

CONSTRUCCIÓN DE UN CORPUS DE ARTÍCULOS DE SEMI-DIVULGACIÓN: ASPECTOS TEÓRICOS Y METODOLÓGICOS

Verónica Muñoz

vmunoz@hum.unrc.edu.ar

Universidad Nacional de Río Cuarto – Becaria CONICET

1. Introducción

Las diferentes facetas que caracterizan al área de la enseñanza de lenguas extranjeras han motivado a lo largo del tiempo importantes cambios en la diversidad y rigurosidad de las investigaciones (Hinkel, 2005). Esta variedad a la que se refiere Hinkel (2005) se ha reflejado en la expansión de la pluralidad de contextos y objetivos de investigación, en nuevas dimensiones de estudio, y en el desarrollo e innovación de nuevas metodologías de investigación. Todo ello en pos de responder a nuevos contextos de aprendizaje, nuevos alumnos, nuevas necesidades de aprender una lengua extranjera, y la consecuente necesidad de profundizar el conocimiento sobre las lenguas como así también de conocer y mejorar los procesos de enseñanza y aprendizaje. En cuanto a las metodologías de investigación, diversos enfoques han aportado contribuciones a partir de una gran variedad de diseños y técnicas de recolección y análisis de datos. Una metodología de investigación que ha ganado protagonismo en las últimas décadas es la lingüística de corpus (Sinclair, 2004), cuya riqueza radica en que sus técnicas y procedimientos pueden explotarse para responder una amplia variedad de preguntas de investigación en diferentes áreas, tales como la lingüística, la lingüística aplicada, la sociolingüística, y la psicolingüística (Lüdeling y Kytö, 2008; Meyer, 2004). Las preguntas que se pueden responder utilizando la metodología de corpus varían desde el análisis de elementos microtextuales, como la gramática y el vocabulario, hasta elementos macrotextuales, como la organización de textos y mecanismos discursivos. Además del alcance que el enfoque tiene para abarcar una amplia variedad de objetivos de investigación atravesando diversas áreas, su mayor contribución es que permite el estudio empírico de grandes bases de datos lingüísticos fundados en el uso real de la lengua, superando de este modo las limitaciones de antiguas técnicas basadas en la intuición de los investigadores (Baker, 2006, 2010; Conrad, 2005; Kennedy, 1998; Sinclair, 1991, 2005).

En este trabajo se introducirán brevemente aspectos y conceptos básicos que caracterizan este enfoque como metodología de investigación empírica de la lengua y se describirán los lineamientos para la construcción sistemática de un corpus. En segundo término, se describirán los pasos seguidos para la construcción de un corpus de artículos de semi-divulgación científica del área de las ciencias agrarias. Dicho corpus fue construido para responder a las necesidades de un curso de lectura en inglés destinado a alumnos de Ingeniería Agronómica. El corpus se construyó para estudiar el vocabulario que caracteriza al artículo de semi-divulgación en dicha disciplina, con el objetivo de profundizar el conocimiento sobre este género y de este modo optimizar la enseñanza de lectura en inglés.

2. Lingüística de corpus

2.1 Contextualización

La lingüística de corpus es una metodología de investigación relativamente nueva, cuyos orígenes se remontan a las décadas del 60 y 70, con la creación de los corpus *Brown* y *LOB*, a partir de los cuales el enfoque se comenzó a desarrollar como una nueva manera de abordar el análisis lingüístico (Gavioli, 2005). En 1988 Biber publica el libro *Variation across speech and writing*, en el que estudia similitudes y diferencias entre el inglés escrito y el inglés hablado, utilizando técnicas y herramientas provistas por la computadora para el análisis de diferentes géneros. Esta publicación es una de las primeras que utiliza esta nueva metodología de análisis lingüístico. Pocos años más tarde, Sinclair (1991) publica el libro *Corpus Concordance Collocation*, en el que puntualiza sobre esta nueva manera de describir la lengua detallando sus diferentes técnicas y sus nuevas formas de recolección, exploración y presentación de

