



PONTIFICIA UNIVERSIDAD
CATOLICA
DE VALPARAISO



Lingüística de Corpus, Computacional y Procesamiento del Lenguaje Natural en la Pontificia Universidad Católica de Valparaíso

Dr. René Venegas
rene.venegas@ucv.cl

Dr.(c) Rodrigo Alfaro
rodrigo.alfaro@ucv.cl



*IngLing2011 I Workshop Virtual
en Ingeniería Lingüística 15 de
Diciembre 2011-11:30 hrs.*

Agenda

1. Lingüística de Corpus en la Escuela Lingüística de Valparaíso
2. Lingüística Computacional (Software e Interfaz de Interrogación Lingüística)
3. Procesamiento del Lenguaje Natural
 - Análisis Semántico Latente y Clasificación de discurso científico-académico
4. Desarrollos en la Escuela de Ingeniería Informática

Lingüística de Corpus

Algunas Investigaciones

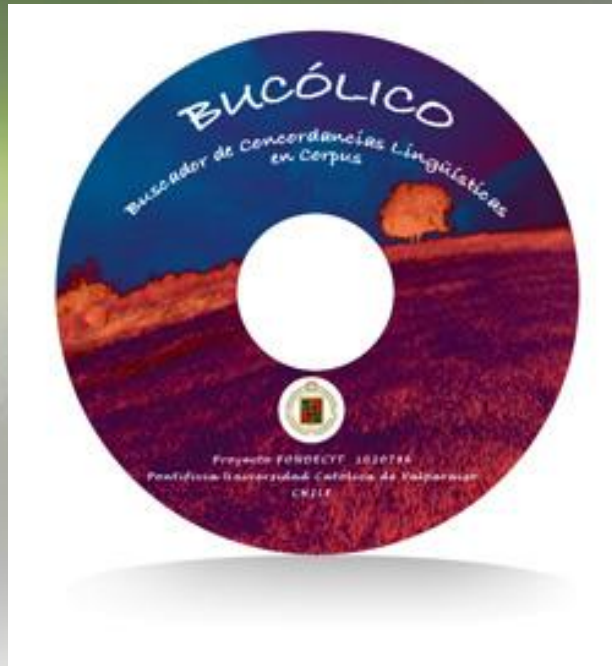
- Fondecyt 1020786: “El análisis del discurso científico y su comprensión lingüística en la formación técnico-profesional” (2003-2005; IR G. Parodi)
- Fondecyt 1060440: “El discurso especializado escrito en el ámbito universitario y profesional: Lingüística de Corpus y Análisis Multidimensional” (2006-2009; IR G. Parodi)
- Fondecyt 1020786 “Evaluación automatizada de resúmenes” (2009-2010; IR R.Venegas)
- Fondecyt 1090030: “Géneros académicos a nivel de doctorado: desde su descripción a la construcción de un perfil multidimensional de la comprensión disciplinar escrita” (2010-2012; IR G.Parodi)
- Fondecyt 1101039: “Caracterización de géneros evaluativos como trabajos finales de grado en licenciatura y magíster a través de cuatro disciplinas: Desde los patrones léxico-gramaticales y retórico-estructurales al andamiaje de la escritura académica disciplinar” (2010-2013; IR R.Venegas)

Lingüística de Corpus

Investigadores de la Escuela Lingüística de Valparaíso

- Dr. Giovanni Parodi (Director de Postgrado en Lingüística)
- Dr. René Venegas
- Dr. Romualdo Ibáñez
- Dr. Cristian González
- Otros doctore(a)s en otras líneas de investigación y un conjunto amplio de ayudantes y tesisistas de licenciatura, magíster y doctorado.

Lingüística Computacional



elgrial
Escuela Lingüística de Valparaíso

ELV Postgrados **El Grial** Revista Signos LECTES Tutor MUSA

Sumérgete en el corazón del Grial
Herramientas de búsqueda

Búsqueda Multiniveles El Manchador

El Corpus
Mapa de Corpus

Comparación de textos

[ver más](#)

www.elgrial.cl

Lingüística computacional

The image shows a screenshot of the website for the Escuela Lingüística de Valparaíso. At the top left is the logo 'elv' with the text 'Escuela Lingüística de Valparaíso'. Below it are navigation links: 'ELV', 'Postgrados', 'El Grial', 'Revista Signos', 'LECTES', and 'Tutor MUSA' (highlighted in an orange speech bubble). On the right side, there are links for 'contacto', 'sitemap', and 'portada'. The main content area features a dark purple background with a green field and a table with chairs. The text 'Este es tu lugar de inspiración' is displayed above the table. A red button with a play icon is located at the bottom left of the main content area.

