

ReLiSCo Rede galega de Recursos lingüísticos
para unha Sociedade do Coñecemento

Xornada
Universidade - Empresa
Minería de Opinión



Universidade de Vigo



Lo más importante

- Para qué estamos aquí
 - para hacer red, hablando de la Red y de la lengua en la Red
 - Red ReLisCo
 - Lengua en la Red
 - *políticos de Sudán* → *políticos no sudan* (Bitext)
 - *quiero comprar billete para una persona obesa/ adquirir ticket para pasajero gordo/ para persona con sobrepeso* (Inbenta)
 - papel de recursos lingüísticos en el procesamiento de contenidos en la web
 - qué se dice en la web, de qué se habla
 - Minería de opinión
 - interés en conocer la opinión de los usuarios sobre productos y servicios del mercado: *Que hablen de uno es espantoso. Pero hay algo peor: que no hablen.* (Oscar Wilde)
 - palabras con polaridad positiva: *limpieza*
 - inversión de la polaridad: *limpieza nefasta, limpieza brilla por su ausencia*
 - palabras con polaridad negativa: *problema*
 - inversión de la polaridad: *superar un problema*

ReLiSCo

Lo segundo más importante

- Qué vamos a hacer hoy

Xosé Ramón Pichel (Imaxin)

11.30: Pausa café

Enrique Torrejón (Bitext)

Francisco Rangel (Autoritas)

14.00: Almuerzo en el Auditorio

Judith Sastre (Inbenta)

Iván Pino (Llorente & Cuenca)

Diego Cobelo (Classora)

21.00: Cena en Restaurante Acio

ReLiSCo

De qué voy a hablar

- Presentación de las **empresas invitadas**
 - Coruña
 - Imaxin, Classora
 - Madrid
 - Bitext, Autoritas, Llorente & Cuenca
 - Barcelona
 - Inbenta
- Presentación de **ReLiSCo** (Xunta de Galicia CN2011/006): 6 grupos de la universidad gallega
 - UVigo
 - Gigralex, Griles, COLE
 - USC
 - GE, FILGA
 - UdC
 - LyS

ReLiSCo

GIGRALEX

(Grupo de investigación en Gramática y Léxico)

<http://gigralex.uvigo.es>

Coordinador: J.M. García-Miguel

GIGRALEX

Grupo de lingüistas de la Facultad de Filología e Traducción

Líneas de investigación principales

- Análisis sintáctico y semántico
- Construcciones verbales y estructura argumental
- Morfología derivativa y morfología genética
- Lexicología y lexicografía
- Redes semánticas / conceptuales
- Ideología en léxico y en discurso

Proyectos de investigación

- ADESSE / ADESSE II / ALEXSYS
- Red de Morfología Genética en el NDHE

Transferencia y creación de empresas: *Evágora* (<http://evagora.es/>)

ReLiSCo

Objetivos Globales

- Anotación sintáctico-semántica y léxica de un **corpus del español** para la consulta de:
 - **Dependencias sintácticas y Papeles semánticos**
 - **Relaciones léxicas:** sinonimia, hponimia, meronimia
 - **Ontología** de eventos (y entidades)
 - **Polisemia** (y extensiones metafóricas del significado)
 - **Combinatoria léxica**
- **Análisis lingüístico** para representar el contenido transmitido por un texto y el cómo se organiza ese contenido
- **Recursos lingüísticos**
léxicos y redes de patrones construccionales basados en corpus

ReLiSCo

GE

(Gramática del Español)

<http://gramatica.usc.es/>

Coordinador: Guillermo Rojo



ReLiSCo

Gramática de Español

Grupo de lingüistas del Departamento de Lingua Española de la USC

Líneas de investigación principales

- Morfología derivativa y morfología genética
- Estructura gramatical del español
- Lingüística de corpus (textuales y orales)
- Procesamiento del Lenguaje Natural y Extracción de Información

Proyectos de investigación

- Red de Morfología Genética
- Formación de corpus de la lengua oral
- Construcción de un *treebank* sintáctico del español
- Extracción automática de relaciones semánticas entre Entidades

Transferencia y creación de empresas:

- **Cilenis** (<http://www.cilenis.com/>): Spin-Off de la USC sobre tecnologías del lenguaje

ReLiSCo

Herramientas y recursos

Recursos

- Base de Datos Sintáctica (BDS: <http://www.bds.usc.es/>).
- Léxicos y terminologías bilingües extraídas automáticamente de la Wikipedia y otras fuentes textuales.
- Léxicos computacionales para el español, portugués y gallego, adaptados a etiquetadores como FreeLing, Tree-Tagger, Xiada.
- Gramáticas computacionales para varias lenguas.

