

# CILLC 2016

8th International Conference on Corpus Linguistics

Universidad de Málaga  
2-4 de marzo de 2016

BOOK OF ABSTRACTS

---

---

---

## INDEX OF PAPERS

**Paper #1.** Wednesday, March 2. 13:30-14:00. LIMONAR

**Guangwei Hu and Feng Cao.** "A Corpus-Based Comparative Study of Interactional Metadiscourse in Research Articles"

**Abstract:**

Metadiscourse is a construct grounded in the fundamental understanding of discourse as social engagement. It refers to rhetorical and pragmatic expressions whereby "we project ourselves into our discourses by signalling our attitude towards both the content and the audience of the text" (Hyland, 2005). Along with an increasing recognition of their important functions in connecting writer, reader, and text, such self-reflexive expressions have attracted much attention in research on academic discourse. Previous studies have shown that the use of metadiscourse in academic communication can be influenced by such factors as discipline, language/culture, and genre. Building on this line of research, this corpus-based comparative study draws on Hyland's interpersonal model of academic metadiscourse to examine how discipline and research paradigm influence the use of interactional metadiscourse in 120 research articles drawn from three social science disciplines (i.e., applied linguistics, education, and psychology) and two research paradigms (i.e., quantitative and qualitative research).

**Paper #7.** Wednesday, March 2. 18:00-18:30. LIMONAR

**Anja Hennemann.** "A corpus analysis of sentence-initial *Lo que*. A constructionist approach"

**Abstract:**

The present paper adopts a constructionist approach in terms of frequency and entrenchment to the rise and development of sentence-initial *Lo que* 'What' – a so called "copulativa enfática de relativo" (NGLE 2009: 3024). The focus marker *Lo que* is considered a construction because its form, function and even meaning is not strictly predictable from its component parts (Goldberg 2006: 5). Focus marking constructions are said "to put forward a particular linguistic entity, which contrasts with other possible entities" (Van den Steen 2005: 279) or "to mark focus on the clefted constituent" (Dufter 2008: 31). The work with the CREA and CORDE shows that the construction's frequency has increased over time so that the construction can nowadays be said to occur "with sufficient frequency" (Goldberg 2006: 5). A further goal of the study is to describe the construction in more detail, showing its characteristics, its use and its restrictions of use.

**Paper #13.** Wednesday, March 2. 18:30-19:00. LIMONAR

**Oscar García-Marchena.** "The Structure of Spanish Verbless Sentences"

**Abstract:**

The variety of structures in Spanish verbless sentences has not yet been fully described and can only be accounted by corpus data. CORLEC corpus of contemporary oral Spanish furnishes a considerable amount of examples that makes possible the analysis of these syntactic structures. An interesting case is the identification of the syntactic status of the SN selected by the non verbal head : a description of its syntactic properties is required in order to determine whether it constitutes its subject or it is just a peripheric argument. This paper examines the properties which allow us to identify the function of the SN, and completes the description of the syntactic structures found in the corpus by analysing the different word orders as a result of different informational structures. This analysis is applied to the five types of predicative verbless sentences found in the CORLEC corpus.

**Paper #14.** Wednesday, March 2. 11:30-12:00. LIMONAR

**Isabel Moskowich and Begoña Crespo.** "Categories and genres in CHET and CECHeT"

**Abstract:**

The classification of text samples into larger categories has been a perennial problem for corpus compilers, and there is frequent debate about what features to take into consideration in assigning a text to a category. The apparently simple task of collecting text samples always requires some kind of categorization. Previous research has pointed out that there is a relationship between the kind of communicative format selected by the author, the language used, her/his cultural background or professional activity and the intended audience (Duszak, 1997; Littlejohn and Fox, 2008). One of the aims of this paper is to present both the terminological mess for text categories and the hypothetical independence of form and function.

**Paper #15.** Friday, March 4. 10:30-11:00. LIMONAR

**Anna Matamala and Marta Villegas.** "Designing a new platform for AD analysis: the VIW project"

**Abstract:**

Audio description (AD) is an intersemiotic translation in which images are translated into words. Research on AD is recent, and very few projects have approached the creation of a corpus of AD (TIWO, TRACCE). However, two frequent problems are found: processing multimodal content and dealing with copyright. This is probably why there is not a freely available corpus of AD. Inspired by Chafe's Pear Stories project, VIW aims to develop a platform which will build an ad hoc corpus of ADs. Departing from a shared stimulus, a wide corpus of ADs in multiple languages will be developed, allowing for contrastive studies. In our paper we will be presenting the project rationale, the platform architecture and its development. Key aspects that will be discussed are the creation of corpora from authentic versus created materials, the need for open access copyright-free platforms, as well as the challenges of processing multimodal multilingual data.

**Paper #16.** Thursday, March 3. 15:30-16:30. CALETA

**Naoki Nakamata.** "Productivity and Learning Difficulty of Verbs: A Case for Verbs Related to "Eating" in Japanese"

**Abstract:**

This study suggests that the productivity of verbs in Japanese has a relation with the difficulty of acquiring them as second or foreign language. Productivity is the degree of variety of other items that a target item can combine with. It can be calculated by  $TYPE \div \sqrt{TOKEN}$ . First, 125 Japanese verbs related to "eating" are chosen from vocabulary list for learners. Next, productivity is calculated by using an extensive corpus (nouns marked accusative for transitive verbs and nouns marked nominative for intransitive verbs). Finally, using learners' corpora, the study surveyed the frequency with which the learners used each verb and the level at which they began to use it. The result shows the verbs with high productivity appear at a very early stage, those with very low productivity also appear early, and some verbs with middle productivity appear at a very late stage.