Presentación

Es una herramienta computacional, desarrollada en el seno del centro de Escritura Académica de la Escuela Lingüística de Valparaíso, cuyo propósito es asistir, a través de la Internet, en la escritura científica y académica a los estudiantes de postgrado. El proceso de escritura es guiado a través de la organización estructural de los textos y es asistido por un tutor especialista, quien retroalimenta los avances del escritor.

Este es tu lugar de inspiración

Te invitamos a descubrir tu musa

Accede aquí a tu cuenta Tu-Musa para conocer a tu lugar de inspiración

Proyecto TU-MUSA: Objetivos

- Motivar y asistir la escritura de géneros académicos a nivel de postgrado en lingüística.
- Diseñar e implementar un sistema tutorial en línea, de tipo asincrónico, con soporte en web para la asistencia de la escritura de artículos de investigación científica (AIC) en lingüística.

Estoy en mi **Escritorio**, revisando

¿Tengo mensajes de mi Tutor?
Envío mis mensajes

Necesitas inspiración

Bienvenido
Rene Venegas
Ver mis trabajos
Editar mi perfil
Modificar

Me voy a casa

Aquí estoy > Mis Documentos

Estos son mis Documentos

Ver Tipos Artículos

Nuevo Documento

En qué estado están?

Personalizar

Integrar

Proyecto 123

En espera



Doc Cristian

En desarrollo



Artículo de Corpus

Vacio



REDES SISTÉMICAS

Inicio

• Encabezado

• Introducción

• Cuerpo del trabajo

• Cierre

• Referencias Bibliográficas

> Título

> Resumen
(250-300 palabras)

> Palabras Clave
(4 o 5 conceptos)

- Propósito
- Mención a las bases teóricas
- Metodología
- Resultados
- Conclusiones



Visualización de la red Sistémica

Conozca cómo se constituye la red sistémica de los tipos de artículos de investigación científica.

Procesamiento del Lenguaje Natural

- Análisis Semántico Latente

el Grial

Acerca de | Quiénes somos | Qué Hacemos |
Dónde trabajamos

interfaz de etiquetaje e interrogación de corpus textuales



Búsqueda Simple

Búsqueda Compleja

El Manchador

Visualización del Out-Put
del análisis

Comparación de textos

Interfaz para el cálculo

Similitud léxico semántica en español.

