

**REPORT ON THE 1st INTERNATIONAL CONFERENCE
ON CORPUS LINGUISTICS (CILC-09)**

MURCIA, SPAIN, MAY 7-9 2009

Gotzon Aurrekoetxea

University of the Basque Country (UPV-EHU)

gotzon.aurrekoetxea@ehu.es

The “First International Conference on Corpus Linguistics-CILC-09” was organized by AELINCO (Asociación Española de Lingüística de Corpus), which is intended to “begin a conference series devoted to Corpus Linguistics as a discipline” (from the conference website). The creation of this society and the organization of the conference have been two simultaneous acts of a group of lecturers and researchers of the Department of English Philology of the University of Murcia.

During two days and a half, together with six parallel sessions, 117 papers and posters were presented (58 in English and 59 in Spanish). The topics ranged from the presentation of different corpora to the use of different applications. As the promoters proposed, a wide range of topics was presented, but now it is not easy to give account of each of them appropriately as it would be suitable in a short report. Thus, it does not represent an exhaustive review.

Apart from four plenary lecture and the round tables, the papers presented can be classified in these main subjects:

- Presentations of diverse corpus
- Corpus related with specific purposes
- Corpus and teaching
- Corpus and linguistic description
- Corpus and discourse
- Corpus and lexicography
- Using corpora in contrastive and translated studies
- Corpus and stylistics
- Corpora and linguistic variation

Following this classification, we will comment enumerate the content of presentations.

1. Presentations of diverse types of corpus

This section reports papers that focused on types of corpus.

1.1. Corpus based on ancient English literature

The old English literature written in dialects was represented by three different corpora: the SCONE Corpus, the Coruña Corpus and the Salamanca Corpus.

The “SCONE Corpus” was showed by J. Gabriel Amores / Julia Fernández / Luisa García (*Annotating Early Modern English Texts in the SCONE Corpus*). It is about the northern English dialect literature from the 7th to the 16th centuries. The corpus covers the edition of the manuscripts and information about the language at different linguistic levels.

The “Coruña Corpus” compiles texts from 17th to 19th centuries, and it was presented by Inés Lareo (*El Coruña Corpus. Proceso de compilación y utilidades del Corpus of English Texts on Astronomy (CETA). Resultados preliminares sobre el uso de Predicados Complejos en CETA*), Moskowich (*Sobre la recoplación de corpus: la experiencia del Coruña Corpus*) and Crespo (*Tras la recopilación: el diseño de La Coruña Corpus tool*). The former shows the criteria for the compilation and selection of writers.

There were also three papers using “Salamanca Corpus”: M^a Fuencisla García-Bermejo (*The Nineteenth century Representation of Yorkshire dialect in the Salamanca Corpus*) presented a description of the 18th and 19th centuries texts of Yorkshire and described the techniques used by writers to use a regional dialect contributing to the knowledge of Yorkshire English from that period). Pilar Sánchez García (*Depicting Northern English dialects through spelling. A study based on the Salamanca Corpus*) showed one of the methods used by Northern English writers on the depiction of the dialect by spelling). Francisco Javier Ruano García (*Looking at suffixation in Early Modern northern English: A preliminary approach using The Salamanca Corpus*) focused on suffixation with two main objectives: see the word-formation patterns of the 16th and 17th centuries, and correct the chronology of certain derivatives suffixes).

1.2. Diverse corpora of Spanish

The paper presented by Isabel De la Cruz & Cristina Tejedor (*Criterios para el diseño de una base de datos de préstamos ingeleses en el español peninsular contemporáneo*) is based on the works of the aLiLex research team and describes the parameters of the design of the data-base.

José Luis de Lucca's paper (*PHRASENET. Detección y extracción de unidades fraseológicas a partir de un corpus textual*) focused on how to extract phraseological units from textual corpora using algorithms and statistical methods.