**Paper #17.** Thursday, March 3. 16:30-17:00. RETIRO

**Henry Hernández Bayter.** "Estrategias discursivas en el discurso de investidura en América Latina: aportes de la lexicometría y de la lingüística de corpus"

**Abstract:**

El discurso de investidura corresponde a todo discurso pronunciado durante la toma de posesión de una función, en nuestro caso nos interesamos en el discurso de investidura presidencial en América Latina, en particular en Venezuela, Bolivia y Ecuador. Este tipo de discurso, como todo discurso político, posibilita la implementación de ciertas estrategias discursivas (Charaudeau, 2005) por parte del locutor político por medio de la palabra (Fernández Lagunilla, 1999; Charaudeau, 2013) y permite la creación de una progresión argumentativa (Gréciano, 1998) que busca la adhesión del público (Plantin, 2012; Amossy, 2000; Perelmann, 2000). Nuestro objetivo es comparar los discursos de investidura pronunciados por: Hugo Chávez, Evo Morales y Rafael Correa. De esta manera, respondemos a la necesidad de contrastar de los estudios lexicométricos, empleados en nuestra investigación.

**Paper #18.** Wednesday, March 2. 12:00-12:30. RETIRO

**Miguel Fuster-Márquez and Carmen Gregori-Signes.** "Looking into the flexibility of lexical bundles in the discourse of US hotel websites"

**Abstract:**

This contribution focuses on the analysis of sequences known as lexical bundles, in US hotel websites. Research on bundles has kept growing in the last two decades and phraseological inventories have been proposed. The extraction of bundles is frequency-driven, after which a more qualitative analysis is typically carried out. The corpus used for this study is part of COMETVAL, a multilingual database of tourism websites, compiled at the University of Valencia. We used a US subcorpus of hotel sites which contains 460,000 words. This discourse has few technical terms and its goal is to persuade customers to acquire services and products. The presence of bundles shows that writers resort to them to achieve their goals. The analysis also shows that bundles are flexible, therefore revealing the existence of a greater phraseological wealth through the addition of less canonical forms.

**Paper #19.** Thursday, March 3. 12:00-12:30. ADUANA

**Javier Calle and Jesús Romero-Barranco.** "Third Person Present Tense Markers in some Varieties of English"

**Abstract:**

In British Standard English, number in the verb phrase is exclusively characterized by the use of the -s inflection with the third person singular in the present tense. World Englishes (Irish English, Scots, New Zealand English and Asian Englishes, among others) present a high level of variation insofar as the uninflected third person singular (i.e. she want to find a job) and the inflected third person plural (i.e. they never goes there) also occur in these linguistic contexts. The

phenomenon, however, has not been hitherto undertaken quantitatively in the emerging varieties of English (Mesthrie and Bhatt 2008: 66-67) and therefore the present paper pursues the following objectives: a) to analyse the use of the present third person inflections (singular and plural) and compare its distribution both in the different varieties of English and across speech and writing and across text types; b) to classify the instances in terms of the typology of the subject; and c) to evaluate the impact of intervening elements in the choice of inflection. The data used as source of evidence come from The International Corpus of English (i.e. New Zealand, India, Singapore and Hong Kong).

**Paper #21.** Friday, March 4. 12:30-13:00. LIMONAR

**Stephane Patin.** “La traducción de las metáforas corporales en los debates parlamentarios del corpus paralelo español / francés del *europarl*”

**Abstract:**

This study examines the strategies used by UE institutional translators when translating metaphors whose source domain involves the human body (the heart of democracy, the political arm of ETA...). In order to elicit such examples, we explore the Spanish and French sections of the EUROPARL corpus (ES 54 806 927 tokens, FR 54 202 850 tokens) using search requests formulated in Corpus Query Language (CQL). In the case of *corazón* and its French equivalents, the main translation strategies consist in a) translation by the same metaphor (el corazón de la Unión Europea < > le coeur de l'Union européenne), b) use of another metaphor (el corazón de la propuesta < > le noyau de la proposition), c) expansion or explicitation (en el corazón de Europa < > au centre de l'Europe), among other strategies.

**Paper #22.** Thursday, March 3. 11:30-12:00. RETIRO

**Julia Lavid, Marta Carretero and Juan Rafael Zamorano.** “Designing and validating an annotation model of modality for English and Spanish: issues and problems”

**Abstract:**

In this paper we report on current work on the design and empirical validation of an annotation model for the category of modality in English and Spanish in the context of the MULTINOT project. While there exist proposals for the annotation of modality in other languages, to date there are no practical annotation schemes for the contrastive annotation of English and Spanish. In this paper, therefore, we intend to fill a gap in this area, by presenting an annotation model for modality in English and Spanish and describing current work in its empirical validation through two annotation experiments. Our annotation model works with two dimensions: the functional-semantic dimension (capturing the modal meanings) and the linguistic dimension (capturing the linguistic encodings of those modal meanings in both languages). We also report on two annotation experiments using the annotation schemes proposed. We are planning to use these validated schemes for the large-scale annotation of the MULTINOT corpus, which will, hopefully, be a useful resource for a range of Natural Language Processing applications, such as Machine Translation and Computer-Assisted Translation systems.

