

# **AELINCO 2015 BOOK OF ABSTRACTS**

7th Conference on Corpus  
Linguistics Valladolid (Spain)  
5-7 March 2015

### **Organizing Committee**

Pedro A. Fuertes Olivera (Chair)

Esther Álvarez de la Fuente

Raquel Fernández Fuertes

Pilar Garcés García

Belén López Arroyo

Marta Niño Amo

Isabel Pizarro Sánchez

Ana Sáez Hidalgo

Ángeles Sastre Ruano

Marisol Velasco Sacristán

### **Student Helpers**

Tamara Gómez Carrero

Noelia Recio

Silvia Sánchez Calderón

### **Scientific Committee**

Francisco Alonso Almeida (Spain)

Theo Bothma (South Africa)

Pascual Cantos Gómez (Spain)

Gloria Corpas Pastor (Spain)

Raquel Criado Sánchez (Spain)

Danie Prinsloo (South Africa)

Teresa Fanego Lema (Spain)

María Luz Gil Salom (Spain)

María de los Ángeles Gómez González (Spain)

Sylviane Granger (Belgium)

Andrew Hardie (UK)

Ulrich Heid (Germany)

Julia Lavid López (Spain)

María José López Couso (Spain)

Juana I. Marín Arrese (Spain)

Antonio Moreno Sandoval (Spain)

Isabel Moskowich-Spiegel Fandiño (Spain)

José Luis Oncins (Spain)

Javier Pérez Guerra (Spain)

Emilio Ridruejo Alonso (Spain)

Jesús Romero Trillo (Spain)

Chelo Vargas Sierra (Spain)

### **And the members of the AELINCO Board**

Aquilino Sánchez Pérez (Spain)

M<sup>a</sup> Luisa Carrió Pastor (Spain)

Miguel Fuster Márquez (Spain)

Antonio Moreno Ortiz (Spain)

### **Sponsors**

Financial support for this event has been received from

Grant no FFI2011-22885: Ministerio de Economía y Competitividad; Principal Investigator: Pedro A. Fuertes Olivera

Grant no VA067A12-1: Dirección General de Universidades e Investigación; Junta de Castilla y León; Principal Investigator: Pedro A. Fuertes Olivera

### **Institutional Sponsors**

Universidad de Valladolid:

Departamento de Filología Inglesa

Research Unit: International Centre for Lexicography

Ministerio de Economía y Competitividad



Dear AELINCO participants,

Last year in Las Palmas we invited you to Valladolid. Now we would like to welcome you warmly to the 7th International Conference on Corpus Linguistics (5-7, March 2015) and to our old and beautiful city in hopes that you will profit from the very promising conference sessions. This conference will confirm that *corpus linguistics* is at a crossroads and will enable possible future ways of advancing, either by accepting the coming of big data or by denying its relevance. Arguments on both sides will be discussed providing us with inspiring and stimulating debates on the subject.

We also encourage you to find some spare time here and there and do some sightseeing: to enjoy the city cosy places, our food (e.g. tapas) and wine, and the friendly and lively atmosphere of Valladolid. We do hope that the weather will be “merciful” to us, but we should not forget that we are approaching spring time, when weather changes are as common as the old Spanish adagio: “marzo marzuelo, un día malo y otro bueno” which reminds us that in March the weather shifts, from bad to good in days or even in hours.

The AELINCO organizing committee

Pedro A. Fuertes Olivera, Esther Álvarez de la Fuente, Raquel Fernández Fuertes, Pilar Garcés García, Belén López Arroyo, Marta Niño Amo, Isabel Pizarro Sánchez, Ana Sáez Hidalgo, Ángeles Sastre Ruano, Marisol Velasco Sacristán

**(133)**

**Wandl-Vogt, Eveline (Austrian Academy of Sciences, Austria), O'Connor, Alexander (Trinity College Dublin, Ireland), Theron, Roberto (Universidad de Salamanca, Spain) & Kieslinger, Barbara (Zentrum für Soziale Innovation, Austria): exploreAT! Re-thinking lexicography**

## PANEL: CORPUS-BASED LEXICOLOGY AND LEXICOGRAPHY

This paper aims to discuss the lexicography paradigm by introducing into a new project, established at the Austrian Academy of Sciences in Vienna (funding: 11.2014; project start: 4.2015).

The project "exploreAT! exploring austria's culture through the language glass" aims to reveal unique insights into the rich texture of the German Language, especially in Austria, by providing state of the art tools for exploring the unique collection (1911-1998) of the Bavarian Dialects in the region of the Austro-Hungarian Empire. This corpus is large and rich, estimated to contain 200,000 headwords in estimated 4 Million records. The collection includes a five-volume dictionary of about 50,000 headwords, covering a period from the beginning of German language until the present (DBÖ, WBÖ).

In order to create enduring value from this resource, the project will apply open science and citizen science techniques to improve access and to leverage the crowd's wisdom. The engagement of users with the system will be the subject for mind-brain studies, and the results and records will be enriched and interlinked using the best practices of semantic content publishing of Linked Open Data on the Web of Data.

The key tasks are to:

1. Explore Austrian culture within a Pan-European and international setting, concerning both concepts of rural life of the multicultural Habsburg Empire, as well as supplementing historical and sociological inquiry with an understanding of the role and implementation of Lexicography over time.
2. Discover challenges and chances of a transformation process from a traditional lexicography project to an open cultural knowledge base and the role of lexicographic knowledge, especially with respect to automatic and semi-automatic techniques for publishing the results as five-star linked data.
3. Invite researchers, professionals, academics and amateurs, to participate, share and grow in the framework of up to date collaborative lexicography.
4. Create an open multilingual infrastructure for all to explore the world described by the corpus as documented in languages.
5. Reflect best practice for publishing multi-lingual linked open data, connecting the lexical, temporal, geographical and historical features of the corpus with the global and European knowledge web.

The interdisciplinary, international team is collaborating to reach the visaged goals by the following steps:

1. Create a web-based collaborative, multilingual infrastructure for archiving, editing, publishing and analysing non-standard data (historical, dialectal), it's lexicographic output and it's knowledge resources for scientific as well as amateur purposes.
2. Create links between different lexicographic data sets to foster exchange and interoperability at international level.
3. Engage the larger public in exploring and contributing to lexicographic data in a playful and educational approach.
4. Challenge further research for innovative ways to explore the data (e.g. visualisations, games).
5. Connect to other initiatives across the globe, e.g. DARIAH.EU, COST ENeL, EUROPEANA, LIDER, W3C Ontology Lexica Community Group, SOCIENTIZE, European School of Social

Innovation (ESSI), opendataportal.at, Open Knowledge Foundation (OKFN), European childrens universities network (EUCU) and WIKIMEDA.AT.

In this presentation we give an overall introduction into our project aims, focussing on the main issues concerning the development of collaborative, linked-open-data e-lexicography in an open science framework.