Esteban T. Montoro del Arco & Emilio J. García Wiedemann (*El Corpus del Español Hablado en Granada: Proyecto para el Estudio Sociolingüístico del Español de España y América (PRESEEA)*) showed the published works and the current situation of the works from their project.

M^a Pilar Valverde Ibáñez and Eckhard Bick (*A Web Corpus of Spanish Automatically Annotated with Semantic Roles*) spoke about an Internet corpus that has been automatically syntactically annotated with semantic role information; they provided also some statistics about the relationship between semantic and syntactic roles.

1.3. Corpora of other languages

Gotzon Aurrekoetxea, Jon Sánchez and Igor Odriozola presented (*EDAK: A Corpus to Analyse Linguistic Variation*) the linguistic features (geolinguistic and sociolinguistic variation) and the data structure (SQL and TEI structure) of the EDAK corpus –The Dialectal Oral Corpus of the Basque language- project.

Joan Torruella y Casañas (*Los ejes principales en el diseño de un corpus diacrónico: el caso del CICA*) presented the computerised corpus of old Catalan which has diachronic, typological and dialectal aspects.

1.4. Other corpora

Antonio Miranda-García & Javier Calle-Martín (*An Annotated Corpus of The Federalist Papers*) presented a twofold paper: the description of the salient features of

the corpus and the presentation of its potential by using *Text SearchEngine* software, designed to retrieve morphosyntactic and lexical information.

Eva Revilla, Marta Villegas, Carla Parra, Núria Bel (*El Proyecto CLARIN: Una infraestructura de investigación científica para las Humanidades y las Ciencias Sociales*) presented an European project which aim is to boost the use of technologies in the research in humanities and social sciences, by creating a GRID net structure.

Lydia-Mai Ho-Dac, C. Fabre, M.-P. Péry-Woodley and J. Rebeyrolle (*Corpus annotation for macro discourse structures*) presented ANNODIS, a discourse annotation project that combines bottom-up and top-down points of view on discourse.

2. Corpus related with specific purposes

One of the most prolific subjects was corpora for medical purposes. M^a Isabel Navarro (*Estudio Comparativo de las habilidades metalingüísticas de un niño con Trastorno Específico del Lenguaje*) showed two corpora, one about children (from 1,8 to 6,1 years old) language with normal language development and another with a specific disorder; the outputs could be applied in clinic field. Teresa Fernández de Vega, Elena Cabeza, Gema García, and Sara Llena (*Planteamiento de Corpus Lingüístico de una población con discapacidad intelectual*) worked with intellectually disabled scholar children within applied linguistics.

Nevertheless, Herminia Peraíta and Lina Grasso's work (*Corpus lingüístico de definiciones de categorías semánticas de sujetos ancianos sanos y con la enfermedad de Alzheimer. Una investigación transcultural hispano-argentina*) was about older people (60-90 years old) with memory pathologies and its outputs could be used in psychology, linguistics and neuropsychology among other sciences.

Tamara Varela (*Córpore ad hoc en la práctica traductora especializada: aplicación al ámbito de las enfermedades neuromusculares*) presented a methodology to create corpora *ad hoc*, directed to specialised translation work, with examples of medicine.

Alba Fernández (*La expresión de la modalidad epistémica en el español científico-médico y en el español conversacional. Análisis contrastive*) proved that resources of expression change according to particular needs.

Camino Rea Rizzo (*Bandwidth: A corpus-based detailed analysis*), based on specialized vocabulary in telecommunication engineering of English, checked the effectiveness of the automatic classification and detailed the established lexical associations.