datos. Con su libro, Sinclair marca un momento crucial en el desarrollo de este marco metodológico, estableciendo las bases sobre las que la lingüística de corpus comenzaría a desarrollarse como una rigurosa metodología de investigación. Otras publicaciones como las de Barnbrook (1996), McEnery y Wilson (1996), Biber, Conrad y Reppen (1998), Kennedy (1998), y Hunston (2002) fueron importantes contribuciones para seguir definiendo la lingüística de corpus como metodología. Estas publicaciones acentuaron un aspecto fundamental que diferencia la lingüística de corpus de otros marcos metodológicos: la posibilidad de cuantificar el uso de la lengua utilizando métodos estadísticos para el análisis de textos. En la actualidad, dos décadas después de la publicación del libro *Corpus Concordance Collocation*, la lingüística de corpus se ha establecido como una metodología de investigación sólidamente fundada, revolucionando las técnicas y herramientas que han sido tradicionalmente utilizadas para el procesamiento de datos lingüísticos (Ådel y Reppen, 2008; Renouf y Keohe, 2009).

Los estudios de corpus son cada vez más significativos no sólo en la enseñanza de lenguas sino también en áreas tales como la lexicografía, la terminología, y la traducción. El ímpetu de esta metodología de investigación se debe a las nuevas formas de evidencia lingüística que ofrece y a sus novedosas técnicas, procedimientos, y herramientas para la recolección y análisis de grandes cantidades de datos lingüísticos, lo que permite realizar descripciones más sistemáticas sobre tendencias de uso, es decir, sobre lo que es típico y habitual en determinados contextos (Biber *et al.*, 1998; Meyer, 2004; Conrad, 2005; Granger, 2002; Kennedy, 1998; Sinclair, 1991; Stubbs, 2004). A continuación se describirán aspectos básicos relacionados con la construcción de un corpus.

2.2 El corpus

Los estudios enmarcados en esta metodología tienen como objeto de análisis la lengua representada en un corpus. La palabra *corpus* es de origen latín y significa *cuerpo*. En principio, por lo tanto, un corpus podría concebirse simplemente como un mero grupo de textos. Sin embargo, tal como oportunamente lo señaló Sinclair (2001), un grupo de textos adquiere el estatus de corpus cuando se compila sistemáticamente bajo ciertos parámetros, lo que implica un aspecto metodológico fundamental en los estudios de corpus. También es importante agregar que un corpus difiere de un simple conjunto de textos porque es creado para ser sometido a análisis lingüístico (Meyer, 2004). Tradicionalmente, un corpus se ha definido como un conjunto o “cuerpo” de textos auténticos en formato electrónico que se seleccionan y se agrupan sistemáticamente bajo ciertos criterios para ser analizados lingüísticamente. Como se observa, tres características fundamentales emergen de esta definición. Primero, un corpus es estructurado y planificado, en oposición a un conjunto de textos sin un orden o una lógica determinada (Barnbrook, 1996). Segundo, un corpus es un conjunto de textos auténticos creados con un propósito comunicativo real, es decir textos que no son procesados pedagógicamente para el estudio o la enseñanza de la lengua (Conrad, 2005). Tercero, un corpus es una muestra diseñada para representar una totalidad, como por ejemplo un determinado dialecto (Biber, 2008). Como se observa en estas tres características, la sistematicidad que caracteriza la estructura de un corpus permite arribar a conclusiones y realizar generalizaciones sobre patrones lingüísticos a partir de evidencia empírica de la lengua tal como es utilizada en situaciones reales (Barnbrook, 1996; Hunston, 2002; Sinclair, 2005).

2.2.1 Criterios para el diseño y compilación de un corpus

Como se mencionó anteriormente, un corpus es una muestra, es decir un recorte o una representación de una población más amplia sobre la que se harán generalizaciones (Biber, 2008). Es por esta razón que, como lo señala Sinclair (1991), la preparación sistemática de un corpus y la precisión de su estructura interna son aspectos muy importantes que determinan la calidad de los resultados de una investigación. Por ello, los investigadores deben diseñar la estructura de sus corpus definiendo claramente la variedad de la lengua, géneros y registros que se estudiarán, y determinando los procedimientos que se usarán para la selección de los textos (Biber, 2008; Sinclair, 1991). En esta etapa de planificación, que debe ser considerada cuidadosamente, es necesario crear lo que O’keeffe, McCarthy y Carter (2007: 1) denominan “matriz de diseño”, es decir un sistema de muestreo que permita refinar la selección de textos. Para esto, se deben describir sistemáticamente los géneros y registros que serán objeto de análisis, basándose en variables no lingüísticas, las cuales incluyen aspectos relacionados con

