya sea éste construido con fines específicos o de acceso gratuito en variados sitios de Internet. El programa mencionado ofrece herramientas que permiten realizar diferentes tareas. La herramienta “*Concord*” (Concordancia) permite localizar todas las referencias a una palabra o frase específica que se desee estudiar en un corpus, presentándola en líneas de concordancia con la palabra nodal centrada y un número variable de palabras a cada lado de la misma. Esta herramienta permite examinar y describir combinaciones típicas y recuperar de los datos, los patrones o frases en las que la misma se utiliza en los diferentes corpus. La función *WordList*, (listas de palabras) permite obtener, como su nombre lo indica, una lista de todas las palabras que se encuentran en el corpus, ordenándolas alfabéticamente o por frecuencia. La frecuencia de uso de una palabra o frase determinada es básica para la identificación de las palabras claves (*Key*) a campo discursivo.

Para el presente estudio se seleccionarán dentro del grupo de palabras de contenido un número de palabras de alta frecuencia en ambos corpus; y dentro de este grupo se partirá del sustantivo en tanto éste funciona como núcleo del grupo nominal. Esta elección se hace en base a dos premisas: que es precisamente el grupo nominal quien más información contribuye al “acerca de que” se está escribiendo (campo) y que cuanto mayor sea la complejidad en la construcción del grupo mayor será la cantidad de información contenida en el mismo. El grupo nominal se analizará a partir de la descripción del mismo que realiza la Gramática Sistemática funcional que identifica como sus elementos constitutivos a demostrativos, numerativos, epítetos, y clasificadores como pre-modificadores del sustantivo, y calificadores en la función de pos-modificadores del mismo. Intentaremos ver y comparar como utiliza en el experto y el alumno estos elementos constitutivos de la frase nominal y en que se diferencian o asemejan los grupos nominales que producen Sustantivos tales como “enfoque” (approach), tarea (task), y estrategias (strategies), son fundamentales al comunicar mensajes en el área disciplinar y serán el punto de partida para este estudio.

Sin lugar a dudas el presente trabajo es limitado dado que en primera instancia el estudio se circunscribe a la descripción del grupo nominal. Sin embargo, la posibilidad de contar con un dos corpus específicos a nuestras necesidades y problemas hace que los posibles aspectos a estudiar en el futuro (Ej. Descripción del grupo verbal, colocación, etc) sean muchos y variados.

A nivel de aplicación pedagógica y como objetivo a largo plazo, se espera a partir del análisis que se plantea en este trabajo y otros posteriores y de la sistematización de los resultados obtenidos, poder elaborar una propuesta pedagógica que involucre el auto-acceso de los alumnos al corpus experto especializado a través de actividades que lo ayuden a desarrollar conciencia sobre los usos idiomáticos de la lengua en el área disciplinar ya mencionada.

Sin dejar de reconocer las dificultades prácticas que acarrea el trabajar sobre corpora producida por estudiantes, acordamos ampliamente con Flowerdew (en Ghadessy 2001 p. 364),

quien resalta la importancia de contar con cuerpos de evidencia confiable que reflejen la lengua en uso y “en desarrollo” de nuestro alumno- su interlengua- ya que el estudio de la misma aparece como un elemento fundamental para la comprensión de sus necesidades y dificultades.

BIBLIOGRAFÍA

CARTER, R. (1994). *Vocabulary, Applied Linguistic Perspectives*. Routledge: London and New York.

GHADESSY, M. HENRY, A & ROSEBERRY, R. (2001). *Small Corpus Studies in ELT. Theory and Practice*. Studies in Corpus Linguistics. John Benjamins Publishing Company. Amsterdam / Philadelphia.

HATCH, E. & BROWN, Ch. (1995). *Vocabulary, Semantics and Language Education*. Cambridge Language Teaching Library. Cambridge University Press.

HAYMAN, J. (1981). *Investigación y Educación*. Ediciones Paidós. Barcelona.

NATION, I.S.P. (2001). *Learning Vocabulary in another Language*. (series Editors: Michael Long and Jack Richards) Cambridge Applied Linguistic Series. Cambridge University Press.

LEWIS, M. (2000). *Teaching Collocation. Further developments in the lexical Approach*. LTP. Teacher Training.

LÓPEZ MORALES, H. (1994). *Métodos de Investigación Lingüística*. Ediciones Colegio de España.

SINCLAIR, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

SCHMITT, N. (2000). *Vocabulary in Language Teaching*. Cambridge University Press.

SCHMITT, N & McCARTHY, M. (1997). *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge Language Teaching Library. Cambridge University Press.

STUBBS, M. (2002). *Text and Corpus Analysis: Computer assisted studies of Language and Culture*. Studies in Corpus Linguistic Series. John Benjamins Publishing Company. Amsterdam /Philadelphia

Artículos electrónicos

GROOT, P. (2000). Computer Assisted Second Language Vocabulary Acquisition. *Language Learning and Technology*, Vol. 4, No.1, 60-81. en www.llt.msu.edu

CURADO FUENTES, A. (2001). *Lexical Behaviour in Academic and Technical corpora: implications for ESP Development* Vol. 5, No. 3, pp. 106-129 en www.llt.msu.edu