

# La construcción del Corpus EspaDA-UNCuyo<sup>1</sup>

## Objetivos y proyecciones

Iris Viviana Bosio, Carolina Ana Sacerdote  
Universidad Nacional de Cuyo

In this work, we present the EspaDA-UNCuyo Corpus to share findings and future directions of our research. The project covers corpus design stage. Its general aim is to establish external criteria in order to systematize a representative and balanced sample of academic written and spoken discourse in Spanish at UNCuyo (Mendoza, Argentina). Among the most relevant findings, we established external criteria and started a compilation of authentic communicative events. In this presentation, we will make connections between the external criteria system and the performance of each criterion categories in the initial stage of sample compilation.

**Keywords:** Specialized Corpora, Academic Written Spanish, Academic Spoken Spanish, EspaDA-UNCuyo Corpus

### 1. Introducción

En este trabajo presentamos el diseño del Corpus EspaDA-UNCuyo, para compartir resultados y proyecciones a la luz de los objetivos propuestos en nuestra investigación y su grado de cumplimiento. Como objetivo general, planteamos establecer criterios externos (Sinclair 2004) para sistematizar una muestra representativa y equilibrada del discurso académico escrito y oral en español de la Universidad Nacional de Cuyo (Mendoza, Argentina). Esta sistematización se orienta al diseño de un modelo de datos para la futura implementación informática del corpus.

Entre los objetivos específicos, propusimos comenzar la compilación de eventos comunicativos de circulación efectiva en nuestra Universidad y evaluar

---

<sup>1</sup> Proyecto financiado por la Secretaría de Ciencia, Técnica y Posgrado (Universidad Nacional de Cuyo, Mendoza, Argentina), durante el periodo 2016-2018.

la operatividad de categorías utilizadas por modelos probados y exitosos de construcción de corpus académicos. Entre los resultados, establecimos una serie de criterios externos que ha permitido comenzar con la recolección y el registro de los eventos, y ha dado lugar a fructíferas discusiones teóricas y metodológicas sobre clasificación de géneros y conceptualización continua o discreta de la oralidad y la escritura, por ejemplo. La representatividad de los eventos ingresados también ha sido objeto de discusiones y toma de decisiones pendientes, por su diversidad y denominación en función de las disciplinas, unidades académicas y niveles del sistema educativo de la UNCuyo, entre otros factores.

## 2. Resultados del proyecto

En este apartado relacionaremos el sistema preliminar de criterios externos con la operatividad de las categorías correspondientes a cada criterio en la compilación inicial de la muestra; y aquí es pertinente hacer algunas aclaraciones. Primero, si bien la determinación de criterios externos es una instancia inicial del diseño de un corpus orientada a la compilación, es necesaria una muestra provisoria para probar su operatividad; de esta primera muestra provienen los resultados que presentamos. Segundo, dado que no podemos describir exhaustivamente en estas páginas las problemáticas surgidas de la puesta a prueba, presentaremos los criterios externos acordados y nos detendremos solo en los más relevantes en cuanto al trabajo de contextualización en la UNCuyo y a la línea de investigación en discurso académico-científico sostenida ininterrumpidamente por nuestro equipo, desde el año 2000, en proyectos acreditados por la Secretaría de Ciencia, Técnica y Posgrado (UNCuyo)<sup>2</sup>.

---

<sup>2</sup> "Estrategias de comprensión y producción de discurso científico en estudiantes universitarios"- SeCTyP- 2000-2002 (código: 06/G188); "Procesos de comprensión y producción de discurso científico en estudiantes universitarios. Herramientas para su escritura"- SeCTyP- 2002-2004 (código: 06/G263); "Discurso académico-científico: Estrategias de lectura y escritura de textos e hipertextos"- SeCTyP- 2004-2007 (código: 06/G337); "La tesis como clase textual prototípica de los procesos de producción de textos académico-científicos. Estrategias de asistencia a su escritura"- SeCTyP- 2007-2009 (código: 06/G418); "Procesos estratégicos de producción de discursos académico-científicos: la constelación discursiva proyecto-tesis-defensa de tesis"- SeCTyP- 2009-2011 (código: 06/G497) ; "El continuum escrituralidad-oralidad en Discursos Académico- Científicos: la defensa de tesis"- SeCTyP- 2011-2013 (código: 06/G576); "Intersubjetividad y oralidad en géneros de investigación y discusión del ámbito científico-académico"- SeCTyP- 2013-2015 (código 06/G685).