**Paper #24.** Wednesday, March 2. 12:30-13:00. LIMONAR

**Kenichi Kadooka.** “A Study of Japanese Modality Using a Corpus”

**Abstract:**

This paper is to analyze the Japanese explanatory modality expressions using corpus ‘Shonagon’ edited by the National Institute of the Japanese Language. The explanatory modality in Japanese is to add explanations to a proposition. When a simple proposition ‘Kore ga sono hon da (This is the book)’ is followed by the modality expression ‘no da (noun and auxiliary)’, emerges the explanatory style ‘Kore ga sono hon na no da (This is the very book).’ Notice that before the modality expression ‘no da,’ is added an auxiliary verb ‘na,’ which is derived from ‘da’ itself. As an intuition, a simple proposition such as ‘Kore ga sono hon da’ is rare as an actual use because it sounds strong. Hence the modality expression ‘Kore ga sono hon na no da’ is a mitigation of such strength. There are four modality expressions investigated here: ‘no da, wake da, koto da, mono da.’ It is interesting that ‘wake, koto, mono’ are nominals, meaning ‘reason, thing, material’ respectively. Of these three nominals, ‘mono’ is the most concrete because it can signify objects like stone, furniture, iron, food, plant, and so on.

**Paper #25.** Wednesday, March 2. 11:30-12:00. ADUANA

**María Ángeles Zarco-Tejada, Carmen Noya Gallardo, María del Carmen Merino Ferradá and María Isabel Calderón López.** “Identifying corpus-based criterial cohesive devices for proficiency automatic classification”

**Abstract:**

The aim of this study is to identify which and how cohesive devices discriminate among levels of proficiency in order to establish quantitative and qualitative differences among levels for 2L English learning texts for proficiency verification. The main cohesive devices studied are Halliday’s (2012) lexicogrammatical resources, with a special attention to

conjunctions, ellipsis and lexical cohesion. The written A2, B1, B2 and C1 subsets of the CLEC (CEFR-labelled English Corpus), a corpus of 2L English materials built up to train statistical models for automatic proficiency assessment, have been analysed with Coh-Metrix and AntConc computational tools. Establishing differences among levels automatically in terms of textual cohesion using CEFR-levelled learning texts is a first step towards the identification of cohesive devices proficiency-oriented and a procedure to verify if and how cohesion is achieved by CEFR-levelled teaching materials.

**Paper #27.** Wednesday, March 2. 19:00-19:30. ADUANA

**Keith Stuart, Ana Botella and Imma Ferri-Miralles.** "A corpus-driven approach to Sentiment Analysis of patient narratives"

**Abstract:**

This paper describes the linguistic analysis of a corpus of patient narratives about their health care that was used to develop and test software to carry out sentiment analysis on the corpus. There is a growing body of research on the relationship between sentiment analysis, social media and health care (Verhoef et al., 2014), but less research on sentiment analysis of patient narratives (Xia et al., 2009). The motivation for this research is that patient narratives of experiences of the National Health Service provide rich data of the treatment received. The corpus threw up some unexpected results that may be of benefit for researchers of sentiment analysis. The linguistic problems encountered have been divided into three sections: the noisy nature of large corpora; the idiomatic nature of language; the nature of language in the clinical domain. This article describes these linguistic problems that arose out of a project which tried to find a means to automate the analysis of patient feedback on health services. Information gathered through sentiment analysis on a corpus of patient narratives could ultimately make a difference to the running of hospitals and the wellbeing of patients.

**Paper #28.** Thursday, March 3. 11:30-12:00. ADUANA

**Solveig Granath and Magnus Ullén.** "The elevation of sensitivity over truth: Political correctness in Time magazine 1923-2006"

**Abstract:**

This paper is a quantitative and qualitative diachronic study of how the expression politically correct (PC) and related terms are used in the American magazine Time from 1923 through 2006. The data show a dramatic increase in the frequency with which PC terms are used in the early 1990s. From this time onwards, the terms are often used as a means of passing evaluative subjective opinions off as objectively reported facts, especially in reviews of cultural events, where they figure prominently. In contrast to earlier studies, our data show that PC terms are not inherently negative; this applies primarily to discourse on environment, where to be politically correct often implies being environmentally or socially conscious in a positive sense. Nevertheless, negative or ironic uses of the terms predominate. Most often they express criticism of unspoken cultural norms rather than being attempts to close down debate or criticizing the replacement of offensive terms by more neutral expressions.