Seleccione un Espacio

Agregar Comparación

Comparar textos

Exportar a Excel



Examinar

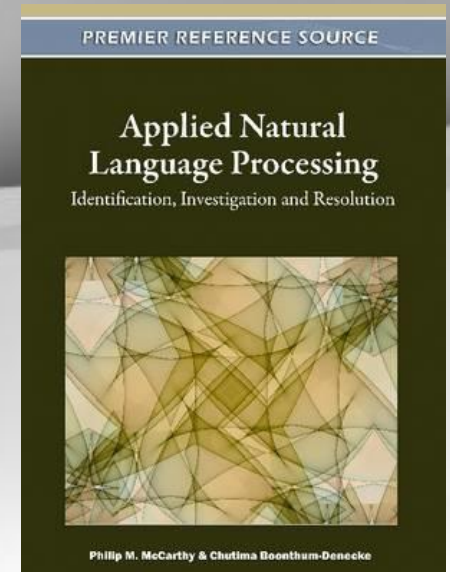
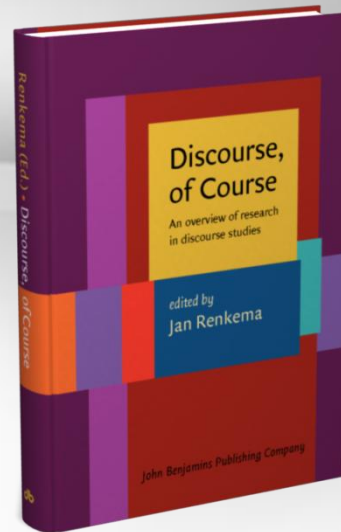
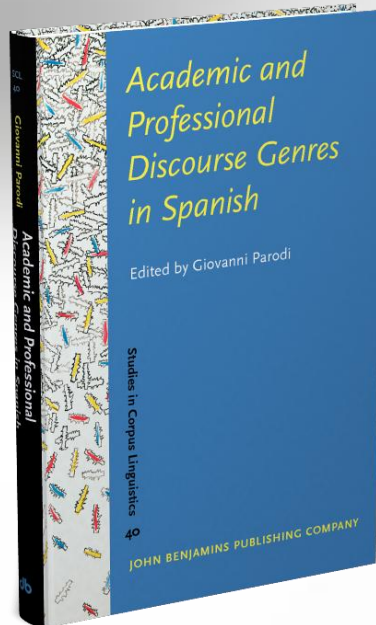
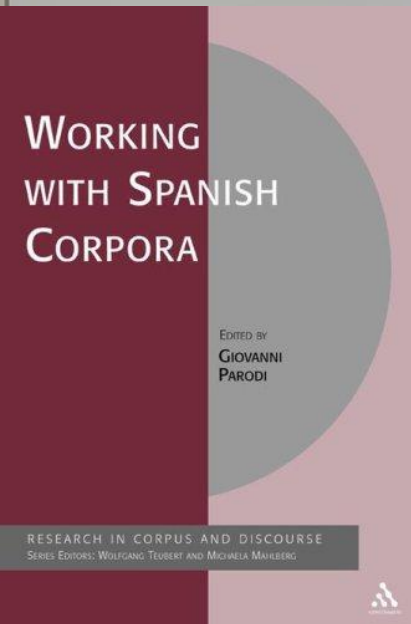
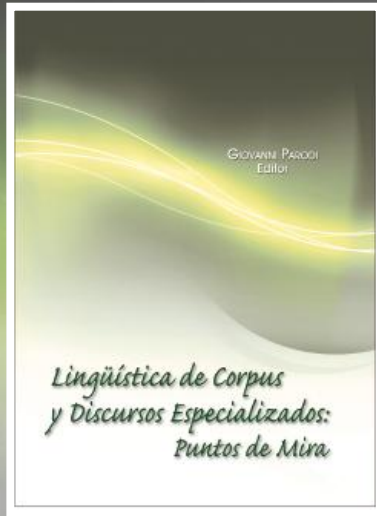
Examinar

INVESTIGACIONES CON LSA

- Comparación de la similitud semántica de Resumen, Palabras Clave, Introducción y Cuerpo del AIC en Ciencias Sociales, Exactas y Naturales (Venegas, 2005).
- Evaluación automatizada de resúmenes de textos expositivos y narrativos (Venegas, 2007, 2009, 2011).
- Diseños de prueba de comprensión en base a LSA (Venegas & Ibáñez, 2011)
- Perfil de Coherencia en Discursos Autoritarios (Venegas, 2011).

Clasificación

- **Clasificación de textos académicos en función de su contenido léxico-semántico (Venegas, 2007)→ Bayes/SVM**
- **Clasificación de Tesis según disciplinas y géneros según Dimensión Informativa (Venegas,2010, 2011)→ Análisis Discriminante**
- **Clasificación de género y disciplinas usando trigramas POS (Tesis/TD/AIC)→Bayes (Venegas, 2011)**



Desarrollo en la Escuela de Ingeniería Informática PUCV

- Clasificación automática de textos, en sus variantes de etiqueta simple y multi-etiquetado, a través de algoritmos de Máquinas de Aprendizaje
- Pre y postgrado: diversas tesis y proyectos de título
- Máquinas de Aprendizaje: Máquina de soporte vectorial, redes neuronales artificiales, bayes ingenuo, vecino más cercano

Aplicaciones de clasificación de noticias, tweets, datos médicos, entre otros, en inglés y español

- Generación de taxonomías de mensajes de tweet
- Clasificador de noticias de interés
- Generación de resúmenes automatizados
- Estimación de índices a partir de noticias

- Investigaciones acerca de diferentes representaciones de textos, medidas de similitud y de cómo éstas afectan el rendimiento de las diferentes máquinas de aprendizaje
- Medidas de similiaridad difusas: fuzzy logic
- Modificación de textos para problemas de multietiqueta
- Selección de características

Otras acciones...

- Creación de la empresa Analitic S.A., dedicada a la generación de información a partir de textos digitales, apoyada con financiamiento CORFO-Innova
- Contratación de un doctor en redes complejas, con financiamiento CONICYT

Centro de Análisis Automatizado del Lenguaje Natural

- Centro interdisciplinar para el estudio del uso del lenguaje natural y para el desarrollo de aplicaciones automatizadas.
- Actividades:
- Integración de los distintos programas de formación e investigación afines
- Establecimiento de relación con organizaciones institucionales y empresariales
- Organización de encuentros científicos y de transferencia de conocimientos
- Workshop en julio 2012