la producción y recepción de textos en contextos específicos (Biber, 2008; Kennedy, 1998; Sinclair, 2005). Los géneros han sido definidos como textos que se caracterizan por tener una determinada organización y convenciones lingüísticas, y son creados para llevar a cabo actividades a través de las cuales los individuos se comunican rutinariamente en contextos culturales específicos (Swales, 1990; Martin, 1992; Miller, 1984). Es decir que esta definición se basa en la concepción del discurso como una actividad socialmente contextualizada. Este énfasis en el rol social de los géneros en contextos culturales destaca el propósito comunicativo de los textos como un aspecto principal (Askehave y Swales, 2001; Bhatia, 1993, 2001; Miller, 1984). Además de definir el propósito comunicativo de los textos, es necesario determinar otras variables que caracterizan el contexto de comunicación en niveles más específicos. Para ello los investigadores deben delimitar el registro, definido desde el marco de la gramática sistémico funcional como una variedad de la lengua determinada por tres variables del contexto específico de situación: tenor, modo y campo (Halliday, 1989). La variable tenor se refiere a la relación social entre los participantes y sus roles. El modo se conceptualiza como el rol de la lengua en la situación de comunicación y el canal de comunicación, siendo la lengua oral parte de la situación y la lengua escrita una reconstrucción de la misma. La variable campo se define como las actividades sociales en las que los participantes están involucrados, e incluye principalmente el tema o dominio de los textos. Es pertinente aclarar que los criterios de selección de textos, tanto a nivel de género como de registro, estarán siempre condicionados por la disponibilidad de acceso a los mismos (Hunston, 2002; Stubbs, 2004).

Una vez que se definen los géneros y registros, es importante determinar la estructura interna de un corpus basándose en ciertos criterios para que los resultados sean confiables y generalizables (Stubbs, 2004). A continuación se presentará una síntesis de criterios sugeridos por diferentes autores, principalmente aquellos propuestos por Sinclair (1991, 2005), Meyer (2004), Biber (2008) y Pearson (1998). Dos conceptos, representatividad y balance, son fundamentales para la construcción de un corpus. Como se mencionó anteriormente, un corpus intenta ser representativo de una totalidad, por lo tanto debe reunir características similares a la lengua que éste representará (Barnbrook, 1996). La representatividad se define como el grado en que una muestra incluye la totalidad de variables de una población (Biber, 2008). Por otra parte, un corpus se considera balanceado cuando está estructurado en diferentes secciones que contienen iguales proporciones de ciertas variables, como por ejemplo número de palabras (Hunston, 2002; Kennedy, 1998; Sinclair, 2005; Stubbs, 2004). La representatividad y el balance de un corpus no son siempre objetivos que se pueden definir y lograr completamente (Sinclair, 2005). Sin embargo, se han propuesto ciertos parámetros que deben considerarse para la composición de un corpus: tamaño, disponibilidad de textos en formato electrónico, número de textos, extensión de los textos, uso de textos completos, período, y características demográficas de los individuos que producen los textos.

El tamaño de un corpus, determinado por el número de palabras, depende de los objetivos de la investigación para la que se construye un corpus, la metodología usada (Barnbrook, 1996; Sinclair, 2001, 2005), la disponibilidad de recursos electrónicos, y el tiempo del que se dispone para la recolección de textos (Meyer, 2004). Hunston (2002) señala que un corpus que contiene entre medio y un millón de palabras se considera usualmente pequeño. Sin embargo, la distinción entre corpus grandes y corpus pequeños no es siempre clara (Gavioli, 2005). Sinclair (1991) inicialmente sugirió que un corpus debe ser tan grande como sea posible. Luego, Sinclair (2001) observó que el tamaño de un corpus no determina la calidad de un estudio, ya que los corpus pequeños son apropiados para estudios que tienen objetivos de investigación bien específicos, contrariamente a los estudios que utilizan corpus grandes para el análisis de fenómenos lingüísticos más generales (Kennedy, 1998; Sinclair, 1991). Por ejemplo, Hunston (2002) y Kennedy (1998) observan que un corpus pequeño es adecuado cuando se analiza vocabulario de alta frecuencia. Otra ventaja de utilizar corpus pequeños es que éstos permiten establecer relaciones entre el corpus y el contexto específico en el cual los textos fueron producidos (Koester, 2010). Además, como sugiere Meyer (2004), los corpus pequeños generalmente son más adecuados para ser abordados manualmente, mientras que los corpus grandes son más apropiados para ser procesados automáticamente por la computadora.