**Paper #30.** Thursday, March 3. 11:00-11:30. ADUANA

**Juan Manuel Hernández-Campoy, Juan Camilo Conde-Silvestre, Tamara García Vidal and Belén Zapata-Barrero.** "Comunidades de Práctica y Estilos en Documentos Epistolares de los Siglos XV-XVIII: Proyecto HiStylVar"

**Abstract:**

La reciente y oportuna complementariedad interdisciplinar entre la Sociolingüística Histórica, la Lingüística de Corpus y la Historia Social está posibilitando que los investigadores se sumerjan en periodos remotos de la lengua inglesa y exploren con mayor precisión su funcionamiento interno y el comportamiento sociolingüístico de los usuarios en su interacción social, además de conferirle validez empírica y fiabilidad histórica. La conservación de colecciones de correspondencia privada con autores de diferente caracterización socio-demográfica y geográfica, así como circunstancias personales, supone una fuente muy valiosa para llevar a cabo investigación sociolingüística tanto cuantitativa como cualitativa. En este sentido, el interés histórico y filológico de los documentos epistolares ingleses pertenecientes a los siglos XV-XVIII resulta excepcional, no sólo porque ofrecen datos sobre la vida política y doméstica de la Inglaterra de esa época, sino también porque se redactaron en periodos decisivos para el desarrollo de la lengua inglesa. El propósito de esta comunicación es la de presentar el Proyecto HiStylVar, diseñado para estudiar estos modelos sociolingüísticos de variación estilística y posibles comunidades de práctica en corpus históricos de correspondencia en lengua inglesa de dichos siglos XV-XVIII, ofreciendo ejemplos de tendencias y patrones de sus resultados preliminares.

**Paper #32.** Thursday, March 3. 12:00-12:30. RETIRO

**Paula Rodríguez-Puente and Julia Williams-Camus.** "The University of Cantabria English Learner Corpus (UCELC): design and compilation techniques"

**Abstract:**

The purpose of this talk is to present the University of Cantabria English Learner Corpus (UCELC), which aims to contribute to the range of the existing corpora in Spain with language produced by native speakers of Spanish who are learning English at university level. More specifically, the data are being collected from first-year university students who have English for Specific Purposes as a compulsory subject in the degrees of Law and Medicine and general English in Primary Education. It is hoped that the UCELC corpus will contribute to existing learner corpora by providing new data about the language produced by Spanish learners of English. A systematic analysis will allow original teaching materials to be tailored to the specific needs of our students.

**Paper #33.** Wednesday, March 2. 16:00-16:30. RETIRO

**Paula Rodríguez-Puente, Cristina Blanco-García and Iván Tamaredo.** "The Corpus of Historical English Law Reports: A progress report"

**Abstract:**

In this paper we present the structure and characteristics of the Corpus of Historical English Law Reports, a diachronic corpus of British English legal texts, as well as the methodology used for its compilation. The new corpus will contain approximately half a million words and cover the years from about 1600 to 2000. The texts included are British English law reports: records of judicial decisions that are "cited by lawyers and judges for their use as precedent in subsequent cases" (EBO s.v. law report). Although other corpora of legal English are currently available, to the best of our knowledge a computerized corpus of historical law reports has not yet been compiled. The Corpus of Historical English Law Reports will, therefore, constitute a new, useful resource for linguists with an interest in legal language, from both a synchronic and a diachronic perspective.

**Paper #35.** Wednesday, March 2. 17:00-17:30. RETIRO

**Francisca Suau Jiménez.** "Customer online reviews from TripAdvisor and their effect on the adjustment of hotel websites: the need for a tailor-made corpus"

**Abstract:**

This presentation discusses the importance of a tailor-made corpus as an undeniable baseline in a research that analyses online review topics of four hotels from TripAdvisor as well as their possible discursive connection with those same topics turned into positive values that enhance hotel services in their websites. To this end, the tailor-made pilot corpus was compiled on the grounds of a series of CL principles used for corpus design, such as: representativeness, being able to produce valid results, the importance of quality and quantity, as well as of several external and internal criteria (Biber et al. 1998, Gelbukh et al. 2002, Bowker & Pearson 2002). 80 negative and positive reviews related to four high and medium standard hotels in New York City and London were collected, with a two-year gap left between the time customer online reviews were published (2010-2013) and the time when hotels presupposedly could have processed them and adjusted their websites (2015).

**Paper #36.** Thursday, March 3. 15:30-16:00. ADUANA

**Miriam Criado Peña.** "They are going tomorrow, isn't it?" On the use of tag questions in Indian English and Hong Kong English"

**Abstract:**

Tag questions in standard British English (PDE) follow a standard pattern consisting of an operator and a subject. This operator generally coincides with the preceding statement while the dummy auxiliary *do* is the choice when the operator is absent. More importantly, a negative tag is generally attached to a positive statement and vice versa (i.e. you know her, don't you?) (Quirk et al. 1985: 810). The Asian varieties of English are an exception insofar as apparently no standard rule is observed. The present paper investigates the use and distribution of regular and irregular tag questions in Indian English and Hong Kong English with the following objectives: a) to analyse the distribution of the construction across speech and writing, text types and other sociolinguistic factors; and b) to determine, if any, the cognitive factors influencing the choice of the non-standard patterns. For the purpose, the Indian and Hong Kong components of the International Corpus of English (ICE-Ind and ICE-HK) will be used as sources of analysis.

**Paper #37.** Thursday, March 3. 15:30-16:00. LIMONAR

**Noelia Ramón and Belén Labrador.** "Describing Cheese in English and Spanish: A Corpus-Based Contrastive Study"

**Abstract:**

This is a corpus-based contrastive study of online cheese descriptions in English and Spanish. For global trade, English is considered the lingua franca, and Spanish dairy industries need to produce accurate descriptions of the cheeses they

manufacture for commercial purposes. This paper provides native speakers of Spanish with the main rhetorical, phraseological and lexical features of this text type in English, highlighting the major differences with the same text type in Spanish. The empirical data have been extracted from a comparable corpus containing 23,089 words in English and 25,111 words in Spanish. The corpus was tagged with rhetorical labels to enable a subsequent analysis of the most common phraseological and lexical resources employed for the description of cheese in both languages. The study offers non-native writers of this subgenre in English guidelines about the structure of these texts, as well as an inventory of ready-made phraseological and lexical units.