Herramientas (licencia GPL)

- Analizador sintáctico de dependencias para español, portugués, inglés, francés y gallego ([DepPattern](#))
- Extractor multilingüe de multipalabras
- Extractor multilingüe de tesauros
- Prototipo de extracción de léxicos bilingües (a partir de corpus comparables)
- Prototipo de extracción de relaciones semánticas (metodología *Open Information Extraction*)

ReLiSCo

Transferencia

Cilenis™

Conjugadores verbales gratuitos con licencia GPL, para gallego, portugués y español: disponibles mediante consulta on-line y para instalar en smartphones (Android)

Evaluador automático de la lengua escrita (Avalingua): analiza diferentes niveles de lengua (ortográfico, léxico y gramatical), detecta errores, los clasifica y evalúa el texto.

Mejoras de recursos y herramientas PLN y adaptación a diferentes lenguas:

- Sistemas de NER (*Named Entity Recognition*) estadísticos e híbridos.
- Léxicos computacionales.
- Terminologías bilingües.
- Creación de corpus especializados “a la carta” mediante técnicas de *crawling*.

ReLiSCo

FILGA

(Grupo de Filoloxía e Lingüística galega)

<http://www.usc.es/es/investigacion/grupos/filga>

Coordinador: Xulio César Sousa Fernández
M.Sol López Martínez



ReLiSCo

FILGA

Grupo de lingüistas y filólogos del Departamento de Gallego de la USC

- Instituto da Lingua Galega (ILG), Centro Ramón Piñeiro (CIRP) y Facultade de Filoloxía
- **Líneas de investigación principales:**
 - Fonética
 - Fonología
 - Gramática sincrónica y gramática histórica
 - Lexicografía
 - Onomástica
 - Dialectoloxía
 - Sociolingüística
 - Historia de la lengua
 - Edición de textos

ReLiSCo

Proyectos de investigación

ILG <http://ilg.usc.es>

LENGUA ORAL

- **Arquivo do galego oral**
- **CORILGA: Corpus Oral Informatizado da Lingua Galega**
- **Amper.** Atlas multimédia prosodique de l'espace roman (en colaboración con varias universidades europeas e americanas)
- **Estudo perceptivo da variación prosódica dialectal do galego**
- **Melloras na versión sonora do Dicionario fonético galego**
- **Tesouro oral informatizado da lingua galega**

LÉXICO

- **Atlas Lingüístico Galego.** Léxico agrícola
- **Tesouro do léxico patrimonial galego e portugués** (en colaboración coas universidades de Lisboa e Coimbra (Portugal), e con sete universidades do Brasil coordinadas pola Universidade Federal do Rio de Janeiro)
- **Dicionario de dicionarios castelán-galego**
- **Dicionario de dicionarios do galego medieval**
- **Dicionario de léxico figurado galego**
- **Léxico dispoñible do galego**
- **Tesouro informatizado da lingua galega e Dicionario da lingua galega**
- **Dicionario dos apelidos galegos**

CORPUS

- **Corpus para o estudo da sintaxe dialectal do galego**
- **Xelmírez.** Corpus lingüístico da Galicia medieval

ReLiSCo

Proyectos de investigación

CIRP <http://www.cirp.es/>

CORGA. Corpus de Referencia do Galego Actual

<http://corpus.cirp.es/corga/>

Xiada. Etiquetador e lematizador do galego actual

<http://corpus.cirp.es/xiada/>

CoToVía. Sintetizador de voz para o galego

<http://www.cirp.es/prx2/sinte.html>

ReLiSCo

Objetivos globales

- Investigación sobre **lingüística e filoloxía galegas** en todas las líneas y todos los períodos históricos.
- Desarrollo de **recursos informáticos** que ayuden a la incorporación de la lengua gallega en el mundo de las **nuevas tecnologías**
- Asesoramiento técnico a instituciones y personas relacionadas con la **normalización del gallego**.

GRILES

(Grupo de investigación de lengua española y lenguas signadas)

webs.uvigo.es/griles

Coordinadora: Inmaculada C. Báez Montero



GRILES

Grupo de lingüistas del Departamento de Lengua española de UVIGO

Líneas de investigación principales

- lengua española
- lenguas signadas (LS)
- **enseñanza/aprendizaje** de las lenguas española y de signos
- estudios de la situación actual de la LSE con el fin de que estas investigaciones contribuyan a mejorar el estatus de las LS en el estado español, en la sociedad española y al **reconocimiento de la LS como lengua natural de la comunidad sorda.**

ReLiSCo

Proyectos de investigación

- **SIGNOGuía** georreferenciada Xacobeo 2010 – Código IN849A 2010/2762. Xunta de Galicia
- **O acceso á educación universitaria** galega en **lingua de sinais** (LS): análise, deseño e elaboración das probas para a avaliación das competencias – Código IN825B2010/20-0 (2202). Xunta de Galicia
- Hacia un **corpus textual digital** de lingua de signos española: accesibilidade, gramática y enseñanza LSE/L2 – FFI2010-20972
- FaiTIC para todos: o **acceso á plataforma de teledocencia** para os xordos e hipoacúsicos. Universidad de Vigo

ReLiSCo

Objetivos globales

- **Lingüísticos**

- Extender y aplicar el conocimiento sobre **la gramática** a formas de expresión de la **LSE**, no estudiadas anteriormente.
- Desarrollar **textos** (tanto actuales como de siglos pasados) **traducidos en lengua de signos** (LS) y **subtitulados** en lengua española (LE)
- Crear materiales para la **actualización del corpus textual** que venimos elaborando desde hace años y pretendemos continuar y poner en red.