Otros aspectos que ayudan a lograr la representatividad de un corpus son el número de textos, la extensión de los textos, y el uso de textos completos. En relación al número y extensión de

los textos, Biber (2008) y Meyer (2004) observan que la representación lingüística se puede lograr de una manera más adecuada incluyendo en un corpus muchos textos cortos y no pocos textos extensos. En relación a la extensión de los textos, Sinclair (2005) señala que no se justifica desde un punto de vista lingüístico seleccionar textos de igual extensión. Como observa Sinclair (1991), en lugar de recolectar textos de igual tamaño, es necesario asegurarse que los textos de un corpus sean documentos completos, los cuales reflejen la extensión típica de los textos que el corpus intenta representar.

Otros dos parámetros que contribuyen a la representación de un corpus son el período o tiempo en el cual los textos fueron producidos y las características demográficas de las personas que produjeron los textos. Con respecto a la variable período, Sinclair (1991) sostiene que es importante recolectar textos que han sido producidos dentro de un determinado tiempo. Kennedy (1998) y Meyer (2004) sugieren que esto es particularmente relevante en el caso de los corpus sincrónicos, los cuales representan la lengua de un período en particular. En relación con las características demográficas de las personas que produjeron los textos, Sinclair (1991), Meyer (2004), y Biber (2008) mencionan aspectos tales como las relaciones sociales entre los participantes, el idioma que hablan como lengua materna, la región geográfica de residencia, la edad, el grupo étnico, género, clase social, nivel de educación, y empleo. Por esto, una amplia variedad de usuarios puede ser representada en un corpus si éste incluye muchos textos producidos por diferentes personas en lugar de pocos textos producidos por pocas personas. Esto es importante para evitar arribar a conclusiones sesgadas por los estilos individuales de las personas (Biber *et al.*, 1998). Una vez que los criterios de identificación de los géneros y registros han sido establecidos, la siguiente etapa involucra la selección y recolección de los textos.

2.2.2 Tipos de corpus

Existe una amplia variedad de tipos de corpus, los cuales son más o menos apropiados para cada estudio dependiendo de los objetivos de investigación. Algunos investigadores utilizan corpus disponibles, es decir aquéllos que ya han sido construidos, tales como el *Lancaster-Oslo/Bergen Corpus*, el *British National Corpus*, *MICASE* y *COBUILD*. Otros investigadores utilizan corpus diseñados para responder preguntas de investigación específicas. Los corpus, según Hunston (2002), pueden clasificarse en diversos tipos: generales, especializados, comparables, paralelos, de alumnos, pedagógicos, de monitoreo, sincrónicos, y diacrónicos.

Una importante diferencia se establece entre los corpus generales construidos con objetivos múltiples y los corpus especializados construidos con objetivos específicos. Un corpus general para objetivos múltiples es un corpus construido para representar toda una lengua o una variedad geográfica. Este tipo de corpus contiene una variedad de géneros y disciplinas, es construido para una variedad de propósitos, y es usualmente más grande que un corpus especializado. Por el contrario, un corpus especializado para objetivos específicos, como lo define Sinclair (1991), representa una variedad específica de la lengua, incluye generalmente un rango limitado de géneros específicos, y restringe el objetivo del análisis para responder preguntas de investigación específicas, tales como la identificación de características particulares de un género o de una disciplina.