3. Corpus and teaching

This framework was well-represented in the conference. There were four papers about Spanish Foreign Language subject: Aymé Pino (*Palabras en interacción: un corpus de aprendices suecos de E/LE*) presented a study based on SFL, where she talked about the frequency of some connectors used by Swedish students and compared them with the examples of CREA and GP4; Irene López and Elena M^a García (*Teaching idiomatic expressions to learners of EFL through a corpus based on Disney movies*); Hui-Chuan Lu and Yi-Chen Wang (*Un estudio de colocaciones a partir de corpus: El análisis de la interrelación entre L1 y L2 en la adquisición del español*); and Sabela Prieto, Estela Mosqueira and Nancy Vázquez (*Córpora y enseñanza de lenguas: se buscan colocaciones*) used CEDEL2 corpus to compare the use of collocations between native Spanish and SFL students, with the aim to elaborate didactic proposals.

Willelmira Castillejos (*Error Analysis in a Learner Corpus. What are the Learners' Strategies?*) presented a ICLE corpus-based paper and analysed the intralingua errors made by native Spanish with a high proficiency level of the English.

Gema Alcaraz (*Vocabulary input in EFL textbooks: Frequency levels*) compared the vocabulary input of a Primary Education coursebook with General English from the British National Corpora.

Raquel Criado and Aquilino Sánchez (*Lexical frequency in textbooks: a necessary condition for consolidating procedural knowledge in the acquisition of language*) analysed EFL textbooks taking into account the cognitive processes of knowledge acquisition, the role of frequency in language, and the distribution of new words in textbooks.

4. Corpus and linguistic description

Miguel Gonçalves (*Corpora y descripción lingüística: de nuevo el caso de por tanto*) based his paper on CRPC, NURC-RJ, and REDIP Portuguese-corpora and found that the main use of this linker is not a conclusive value.

Cristina Suárez (*Development of dependent structures: an instance from the History of English*) illustrated the development of the dependent structure of relative clauses in English.

Lieselotte Brems and Kristin Davidse (*The reanalysis and grammaticalization of nominal constructions with kind/sort of: Chronology and paths of change*), refining Denison's chronology, argued that new constructions emerge as the result of complex interlocking paths.

Marcos García Salido (*Enfoques cuantitativo y cualitativo en el estudio de la alternancia pronombre personal/afijo en español*), based on conversational and survey data, analysed the influence of the former sentence in the alternation of pronoun.

5. Corpus and discourse

Lydia-Mai Ho-Dac (*An exploratory data-driven analysis for describing discourse organisation*) focused on the initial position, emphasising the important function that this position has in discourse organization.

Juana Isabel Marín (*Effective vs. Epistemic Stance and Inter/Subjectivity in Discourses. A Corpus Study*) presented results of the various variables in political and journalistic discourse.

M^a Ángeles Orts and Angela Almela Lafuente (*An approach to the Spanish economic discourse of the crisis: translating the untranslatable, untranslating the translatable*) began analysing an *ad hoc* corpus of economic news-items, and then they tried to clarify the terminological problems in translation.

M^a Ángeles Orts and Ana Rojo (*Metaphor framing in Spanish economic discourse: a corpus-based approach to metaphor analysis in the Global Systemic Crisis*) showed the study based on a corpus of financial articles of the newspaper *El Economista* during 2007 and 2008.

Irina Keshabyan and Pascual Cantos (*“Vengeance” and “Honour” in Shakespeare's Hamlet and Sumarokov's Gamlet: A Collocational Approach*) analysed the perception of these two words using quantitative and qualitative analysis.

Iker Zulaica (*On the Cognitive Status of Antecedents in Spanish Discourse Anaphora*) argued in favour of the status of the anaphora.

Sara Piccioni and M. Giovanna Biscu (*No te puedes perder... Estrategias de implicación del lector en un corpus de folletos turísticos españoles*) presented the design of the corpus and studied the implicational strategies of the texts.

Francisco José Sánchez García (*Usos metafóricos del lenguaje político español. La metáfora estructural en los debates sobre el estado de la nación*) studied the use of metaphors in the speech of politicians.

Inés Olza Moreno (*Estudio cognitivo de un corpus de metáforas sobre el 'proceso de paz'*) analysed the influence of some metaphors over other features.