**Paper #38.** Wednesday, March 2. 18:30-19:00. ADUANA

**Toyese Najeem Dahunsi.** "Development of a 'Nigerian English' Corpus-Based Word Prediction Software for Computers and Mobile Phones"

**Abstract:**

One of the global characteristics of most mobile phones is the in-built capacity for word prediction when composing messages. The choice of one word automatically triggers the display of some option words, which have close contextual and collocational associations with the first word. In this research, a word-prediction software was developed with one million-word corpus of the formal use of English in Nigeria as resource. The prediction model is based on concordance and word frequency principles, which determine the probability ratio of each of the words co-occurring in contexts with the target word. The choice of one word will therefore trigger the display of three other words with which the target word has the highest frequency of occurrence in the corpus. The software was developed using JAVA and PHP. It runs successfully on mobile phones, laptops and desktop computers.

**Paper #39.** Thursday, March 3. 11:30-12:00. CALETA

**Marcus Callies.** "Towards a process-oriented approach to comparing EFL and ESL varieties: A corpus-study of lexical innovations"

**Abstract:**

In the context of current research in the field of corpus linguistics that challenges the traditional division between foreign language / learner varieties of English (EFL) and institutionalized second-language varieties of English (ESL), this talk adopts a process-oriented approach to comparing the two types of varieties and examines to what extent they are driven by general cognitive processes of language acquisition and production. A comparative corpus-study of lexical innovations in derivational morphology based on data from the International Corpus of Learner English and similar ESL data from the International Corpus of English brings to light two general types of innovations: 1) interlingual, L1-based innovations, resulting from cross-linguistic influence, and 2) intralingual, L2-based innovations, resulting from various other processes. While the first type shows clear L1-effects and is virtually absent in ESL varieties, it is in the second type where similar types of innovations in EFL and ESL varieties can be observed. It will be argued that these can be explained in terms of their underlying cognitive processes that serve to create and maximise morphological transparency and increase explicitness of form-meaning relations.

**Paper #40.** Thursday, March 3. 11:00-11:30. RETIRO

**Patrick Goethals, E. Lefever and L. Macken.** "Compiling, processing and analysing Spanish corpora: results and challenges"

**Abstract:**

In this paper we will focus on the methodological problems and challenges that arise in the context of preprocessing corpora of Spanish-written texts. In a recent study on the use of adjectives in tourism texts, new corpora were compiled, Part-of-Speech-tagged and lemmatized with TreeTagger and analysed with Wordsmith Tools. Although these tools are supposed to be part of the conventional toolkit of corpus linguists, we were surprised by the high number of problems that we were confronted with. We will present a structured list of these problems, as well as a systematic account of the strategic decisions that were taken in the correction process of the Spanish TreeTagger output. However, we will also be able to show the first results of a newly trained TreeTagger parameter file, which improves the results of the currently available version. The ultimate goal is to facilitate the compilation of comparably annotated corpora, as a stepping stone to maximally comparable empirical corpus analysis.

**Paper #41.** Wednesday, March 2. 11:30-12:00. RETIRO

**Pedro A. Fuertes Olivera, Pablo Gordo Gómez, Marta Niño Amo, Ángel de Los Ríos Rodicio, Ángeles Sastre Ruano and Marisol Velasco Sacristán.** "Using the Web as a Corpus for Lexicographic Purposes"

**Abstract:**

The coming of age of the internet is reshaping entire industries and even changing the very notion of organisations, firms, and so on. Within this framework, which is referred to as a Cambrian explosion, e.g. a Cambrian explosion in lexicography (Fuertes-Olivera, 2016, 2017), this talk defends that the traditional concept of corpus used in Corpus Linguistics is not adequate for making most types of dictionaries, especially general dictionaries and specialised dictionaries (Fuertes-Olivera & Tarp, 214). Furthermore, the concept of corpus as a balanced collection of texts used for linguistic analysis, usually stored in an electronic database so that the data can be accessed easily by means of a computer, limits or hinders the compiling of some key lexicographic items, especially the crafting of definitions, the description of grammar and the use of culturally-motivated notes. Instead of a balanced linguistic corpus, this talk defends the use of the web as a lexicographic corpus, which is defined as a repository of data that is easily accessible through different internet technologies, e.g. Google, with the aim of making dictionaries aiming at satisfying specific users needs in specific use situations. We will illustrate our views with examples taken from two ongoing lexicographic projects, the Diccionario inglés-español de contabilidad and the Diccionarios en línea de español “Universidad de Valladolid”.