Los corpus comparables consisten de dos o más sub-corpus que representan diferentes idiomas y están diseñados bajo los mismos parámetros. Por otro lado, existen corpus paralelos, que consisten de dos o más corpus de textos traducidos en diferentes idiomas. Ambos tipos de corpus son utilizados generalmente por traductores que tienen como objetivo establecer comparaciones entre lenguas. Otro tipo de corpus es el corpus de alumnos, que incluye textos producidos por estudiantes y es empleado para identificar diferencias y similitudes entre alumnos de diferentes lenguas. En cuanto a los corpus pedagógicos, éstos representan la lengua a la que se expone a un determinado grupo de alumnos en un contexto particular. Los corpus de monitoreo, por su parte, son corpus que se actualizan y se agrandan constantemente y son usados para identificar cambios en el uso de la lengua a través del tiempo. Por otra parte, los corpus sincrónicos se construyen para representar una variedad de la lengua en un momento definido, mientras que los corpus diacrónicos se utilizan para estudiar el desarrollo de elementos lingüísticos específicos y por lo tanto incluyen textos de diferentes períodos.

3. Construcción del corpus CEPAA

A continuación se ilustrará el diseño y construcción de un corpus especializado, denominado CEPAA (*Corpus of English Popularization Articles in Agriculture*), en español *Corpus de Artículos de Semi-divulgación de las Ciencias Agrarias*. El corpus es representativo de un género (artículos de semi-divulgación), un registro (científico), una disciplina (ciencias agrarias), un tema específico (producción de granos), y un idioma (inglés). Para diseñar el corpus, se consideraron dos criterios principales, balance y representatividad. El balance del corpus descrito en este trabajo, el CEPAA, está determinado por su estructura interna homogénea ya que, como se explicó anteriormente, es específico con respecto al género, el registro, la disciplina, el tema de los textos, y el canal de comunicación. La representatividad del CEPAA se consideró siguiendo los lineamientos desarrollados en el apartado 2.2.1 de este trabajo: diversidad de fuentes de publicación, disponibilidad de textos en formato electrónico, período, tamaño, uso de textos completos, variedad de escritores, y dialecto.

En cuanto a la diversidad de fuentes de publicación y la disponibilidad de textos en formato electrónico, los artículos fueron recolectados de doce revistas de divulgación publicadas online por once universidades de los Estados Unidos (Tabla 1). Para lograr una selección sistemática de las revistas y los textos, se procedió a consultar la sección “editorial” de revistas publicadas por el área de extensión de diferentes universidades. Se seleccionaron revistas que respondían a las características que definen al artículo de semi-divulgación como género, considerando específicamente los siguientes criterios: fuente de publicación, contexto de circulación de las publicaciones, objetivo de comunicación, emisores y audiencia. Dichos criterios fueron determinados de acuerdo a Ciapuscio y Kuguel (2002) y Gallardo (1998), quienes identifican el artículo de semi-divulgación como un género que se distingue del artículo de investigación publicado en revistas altamente especializadas y del artículo de divulgación publicado en diarios y revistas de interés general. En cuanto a la fuente de publicación, los artículos de semi-divulgación se publican en revistas de divulgación científica con cierto nivel de especialización. Dichos artículos circulan tanto en instituciones académicas y de investigación como en ciertos sectores específicos de la sociedad. Estos artículos tienen como objetivo informar sobre resultados científicos, recursos, herramientas y nuevos desarrollos. Los emisores de los artículos de semi-divulgación son investigadores, especialistas y docentes universitarios dedicados a la investigación, con ocasional revisión de editores o divulgadores. Estos artículos están dirigidos a lectores semi-legos, definidos por Ciapuscio (2003) como aquellos lectores que tienen un grado intermedio de competencia entre el lector experto y el lector lego. Es decir que los lectores semi-legos representan una audiencia más amplia que las comunidades científicas que leen artículos de investigación altamente especializados, y una audiencia más delimitada que el público general y masivo que lee artículos de divulgación en diarios y revistas. Los lectores semi-legos pueden ser alumnos universitarios con formación en una determinada disciplina, profesionales, personas con cierta experiencia en una disciplina, e investigadores especialistas en áreas de conocimiento diferentes o relacionadas.

Las revistas seleccionadas para la construcción del CEPAA son publicadas por programas de extensión de las universidades, los cuales realizan investigaciones para desarrollar y ofrecer soluciones en respuesta a necesidades locales. Las revistas, de circulación dentro el ámbito de las universidades y fuera de éstas, tienen como objetivo transmitir información basada en investigaciones, comunicando resultados de estudios, y ofreciendo recomendaciones y recursos asociados con diferentes actividades de extensión. En cuanto a los autores de los artículos, la mayoría de los textos seleccionados fueron escritos tanto por docentes universitarios involucrados en diferentes investigaciones como por especialistas involucrados en actividades de extensión. Algunos textos fueron escritos por los editores de las revistas. La audiencia de las revistas son estudiantes universitarios con formación en el área de las ciencias agrarias, especialistas, y no especialistas con cierta experiencia en la disciplina.