## 2.1 Registro y organización de la muestra

El diseño de un formulario de Google Docs ha permitido poner a prueba los criterios externos, ya que cada una de las respuestas que se debe completar al ingresar un evento corresponde a un criterio. En algunas respuestas usamos siglas de tres letras, para hacer transparentes las etiquetas y facilitar la tarea del usuario. A continuación presentamos los criterios, aclarando si las respuestas son opciones predefinidas o campos de completamiento manual:

- a. Fecha del evento (calendario automático).
- b. Español (opciones predefinidas): Español Lengua Materna (ELM), Español Lengua Segunda (ELS); Español Lengua Extranjera (ELE). Si bien la totalidad de los eventos ingresados a la fecha corresponden a hablantes ELM, esperamos incorporar eventos ELS y ELE.
- c. Medio (opciones predefinidas): Oral / Escrito. A diferencia de la mayoría de corpus académicos publicados, EspaDA-UNCuyo contiene eventos orales (48,4%) y escritos (51,6%), dado que ambos constituyen nuestro ámbito de estudio en discurso académico.
- d. Modo (opciones predefinidas): Monomodal / Multimodal. Si bien asumimos que, en principio, ningún discurso es monomodal, incorporamos esta opción para cumplir con la categorización binaria de criterios externos (Sinclair 2004). Entendemos el modo *multimodal* como el que comprende codificaciones lingüísticas y no lingüísticas.
- e. Tipo de evento comunicativo (opciones predefinidas): Los eventos comunicativos orales y escritos consensuados hasta el momento para incorporar al corpus los presentamos en la Tabla 1 y en la Tabla 2, respectivamente. Para el registro de los eventos orales, se trabajó sobre la propuesta de MICASE –Michigan Corpus of Academic Spoken English– (Simpson-Vlach & Leicher 2009), complementada y contextualizada en relación con los géneros que efectivamente circulan en la UNCuyo:

**Tabla 1.** Eventos comunicativos orales

<b>Tipo de evento</b>	<b>Sigla</b>	<b>Descripción</b>
Clase	CLA	Interacción áulica de docentes y estudiantes de grado o posgrado en el proceso de construcción de conocimientos
Presentación áulica	PRA	Exposiciones de alumnos, puestas en común
Grupo de estudio	GRE	Grupo pequeño de alumnos, graduados y/o profesores que se reúne regularmente en la Universidad para estudiar

Ponencia en eventos académico-científicos	POE	Exposición breve (aproximadamente 20') individual o grupal de los resultados de una investigación. Incluye la ronda de preguntas
Presentación de póster en eventos	PPE	Interacción del autor de un póster con la audiencia
Conferencia	CON	Exposición extensa individual en un evento académico
Panel	PAN	Exposición grupal, panel de discusión, mesa redonda, foro, coloquio, etc. en un evento académico
Defensa de tesis	DET	Exposición individual final de una tesis. Incluye coloquio con jurados
Defensa de proyecto	DEP	Exposición individual final de un proyecto. Incluye ronda de preguntas
Reunión de trabajo	RET	Grupo de alumnos, graduados y/o profesores que se reúne en la universidad para trabajar en temas académicos
Horario de consulta	HOC	Interacción informal entre profesor y alumno/s en relación con el espacio curricular
Examen parcial de asignatura	EPA	Evaluación de los contenidos de un espacio curricular, desarrollados en un período determinado del ciclo lectivo
Examen final de asignatura	EFA	Evaluación que permite la acreditación de un espacio curricular
Reunión de Consejo	REC	Sesión de miembros del Consejo superior (gobierno universitario) o de Consejos directivos (gobierno de las unidades académicas)

En cuanto a los eventos comunicativos escritos, se comenzó a trabajar sobre la propuesta de BAWE –British Academic Written English– (Gardner & Nesi 2013; Heuboeck, Holmes & Nesi 2010; Hernández Navarro & Castelló 2014), pero aun los resultados de clasificación y contextualización son preliminares, hasta tanto se ingrese mayor diversidad de géneros escritos que permita discutir su caracterización:

**Tabla 2.** Eventos comunicativos escritos

<b>Tipo de evento</b>	<b>Sigla</b>	<b>Descripción</b>
Informe de investigación	IDI	Presentación y discusión de resultados de una investigación
Informe de visita a obra	IVO	Descripción técnica de una visita a una obra en construcción
Tesis de maestría	TEM	Trabajo final de una carrera de maestría
Tesis de doctorado	TED	Trabajo final de una carrera de doctorado
Tesis de grado	TEG	Trabajo final de una carrera de grado universitario

Programa de asignatura	PRA	Instrumento curricular que especifica la labor de docentes y alumnos en una asignatura
Artículo de revista	ART	Artículos publicados en revistas de unidades académicas
Revisión bibliográfica	REB	Compendio de antecedentes de investigaciones sobre una temática específica
Ordenanza	ORD	Normativa general perdurable en el tiempo, emitida por Consejo Directivo, Consejo Superior, Decano o Rector
Resolución	RES	Normativa particular, emitida por Consejo Directivo, Consejo Superior, Decano o Rector

- f. Versión (opciones predefinidas, solo para eventos escritos): Final, Preliminar 1, Preliminar 2, Preliminar 3. Incluimos este criterio porque las versiones de un mismo evento proveerán datos para el análisis del proceso de producción y contribuirán al diseño de estrategias didáctico-pedagógicas.
- g. Número de participantes (opciones de rangos predefinidos): Solo para eventos orales.
- h. Sexo (opciones predefinidas).
- i. Edad (opciones de rango predefinidas).
- j. Número de autores (opciones predefinidas, solo para eventos escritos): Individual / Grupal.
- k. Número de hablantes (campo de completamiento manual): Solo para eventos orales.
- l. Grado de interacción (opciones predefinidas, solo para eventos orales): Altamente monológico, Mayormente monológico, Mayormente interactivo, Altamente interactivo, Mixto.
- m. Rol y nivel académico (opciones predefinidas): Este criterio se trabajó a partir de MICASE (Simpson-Vlach & Leicher 2009) y luego se contextualizó según el relevamiento de eventos orales y escritos. En la Tabla 3 presentamos las categorías incorporadas hasta el momento:

**Tabla 3.** Rol y nivel académico

<b>Rol y nivel</b>	<b>Sigla</b>	<b>Descripción</b>
Estudiante de grado inicial	EGI	Estudiantes de 1° y 2° año
Estudiante de grado avanzado	EGA	Estudiantes de 3° y 4° año
Estudiante de posgrado inicial	EPI	Estudiantes de diplomaturas y especializaciones
Estudiante de posgrado avanzado	EPA	Estudiantes de maestrías y doctorados
Docente inicial	DOI	Docentes adscriptos, ayudantes y jefes de trabajos prácticos

---

Docente avanzado	DOA	Docentes titulares, asociados y adjuntos
Investigador inicial	INI	Becarios, alumnos
Investigador avanzado	INA	Investigadores con título de maestría o doctorado, investigadores categorizados en programas acreditados
Administrativo	ADM	Personal en función administrativa o de gestión
Institucional	INS	Personal en función de gestión institucional (Rector, Decano, Secretario, Director, etc.)
Mixto	MIX	Participantes con distintos roles y niveles académicos

---

- n. Unidad Académica (opciones predefinidas): La UNCuyo cuenta con 15 unidades académicas. Si bien hasta el momento hemos registrado mayor variedad y cantidad de eventos (27,4%) en FFYL (Facultad de Filosofía y Letras) debido a que aquí se encuentra radicado nuestro proyecto, hemos comenzado a incorporar eventos en FCA (Facultad de Ciencias Agrarias), FCEN (Facultad de Ciencias Exactas y Naturales), FED (Facultad de Educación), FCE (Facultad de Ciencias Económicas), FO (Facultad de Odontología) y FI (Facultad de Ingeniería)<sup>3</sup>.
- o. Carrera (completamiento manual): solo para eventos de docencia.
- p. Disciplina y subdisciplina (completamiento manual): Para su codificación, el formulario enlaza el “Listado de códigos de disciplinas desagregadas” del sistema SIGEVA-CONICET<sup>4</sup>.
- q. Asignatura (completamiento manual): Solo para eventos de docencia.
- r. Evento sometido a evaluación (opciones predefinidas): Sí / No.
- s. Calificación (opciones predefinidas): Solo para eventos evaluativos de grado y posgrado, según la reglamentación vigente en la UNCuyo.

### 3. Proyecciones de la investigación

A partir de los resultados presentados, entre las proyecciones de esta investigación destacamos los siguientes aspectos: (1) la necesidad de avanzar en la homologación de criterios que se plasmen claramente en la identificación, el encabezado y la anotación de cada evento incorporado al corpus; (2) la

---

<sup>3</sup> Hemos respetado las siglas oficiales de la UNCuyo, razón por la cual en el formulario las siglas correspondientes a algunas unidades académicas no siguen el criterio de tres letras acordado en el proyecto.

<sup>4</sup> El Sistema Integral de Gestión y Evaluación (SIGEVA) es una aplicación desarrollada por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) que gestiona el proceso de otorgamiento de subsidios a proyectos y becas de investigación.

focalización en resolver la problemática de la representatividad y el equilibrio de la muestra, fundamentalmente respecto de los criterios *Medio* (evaluación de la relación cualitativa y cuantitativa entre eventos orales y escritos), *Tipo de evento* (posibilidad de formar familias de géneros y de considerar géneros emergentes), *Disciplina y Unidad Académica* (incorporación de mayor número y diversidad de eventos no relacionados con disciplinas ni unidad académica de radicación de este proyecto).

## Referencias

- Gardner, S. & Nesi, H. 2013. A classification of genre families in university student writing. *Applied Linguistics* 34(1): 25-52. DOI:10.1093/applin/ams024.
- Hernández-Navarro, F. & Castelló, M. 2014. Análisis de los géneros de escritura académica, de acuerdo con la percepción de los estudiantes universitarios. *Innovación Educativa* 14(65): 61-80. <http://bit.ly/2oQmaHV> (consultado Mayo 17, 2018).
- Heuboeck, A., Holmes, J. & Nesi, H. 2010. *The BAWE Corpus Manual (Version III)*. Publicado por Universidad de Coventry. <http://bit.ly/2MjIreo> (consultado Junio 18, 2018).
- Simpson-Vlach, R. & Leicher, S. 2009. *The MICASE Handbook. A Resource for Users of the Michigan Corpus of Academic Spoken English*. Ann Arbor: The University of Michigan Press.
- Sinclair, J. 2004. Corpus and Text - Basic Principles. En M. Wynne (ed), *Developing Linguistic Corpora: a Guide to Good Practice*. Publicado por Oxbow Books. <http://bit.ly/2AOFVsi> (consultado Junio 18, 2018).