**Paper #42.** Thursday, March 3. 12:00-12:30. CALETA

**Shintaro Torigoe.** “Seeking the Portuguese Vocabulary Profile”

**Abstract:**

In this study, the researcher will propose a Portuguese vocabulary reference list for learners by analysing CEFR-based learner corpora, as the Portuguese version of the English Vocabulary Profile (Capel 2010, 2012) or the CEFR-J Wordlist (Tono 2013). The databases used in this study are the Corpora do PLE constructed by the University of Lisbon and the Corpus de PEAPL2 by the University of Coimbra. The researcher classifies them into subcorpora within the CEFR A1-A2, B1-B2, and C1-C2 levels, and identifies words significantly used by learners from each level. The most frequent 500 words are defined as elementary words; intermediate words and advanced words are identified from the next 1,000 words by chi-square testing between the B1-B2 and C1-C2 levels as determined by learners’ usage. The profile is aimed particularly at improving the quality of Portuguese textbooks in Japan, most of which seem to be written based merely on authors’ intuition.

**Paper #43.** Friday, March 4. 13:00-13:30. CALETA

**Toyese Najeem Dahunsi.** “An Investigation of Lexical Densities in Selected Speeches of President Barack Obama”

**Abstract:**

This is a lexical research with focus on lexical densities in selected speeches of the American President, Barack Obama. The study objectives were to investigate whether President Barack Obama’s speeches are lexically dense or not; whether he has varying lexical density style and whether such variations (if established) are according to audience types, situations or any other identifiable factor(s). President Obama’s speeches were categorised into three broad types (speeches made on the floor of Senate on legislative matters; speeches to non-legislative audiences; and United States nation-wide speeches). Two speeches were selected from each category. The wordcount and word frequency per speech were analysed using AntConc, a corpus analysis tool. The words in each speech were first manually classified into their respective parts of speech using the 21st Century Chambers English Dictionary, and subsequently categorised into lexical or grammatical words as appropriate. Using the established lexical density formula, it was found that President Obama’s speeches were not lexically dense, a feature which enhances readability and comprehensibility. It was also found that President Obama varied lexical densities in his speeches according to situation and audience types.

**Paper #44.** Friday, March 4. 11:00-11:30. LIMONAR

**Patricio Moya and María Luisa Carrió-Pastor.** “Estrategias de intensificación en los comentarios sobre las noticias digitales: un análisis comparativo entre España y Chile”

**Abstract:**

The aim of this study is to describe the mechanisms of intensification used by Spanish and Chilean writers in comments about the news included in two digital newspapers: El País (Spain) and Emol (Chile), and in two different subject areas (political and sports). For our purposes, intensification is understood as a category of analysis within interactional metadiscourse, this includes those elements which allow writers to express certainty in a given text. To achieve this objective, we analyzed a corpus of 2,400 comments (600 comments for each writer’s country and each subject area). The results showed that, on the one hand, the comments of Spanish writers presented more boosters (intensification devices) (138.6 words for 10,000 words) than those written by Chilean writers (123.4 for 10,000 words). On the other, we found that the subject area did not influence decisively in the use of boosters in the four sub-corpora.

**Paper #45.** Thursday, March 3. 11:00-11:30. LIMONAR



**John Lee.** "Conversational network: A preliminary study on the Chinese Buddhist Canon"

**Abstract:**

We describe a method to analyze characters in a literary text by considering their verbal interactions. This method exploits techniques from computational linguistics to extract all direct speech from the text, and to build a conversational network that visualizes the speakers, listeners and their degree of interaction. We apply this method in a case study on the Chinese Buddhist Canon. With over 50 million characters in over 1500 texts, the Canon is a complex literary collection. We outline the steps in constructing the conversational network, including automatic syntactic analysis, and extraction of direct speech and speaker attribution. We then present a quantitative analysis on the protagonists and their verbal interactions. This method of analysis complements more traditional methods, and can potentially be applied to other works of literature.

**Paper #47.** Friday, March 4. 12:00-12:30. ADUANA

**Jessica Carmona-Cejudo.** "How many people in the making of Sloane 770? A corpus-based approach"

**Abstract:**

During the transcription of a late Middle English manuscript on Medicine (London, British Library, MS Sloane 770), a series of orthographic variations appeared, several of which seemed to be arranged following a predictable pattern. Should this prove correct, it may be a clue to posit the existence of two or more scribes who were involved in the copying of the codex, or else of the dialect of the MS being an example of Mischsprache that combines the dialects of the exemplar MS and that of the scribe. To ascertain whether the MS was written by more than one copyist or whether it is an example of the coexistence of two different dialects, a morphologically lemmatized corpus was built. This paper will present the results obtained after studying that corpus in order to verify either the original hypothesis is linguistically and scientifically based or not.

**Paper #48.** Wednesday, March 2. 12:30-13:00. RETIRO

**Gunnar Bergh.** "Using corpora to compile a dictionary of English and Swedish football vocabulary"

**Abstract:**

The language of football can be seen as the most widespread of all "special languages" (cf. Sager et al. 1980). Despite this, relatively scant scholarly attention has been paid to football language in this capacity (Lavric et al. 2008:5). A comprehensive lexicographic project was therefore started a few years ago in order to chart the central vocabulary and phraseology of this special language. The main aim of the work is the compilation of a bilingual dictionary of English and Swedish football language, where collected corpus data are organized in a formal database and supplied with linguistic metadata, translation equivalents, authentic examples, cross-references, etc. The present paper gives a work-in-progress account of the methodological aspects of this project. It features a description of (i) the means and principles of collecting football language material through a corpus-based approach; (ii) the different lexicographic tools employed, search applications as well as database software; and (iii) the formal organization of the material, including the linguistic field structure applied in the dictionary. So far, some 8 000 lexical entries, words as well as phrases, have been collected as a result of the data mining process.