Las revistas seleccionadas tienen diferentes frecuencias de publicación: quincenal, bimestral, mensual, y semanal. Por lo tanto, el número de artículos seleccionados de cada revista varió en relación a su frecuencia de publicación. Todos los textos se extrajeron de las ediciones correspondientes a los períodos 2008 y 2009 para asegurar que los resultados reflejaran el uso

del vocabulario en un determinado período. En cuanto al tamaño del corpus y al uso de textos completos, se recolectaron 700 artículos de diferentes extensiones, sumando un total de 455.366 palabras. Los artículos se guardaron en su formato original (PDF y/o HTML). Luego, cada texto se guardó como archivo Word. Las versiones Word de los textos fueron preparadas para ser procesadas por el software eliminando tablas, fotos, figuras, ilustraciones y gráficos, como así también nombres, afiliaciones e información personal de los autores, referencias, agradecimientos, títulos de las tablas, leyendas, y encabezamientos. Finalmente, éstos se guardaron en formato TXT para que puedan ser procesados por el software *WordSmith Tools* 4.0. Cada archivo se rotuló con un número para su identificación en el corpus.

Tabla 1 – Revistas, instituciones que publican, y número de textos usados para el corpus

Revista	Institución que publica	Número de textos usados para el corpus
1 - C.O.R.N. Newsletter	<i>The Ohio State University</i>	76
2 - Integrated Crop Management News	<i>Iowa State University</i>	74
3 - CropWatch Newsletter	<i>University of Nebraska</i>	83
4 - Field Crop News	<i>The Pennsylvania State University - PennState</i>	85
5 - Plant and Soil Sciences Extension Newsletter	<i>Oklahoma State University</i>	65
6 - Agronomy e-updates	<i>Kansas State University</i>	94
7 - Pest&Crop Newsletter	<i>Purdue University</i>	58
8 - Crop & Pest Report	<i>North Dakota State University</i>	59
9 - Weekly Crop Update	<i>University of Delaware</i>	28
10 - Kentucky Pest News Newsletter	<i>University of Kentucky</i>	32
11 - Louisiana Agriculture Magazine	<i>Louisiana State University</i>	27
12 - The Plant Disease Alert Newsletters	<i>Kansas State University</i>	19
Número total de textos		700

Otro criterio usado para lograr la representatividad del CEPAA fue la variedad de escritores. Los 700 artículos fueron escritos por un total de 292 autores. Por último, para contemplar el criterio de dialecto, los textos se seleccionaron de revistas publicadas por universidades de los Estados Unidos, que representan la variedad del inglés americano usado en artículos de semi-divulgación de las ciencias agrarias. Para una mejor representación de la variedad, se seleccionaron universidades de once estados diferentes: Ohio, Iowa, Nebraska, Pensilvania, Oklahoma, Kansas, Indiana, North Dakota, Delaware, Kentucky, y Louisiana (Figura 1).