- **Docentes**

- **Formar docentes e intérpretes** que faciliten el acceso de los sordos a la universidad y el desarrollo de nuevos métodos y técnicas que permitan enseñar la lengua española a los sordos y la lengua de signos a los oyentes.
- Elaborar **criterios de evaluación y pruebas en LS** para el acceso de los sordos a la educación, centrándonos en el **acceso a la universidad** y creación de materiales -tanto docentes como lingüísticos- que les permitan una educación en igualdad de condiciones.

ReLiSCo

Objetivos globales

- **Aplicados**

- Realizar una **adaptación de la plataforma de teledocencia** de la Universidad de Vigo (Faitic) para todas aquellas personas que necesitan sobrepasar las barreras de comunicación en el ámbito educativo.
- Experimentar con el formato de video compuesto Polimedia como base para la **creación de contenidos digitales de alta calidad para personas sordas** y con la distribución online y offline para móviles.
- Rentabilizar las tecnologías para la difusión de contenidos científicos en una **lengua de modalidad viso-gestual**.
- Experimentar con posibles **modelos de negocio online** relacionados con los **contenidos signados de alta calidad**.

ReLiSCo

COLE

(Compiladores y Lenguajes)

www.grupocole.org

Coordinador: Manuel Vilares Ferro



Universidade de Vigo

ReLiSCo

COLE

- Grupo multidisciplinar (13 investigadores):
 - Ingenieros en Informática ... inteligencia artificial
 - Lingüistas ... filología, traducción y filosofía
 - Matemáticos ... aspectos formales de la computación
- Líneas de investigación:
 - Procesamiento del lenguaje ... análisis morfológico, sintáctico y semántico
 - Inteligencia artificial ... aprendizaje automático
 - Gestión de información ... recuperación, minería de opiniones
- Reconocimiento como grupo de excelencia y consolidables:
 - Xunta de Galicia ... consolidable (2008-10)
 - Univ. de Vigo ... excelencia (2010-13)

Proyectos de investigación

- Una veintena de proyectos de investigación:
 - Convocatorias públicas competitivas ... **nacionales e internacionales**
 - Algunos proyectos representativos:
 - **Análisis de texto y recuperación de información en minería de opiniones (TIN2010-18552-C03)**
 - **Búsqueda de respuestas empleando meta-gramáticas (HUM2007-66607-C04)**
 - **Recuperación de textos sobre bases documentales de agencias de noticias (07SIN005206PR)**

Contratos de investigación

- Una veintena de **contratos de investigación**:
 - **Generación, extracción y estructuración de la información legislativa mediante técnicas de inteligencia artificial (PGIDIT05SIN044E)**
 - **Creación y aplicación de recursos documentales en enfermedades neuromusculares (PGIDIT04SIN065E)**
 - **Aplicación de la IA para la extracción cognitiva y cualitativa de información en mercados bursátiles (PGIDIT02SIN01E)**
 - **Sistemas de recuperación de información a través de técnicas de lenguaje natural para evaluaciones cognitivas de la información (2000-02)**



(Lengua Y Sociedad de la Información)

<http://www.grupolys.org>

Coordinadora: Margarita Alonso Ramos

LyS

- Grupo interdisciplinar formado por **informáticos y lingüistas** de la UDC
- **Líneas de Investigación:**
 - Procesamiento automático de la lengua natural
 - **Minería de textos**
 - Recuperación de la Información
 - RI en español
 - RI multilingüe
 - RI robusta
 - Lexicografía computacional
 - Aprendizaje de lengua asistida por ordenador

ReLiSCo

Proyectos de investigación

- **Proyectos activos dirigidos:**

- 2 proyectos del Plan Nacional I+D+i

Análisis de textos y recuperación de la información para la MINERÍA DE OPINIONES: análisis de enunciados y extracción de relaciones (TIN2010-18552-C03-02.)

Herramienta de ayuda a la redacción en español: fundamentos lingüísticos para el procesamiento de colocaciones (FFI2011-30219-C02-01)

- 1 convenio de investigación: Centro Ramón Piñeiro

- **Redes activas:**

- Europea: Evaluating Information Access Systems (ELIAS)
- Autonómica: ReLiSCO

ReLiSCO

Herramientas y recursos

- **Recursos lingüísticos:**
 - *Diccionario de Colocaciones del Español (DiCE)*:
www.dicesp.com
 - **Corpus de aprendices** de español con etiquetado de colocaciones
- **Herramientas**
 - Etiquetación y lematización: **MrTagoo**
 - Análisis sintáctico: **COMPAS** y análisis de dependencias

Doy paso ya a las empresas matutinas:

- *Imaxin*
- *Bitext*
- *Autoritas*

Gracias por su atención

ReLiSCo