**Paper #49.** Thursday, March 3. 16:00-16:30. ADUANA

**Elena Fernández de Molina Ortés.** " "Acho, qué perrino más chico". Recursos morfológicos y expresivos de los habitantes de Mérida (Badajoz)"

**Abstract:**

In this research we will present a study about the morphological and linguistics expressions features of an urban area, Mérida (Badajoz), in which we will compare traditional linguistics results about extremeñian variation with those currently obtained in this location. Also, and to test which is the use of the variants under study, we will use a sociolinguistic perspective allowing us to corroborate which features are specific of the Extremadura's main dialect and which ones are restricted to specific social groups.

**Paper #50.** Thursday, March 3. 17:00-17:30. LIMONAR

**Gustavo Adolfo Rodríguez Martín.** "The Times They Are A-changin' – and so are corpora"

**Abstract:**

Popular culture texts often make their way into the 'phrasicon' (i.e., the collection of phraseological units of a speech community). That seems to be the case with Bob Dylan's album and title song "The Times They Are a-changin'". This popular phrase, in addition, has spread beyond the borders of American folklore and has been borrowed and adapted into Spanish. The linguistic and cultural differences account for disparities in the use and stylistic meaning of the phrase in both languages, a fact that will be appraised through the use of corpus data.

**Paper #51.** Wednesday, March 2. 11:30-12:00. CALETA

**Mark Donnellan.** "An Introduction to Using Corpora with EFL Learners"

**Abstract:**

The use of corpora in the classroom represents an innovative way to enable English language learners to undertake independent study of lexical and grammatical patterns; however, only a limited amount of investigation into the use of corpora with students exists. This paper will first briefly introduce pertinent literature, which will give a basic overview of corpus linguistics. The paper will report on the use of the British National Corpus (BNC) and other corpus tools with students in a semester long course in a Japanese university with advanced EFL learners. These students undertook a series of tasks and projects, which allowed them to achieve the overall course goal of being able to conduct independent research into lexical and grammatical patterns. In order to assess student progress and to gather student opinions about the course and the use of corpora, data was collected in two ways: pre and post CEFR style student self-assessments, and a course reflection and evaluation survey. The results indicated that the students had progressed throughout the semester and that they had a largely positive opinion of the course. In concluding the paper, suggestions for teachers wishing to use corpora and corpus tools with their students will be offered.

**Paper #52.** Friday, March 4. 10:30-11:00. CALETA

**María José Barrios Sabador.** "A lo mejor, igual y lo mismo. Epistemic and non-epistemic meanings"

**Abstract:**

In this presentation we discuss the use of a *lo mejor* (maybe, perhaps) in oral texts – *Corpus de Referencia del Español Actual* – where the speaker doesn't show any lack of commitment to the proposition. In some of these texts, the speaker gives an example to substantiate his position; in others, he talks about everyday actions – characterized by showing an effective state of affairs and, as such, not subject to doubt –. There are also some occurrences of two other probability markers (*igual y lo mismo*), but they are only occasional in the corpus analyzed. In some contexts, a certain meaning of doubt can be understood, whereas in others such a meaning is unnoticeable. To account for these uses – not described in grammar – we provide some explanatory hypotheses of these markers based on the primary meaning of their components, their stage of grammaticalization and their pragmatic motivation.

**Paper #53.** Thursday, March 3. 13:00-13:30. LIMONAR

**Laura Hernández Lorenzo.** "El léxico poético de Fernando de Herrera: Un acercamiento desde la lingüística de corpus y computacional / The poetic word of Fernando de Herrera: An Approach through corpus and computational linguistics"

**Abstract:**

Great advances in Corpus Linguistics have led to new approaches in literary studies. At this work, we apply these new tools to the analysis of Golden Age Spanish poetry written by Fernando de Herrera, the author of "Anotaciones a Garcilaso de la Vega" (1580) and one of the greatest poets of his time. Through a keyword method combined with lexical concordances, we will try to overview principal characteristics and differences between subgenres in Herrera's poetry, dealing with the poems he published in life (known as H) and getting results which help in the academic debate about this poet's works and style.

**Paper #54.** Thursday, March 3. 13:30-14:00. ADUANA

**David Britain, Dominique Bürki, Tobias Leonhardt, Sara Lynch, Laura Mettler and Kazuko Matsumoto.** "English in las Islas de los Garbanzos: developing a multi-locality corpus of Micronesian Englishes"

**Abstract:**

We outline here the construction of a substantial corpus of Micronesian English. Fieldwork trips to five Micronesian Islands enabled the collection of a large corpus of recordings (>200 hours, >250 islanders, old/young, male/female, more/less educated and travelled). Recordings were transcribed using ELAN. We also report on some of the linguistic features being investigated: future tense marking; /d/ frication; postalveolar fricative devoicing; non-etymological /h/ insertion; grammaticalisation of terms of address into discourse markers. Investigating Micronesian English is important for a number of reasons. Firstly, it is an unexplored area, dialectologically. Secondly, attempts to present unified accounts of the development of Englishes worldwide are based on British rather than American colonialism. Finally, few Micronesian islands experienced significant English-speaking settlement. The linguistic motivation of the project is to investigate the emergence of unresearched new Englishes formed with little contact with the superstrate language. Significant evidence of substrate influence in the emerging Micronesian varieties is expected, therefore.