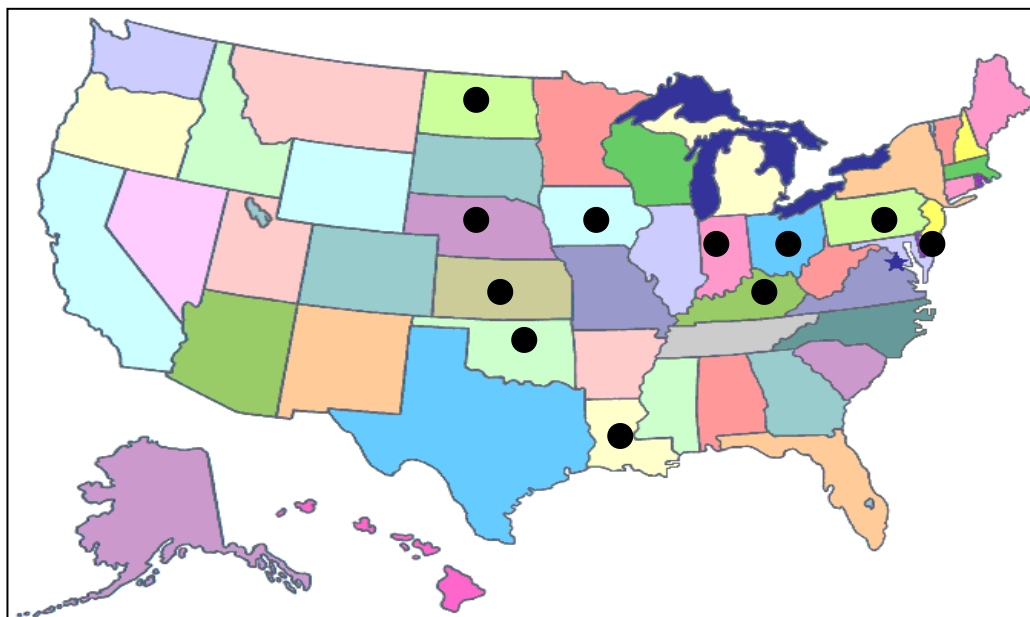


Figura 1 – Estados representados en el CEPAA

En la Tabla 2 se presenta una síntesis sobre el diseño del CEPAA (adaptado de Biber, 2008).

Tabla 2 – Descripción del CEPAA

	CEPAA
Género	Artículos de semi-divulgación
Registro	Científico
Disciplina	Ciencias agrarias
Tema	Producción de granos
Fuentes de publicación	Doce revistas publicadas online por once universidades de Estados Unidos
Período (fecha de publicación)	2008-2009
Tamaño	700 textos - 455,366 palabras
Escritores	Principalmente docentes universitarios - 292 escritores
Lectores	Estudiantes universitarios, especialistas, y no especialistas con cierta experiencia en el área de las ciencias agrarias
Dialecto	Inglés americano en artículos de semi-divulgación de las ciencias agrarias
Contexto de circulación de las revistas	Institución-público
Canal de comunicación	Escrito
Formato	Publicado
Factualidad	Información factual
Propósito comunicativo	Informar, instruir, explicar

4. Conclusión

En este trabajo se realizó una breve introducción a la lingüística de corpus como metodología de investigación y se describieron los lineamientos para la construcción sistemática de un corpus especializado, el CEPAA. El mismo fue construido para analizar el vocabulario de artículos de semi-divulgación de las ciencias agrarias. Este corpus permitió definir las características de un género en particular que los alumnos de Ingeniería Agronómica de la Universidad Nacional de Río Cuarto necesitan leer. A modo de conclusión, es importante destacar que la posibilidad de explorar géneros específicos a través de la construcción de un corpus, como el descrito en este trabajo, permite identificar los elementos lingüísticos que caracterizan un determinado género y que deben enseñarse en contextos particulares para optimizar el proceso de enseñanza-aprendizaje. Tal es el caso de los cursos de Inglés con Fines Específicos que tienen como objetivo asistir a los alumnos en el desarrollo de la habilidad lectora.

