

# Implementación digital del corpus EspaDA-UNCuyo

## Desde la transcripción al desarrollo de software

Acosta, Nicolás<sup>1</sup>    Espeche, Paula<sup>2</sup>

<sup>1-2</sup> Facultad de Filosofía y Letras  
Universidad Nacional de Cuyo  
Mendoza, Argentina

<sup>1</sup>nmac.1996@gmail.com, <sup>2</sup>pau\_espe20@hotmail.com

III Congreso Internacional de la Asociación Argentina de Humanidades Digitales  
7 – 9 de noviembre de 2018

# Objetivo de la presentación

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

Buscamos destacar la importancia del trabajo horizontal entre el lingüista computacional y el transcriptor, en el marco de la construcción del corpus EspaDA-UNCuyo.

# ¿Qué es un corpus?

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- Definición (Sinclair, 2004).
- Enfoque.
- Noción de corpus para el proyecto.

# Información del proyecto

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- **Título del proyecto:** Construcción de un corpus de discurso académico escrito y oral de la UNCuyo con metodología de lingüística de corpus. Proyecciones hacia ámbitos científicos y didáctico-pedagógicos.
- **Directora:** Dra. Iris Viviana Bosio.
- **Co-directores:** Prof. Esp. Carolina Sacerdote, Dr. Luis Aguirre.
- **Marco institucional:** Proyecto de Investigación de la SIIP.

# Descripción del proyecto

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- **Problema:** No existen corpus de Discurso Académico en español.
- **Hipótesis:** Los estudios del lenguaje se benefician de evidencias empíricas de diversas fuentes para contribuir al desarrollo acumulativo del conocimiento científico.
- **Objetivo:** Construir un corpus de español académico en el marco de la producción de la UNCuyo con distintos tipos de anotación (fonética, morfológica, léxico-semántica, pragmática y discursiva).
- **Metodología:** Recolección de textos orales y escritos en formato digital escrito (textos producidos de manera escrita digital y transcripción), observación de otros modelos de corpus de textos académicos en inglés.

# Características del corpus

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- Corpus especializado en géneros académicos.
- Versatilidad del corpus para abarcar diversas áreas de la investigación lingüística.
- Disponibilidad de los textos en formato plano y para consulta a través de una interfaz web.

# Importancia de la metodología de lingüística de corpus

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- Representación empírica de la lengua en uso.
- Procesamiento digital de los textos en base a criterios lingüísticos.
- Posibilitar la realización de análisis cualitativos y cuantitativos de rasgos lingüísticos (Gries, 2015).
- Brindar a la comunidad científica resultados acumulativos y replicables.

- ¿Qué es la transcripción?
  - Representación de elementos fonéticos, fonológicos, léxicos o morfológicos de una lengua o dialecto mediante un sistema de escritura.
  - Foco en la transcripción desde el medio oral al escrito.

# La transcripción como una labor subjetiva

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- Noción de la transcripción como un simple proceso de transferencia (Tilley & Powick, 2002).
- "Transcribir datos orales es una tarea laboriosa y delicada que requiere entrenamiento y práctica. Es donde comienza el tratamiento de los datos e implica tomar una serie de decisiones que forman ya parte del análisis que lo afectan [...] Según cuáles sean los objetivos del análisis, se puede optar por unos criterios u otros, por unas convenciones u otras" (Calsamiglia Blancafort & Tusón Valls, 2001, p. 357).
- La transcripción como un acto de poder (Bucholtz, 2000).
  - Decisiones sobre el contenido.
  - Decisiones sobre la forma.

# Trabajo con las etiquetas

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- "El reciente crecimiento del uso de las nuevas tecnologías de la información está obligando también a la filología y a la lingüística a adaptarse con la nueva forma de hacer ciencia, enfocándolas naturalmente desde su propia perspectiva" (Fernández, 2015, p. 24).
- Transcripciones del tipo ortográficas.
- Fenómenos de la oralidad representadas: interrupciones, solapamientos, extranjerismos, etc.
- Familiarización con las etiquetas.
- Límites de la transcripción: toma de decisiones frente a qué representar.

# Programación de la interfaz

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

**Interfaz web**

Conclusiones

Referencias

- Perl y módulo CGI.
- Diseño orientado a la consulta web.

# Traducción de los metadatos a HTML

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

Etiquetas y metadatos	Código HTML
<code>\t&lt;[A-Z]{3}=[^&gt;]+&gt;</code>	<code>&lt;b&gt;CATEGORÍA:&lt;/b&gt; ETIQUETA</code>
<code>^&lt;S([0-9] D-[fm])&gt;</code>	<code>&lt;b&gt;(Sx SD-x):&lt;/b&gt;</code>
<code>&lt;SOLAP&gt;...&lt;/SOLAP&gt;</code>	<code>&lt;font color=#1922d7&gt;...&lt;/font&gt;</code>
<code>&lt;INT S="Sx"&gt;...&lt;/INT&gt;</code>	<code>&lt;font color=#ef8609&gt;...&lt;/font&gt;</code>
<code>&lt;EXTR&gt;...&lt;/EXTR&gt;</code>	<code>&lt;i&gt;...&lt;/i&gt;</code>

# Método de indización del corpus

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- Extracción de los metadatos.
- Índice de los textos para búsqueda.

# Demostración del funcionamiento

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- Buscador del corpus.

# Conclusiones

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- La importancia de continuar y promover el trabajo colaborativo entre el lingüista computacional y el transcriptor.
- La necesidad de generar una metodología de transcripción clara.
- Documentar los sesgos personales de la transcripción: creación de bitácoras de transcripción.
- Evaluación del prototipo.
- Proyecciones hacia una versión final del *software*.

# Referencias

Corpus  
EspaDA-  
UNCuyo

Acosta, N. &  
Espeche, P.

Objetivo

Corpus  
EspaDA-  
UNCuyo

La  
transcripción

Interfaz web

Conclusiones

Referencias

- Bucholtz, M. 2000. The politics of transcription. *Journal of Pragmatics* 32(2000). 1439–1465.
- Calsamiglia Blancafort, Helena & Amparo Tusón Valls. 2001. *Las cosas del decir*. Barcelona: Ariel.
- Fernández, Patricia. 2015. La creación de corpus lingüísticos diacrónicos: la perspectiva del transcriptor. *Enlce: Revista Venezolana de Información, Tecnología y Conocimiento* 13(3). 23–47.
- Gries, Stefan Th. 2015. Quantitative designs and statistical techniques. En Douglas Biber & Randi Reppen (eds.), *The Cambridge Handbook of English Corpus Linguistics*, 50–72. Cambridge: Cambridge University Press.
- Sinclair, John. 2004. Corpus and text – basic principles. En Martin Wynne (ed.), *Developing linguistic corpora: A guide to good practice*, 1–16. Oxford: Oxbow Books.
- Tilley, S. A. & K. D. Powick. 2002. Distanced data: Transcribing other people's research tapes. *Canadian Journal of Education* 27(2/3). 291–310.

## ¡Muchas gracias!

Acosta, Nicolás<sup>1</sup> & Espeche, Paula<sup>2</sup>

<sup>1-2</sup> Facultad de Filosofía y Letras  
Universidad Nacional de Cuyo  
Mendoza, Argentina

<sup>1</sup>nmac.1996@gmail.com, <sup>2</sup>pau\_espe20@hotmail.com