An Vande Casteele (*El papel comunicativo de las expresiones nominales con referente humano en los titulares de artículos periodísticos*), based on a newspaper, analysed the headlines that have a human reference.

Tony Bastow (*Some discourse features of the modal adverbials of course, clearly, obviously, and certainly in a corpus of defence speeches*) studied the different features that these adverbs have in discourses.

6. Corpus and lexicography

José Luis Oncins (*Corpus Linguistics as an aid to lexicographers: Towards a better informed description of Anglicisms in Contemporary Spanish*), based on CORDE and CREA corpora, suggested a more accurate way to study different aspects such as the degree of penetration or the speed of spreading of loans.

Isabel De la Cruz, Guzmán Mancho and Cristina Tejedor (*Análisis de un corpus de textos turísticos: La incorporación, difusión e integración de los préstamos ingleses en los textos turísticos*) used a subcorpus of the aLiLex research team concerning tourism lexicon and studied the presence of Anglicisms and its percentages.

Two papers were presented by the research project FRASESPAL: Ana Mansilla (*Las entradas léxicas en un corpus onomasiológico de español-alemán*) showed the

creation of a phraseologic bilingual corpus and the pitfall to identify the lexical entry. Carmen Mellado (*Utilidad y limitaciones de los corpora informáticos en la elaboración de un tesaurus fraseológico (alemán-español)*) dealt with the difficulties of the interpretation of data-bases, to make statements, definitions and examples.

Andjelka Pejovic (*De la idea a la locución; hacia la propuesta de un nuevo diccionario fraseológico basado en los ejemplos procedentes de corpus textuales*) analysed the possibility to elaborate a dictionary of idioms, using not their formal structure, but their meaning.

Aquilino Sánchez, Pascual Cantos and Moisés Almela (*In search of a new model for WSD: a multilayered lexical constellation model*), based on previous researches, and regarding WSD algorithm, concluded that the model of meaning organization used by most of them is not adequate.

Pedro Fuertes (*Working with English Specialized Corpora: lexical bundles in Written Scientific and Technical English*) focused on some three-word lexical bundles and confirmed the Hyland's hypothesis of the discourse of the academy.

Lourdes Cerezo and M^a Dueñas (*Calidad literaria frente a fenómeno editorial: análisis cuantitativo del contenido léxico de dos novelas contemporáneas*), based on two current novels with different literary canons and using *WordSmith Tools* program, analysed the lexical density, the keywords, and the lexical and thematic frequency of both works.

Gloria Vázquez and Ana Fernández Montraveta (*Ampliación del Banco de Datos de Verbos del español SenSem*) showed the innovative aspects in the field of semantics. The work focused on aspectuality and modality.

7. Using corpora in contrastive and translation studies

Among all the papers presented in this section, we have selected a sample: Louise Frances Denyer (*Corpus study carried out on three "legal" verbs to demonstrate their similar and different usage for the purposes of legal translators and lawyer-linguists*) showed her experience taking into account some examples and comparing different corpora as BOE, CREA, CE, CCF, and other dictionaries.

Bárbara Lübke, Pia Sloth Poulsen and Victoria Vázquez Rozas (*La explotación de un corpus analizado sintácticamente en el Diccionario contrastivo de valencias verbales español-alemán*) described the structure and the development of the dictionary explaining the advantages and limitations of the corpora as basic material.

8. Corpora and linguistic variation

Juan Manuel Hernández Campoy, Juan Antonio Cutillas-Espinosa and Natalie Schilling-Estes (*Proactive Practice in Style-shifting: Evidence from Radio Recordings*) presented a study about an unexpected use of local dialect features by a very important local politician.

Esteve Clua, Esteve Valls, Maria Rosa Lloret (*Análisis lingüístico y dialectométrico del Corpus Oral Dialectal (COD)*) showed the results of the linguistic and Cluster Analysis using UPGMA method.

The next International Conference on Corpus Linguistics (CILC-10) organized by the AELINCO society will be held at the Universidade da Coruña in May 2010.