5. Referencias

- Ädel, A., y Reppen, R. (2008). The challenges of different settings: An overview. En A. Ädel, y R. Reppen (Eds.), *Corpora and Discourse. The challenges of different settings* (pp. 1-6). Amsterdam/Philadelphia: John Benjamins Publishing.
- Anthony, L. (2007). AntConc: a Multi-Platform Corpus Analysis Toolkit (Versión 3.2.1) [Software]. Disponible online: <http://www.antlab.sci.waseda.ac.jp/software.html>
- Askehave, I., y Swales, J. (2001). Genre identification and communicative purpose: a problem and a possible solution. *Applied Linguistics*, 22(2), 195-212.
- Baker, P. (2006). *Using corpora in discourse analysis*. London/New York: Continuum.
- Baker, P. (2010). Corpus methods in linguistics. En L. Litosseliti (Ed.), *Research methods in linguistics* (pp. 93-113). London/New York: Continuum.
- Barnbrook, G. (1996). *Language and computers. A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- Bhatia, V. (1993). *Analyzing genre: language use in professional settings*. Essex: Longman.
- Bhatia, V. (2001). Analysing genre: some conceptual issues. En S. M. Hewing (Ed.), *Academic writing in context* (pp. 79-92). Birmingham: University of Birmingham Press.
- Biber, D. (1988). *Variation across speech and writing*. New York: Cambridge University Press.
- Biber, D. (2008). Representativeness in corpus design. En T. Fontenelle (Ed.), *Practical lexicography. A reader* (pp. 63-87). Oxford: Oxford University Press.
- Biber, D., Conrad, S., y Reppen, R. (1998). *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- Ciapuscio, G. (2003). Formulation and reformulation procedures in verbal interactions between experts and (semi-) laypersons. *Discourse Studies*, 5(2), 207-233.
- Ciapuscio, G., y Kuguel, I. (2002). Hacia una tipología del discurso especializado: aspectos teóricos y aplicados. En J. García Palacios, y M. T. Fuentes (Eds.), *Entre la terminología, el texto y la traducción* (pp. 37-73). Salamanca: Almar.
- Conrad, S. (2005). Corpus linguistics and L2 teaching. En E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 393-409). Mahwah: Lawrence Erlbaum Associates.
- Gallardo, S. (1998). Estrategias y procedimientos de reformulación en textos de Divulgación Científica. *Revista de la Sociedad Argentina de Lingüística*, 1, 67-79.
- Gavioli, L. (2005). *Exploring corpora for ESP learning*. Amsterdam: John Benjamins Publishing Company.
- Granger, S. (2002). A Bird's-eye view of learner corpus research. En S. Granger, J. Hung, y S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam: John Benjamins Publishing Company.
- Halliday, M.A.K. (1989). Register variation. En *Language, context, and text: aspects of language in a social-semiotic perspective* (pp. 29-43). Oxford: Oxford University Press.
- Hinkel, E. (2005). Introduction. En E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 375-391). Mahwah: Lawrence Erlbaum Associates.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- Koester, A. (2010). Building small specialized corpora. En A. O'Keeffe, y M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 66-79). London: Routledge.
- Lüdeling, A. y Kytö, M. (2008). Introduction. En A. Lüdeling, y M. Kytö (Eds.), *Corpus linguistics. An International handbook* (Vol. I, pp. v-xii). Berlin: Walter de Gruyter.
- Martin, J. R. (1992). *English text. System and structure*. Filadelfia: John Benjamins Publishing Company.
- McEnery, T. y Wilson, A. (1996). *Corpus linguistics. An introduction*. Edinburgo: Edinburgh University Press.
- Meyer, C. (2004). *English corpus linguistics. An introduction*. Cambridge: Cambridge University Press.
- Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech*, 70, 151-167.
- O'Keeffe, A., McCarthy, M., y Carter, R. (2007). *From corpus to classroom. Language use and language teaching*. Cambridge: Cambridge University Press.
- Pearson, J. (1998). *Terms in context*. Amsterdam: John Benjamins Publishing Company.
- Renouf, A., y Kehoe, A. (2009). Introduction. En A. Renouf, y A. Kehoe (Eds.), *Corpus Linguistics: Refinements and Reassessments* (pp. 1-6). Amsterdam/New York: Rodopi.
- Scott, M. (1992). WordSmith Tools (Versión 4.0) [Software]. Oxford: Oxford University Press.
- Scott, M. y Tribble, C. (2006). *Textual patterns. Key words and corpus analysis in language education*. Amsterdam: John Benjamins Publishing Company.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2001). Preface. En M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT. Theory and practice* (pp. vii-xv). Amsterdam: John Benjamins Publishing Company.
- Sinclair, J. (2004). Introduction. En J. Sinclair (Ed.), *How to use corpora for language teaching* (pp. 1-10). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Sinclair, J. (2005). Corpus and text: basic Principles. En M. Wynne (Ed.), *Developing linguistic corpora: a guide to good practice* (pp. 1-16). Oxford: Oxbow Books: Consultado el 9 de Diciembre del 2009, de <http://ahds.ac.uk/linguistic-corpora/>.
- Stubbs, M. (2004). Language corpora. En A. Davies, y C. Elder (Eds.), *The handbook of applied linguistics* (pp. 106-132). Malden: Blackwell Publishing.
- Swales, J. M. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.