**Paper #55.** Friday, March 4. 12:00-12:30. RETIRO

**Hanna Skorczyńska.** "The emerging parties' manifestos for the 2015 Spanish general elections: a comparative analysis of lexical choices"

**Abstract:**

This study aims to provide quantitative data from the comparison of the election manifestos of two emerging parties: Ciudadanos (C) and Podemos (P). Two texts have been used for the study: "Ciudadanos – nuestras ideas" and "Programa marco". They were queried with Sketchengine's 'compare corpora' tool, which identified the key words in one corpus using the other as the reference corpus, and vice versa. The texts turned out to be very different reaching the comparability score of 4.61 (1.0 indicates that corpora are nearly identical). There were notable differences in the top 35 key words found. In C's text, these key words made references in general to the state organization, the EU, culture, market economy or immigration, while P's text focused more on the specific problems in Spain, their solutions, but also on child education, sports or rural development. C's key word list contained evaluative adjectives expressing necessity y urgency, while P's key evaluative adjectives had negative or positive connotations. Further analysis was concerned with the collocational patterns of selected keywords and key evaluative adjectives.

**Paper #56.** Friday, March 4. 12:30-13:00. RETIRO

**Guadalupe Nieto Caballero and Pablo Ruano.** "Estudio de corpus sobre patrones textuales comunes en la narrativa de Benito Pérez Galdós y de Charles Dickens"

**Abstract:**

In this presentation, Galdós's and Dickens's styles will be explored using a corpus-stylistic approach. More specifically, we will try to demonstrate that there exist stylistic similarities among the two novelists that have been hitherto unnoticed. Using a corpus-based methodology, some of these similarities will be herein discussed. As will be shown, Galdós resorts to some Dickensian strategies when building his fictive universes. The analysis will focus on clusters containing body parts (con las manos en los bolsillos and with his hands in his pockets, for example), which will help to both ratify Dickens's well-known influence on Galdós's style and unveil some unknown aspects in the exegesis of the Spanish author.

**Paper #57.** Friday, March 4. 13:00-13:30. RETIRO

**Guadalupe Nieto Caballero.** "La proyección del discurso en las novelas de Galdós a través de una metodología de corpus"

**Abstract:**

In this analysis, Benito Pérez Galdós's style will be explored using a corpus-based approach. Even though the study will deal with all Galdós's works (c. 6.2 million words), the analysis of examples will be limited to Episodios Nacionales (primera serie) (c. 700,000 words). The study will be divided into two parts. On the one hand, clusters containing the form *dijo* will be scrutinized. As will be seen, the different realizations of this speech verb can contribute to depict characters. On the other hand, the rest of speech verbs will be analyzed. More than 4,000 examples have been retrieved, whose analysis also sheds light on Galdós's techniques of characterization. Texts have been downloaded from the repository Cervantes Virtual and processed with WordSmith Tools 6.

**Paper #60.** Thursday, March 3. 11:30-12:00. LIMONAR

**Carmen Gregori-Signes.** "A corpus-based study of the representation of men in <<3<sup>rd</sup> Rock from the Sun>>"

**Abstract:**

The present paper resorts to Corpus Assisted Discourse Studies (CADS) to study the humorous representation of men in the television series "3rd Rock from the Sun". The study is an example of a cross-disciplinary field that is still somewhat under-subscribed and largely unexplored (Bednarek 2012). Dow (1996) argues that "television is a large part of American culture, and thus it is a significant and important site for sociological analysis." Sitcoms are, still today, one of the most successful TV genres. The linguistic analysis follows appraisal theory in particular the study of judgment, although the nature of the data sometimes demands to be complemented with multimodal analysis in order to provide a meaningful account of the results obtained using CL methodologies. There are many studies on women and how femininity is constructed through the media, however the focus on the humorous representation of masculinity on television sitcoms is still largely unexplored. The analysis expands on previous analyses which focus on women and uses the results as a benchmark against which the representation of men can be compared.

**Paper #61.** Thursday, March 3. 15:30-16:00. RETIRO

**Roger Gee.** "The Defining Vocabulary of Mid-Frequency Vocabulary in Online Dictionaries"

**Abstract:**

The study reported in this session examined the defining vocabulary found in five online dictionaries that was used for target words in the three to five thousand word frequency bands. The target words used for this study come from the 100,000 Word Frequency list developed by Davies (2013). The defining vocabulary corpus was analyzed using AntWordProfiler, once with a reference list of word forms and again with a reference list of word families. A third analysis using lemmas was done with Text Lex Compare. Contrary to what might logically be expected, it was found that a substantial proportion of the defining vocabulary was less frequent than the target words. Assuming that more frequent words are known before less frequent words, this result suggests that the online dictionaries used in this study may not allow for unassisted use by upper intermediate and even many advanced English language students.

**Paper #62.** Wednesday, March 2. 13:00-13:30. RETIRO

**Ryo Tsutahara.** "Corpus-based lexical semantic analysis of minimal pairs of deverbal adjectives with the Spanish -dor/-nte suffixes"

**Abstract:**

In this study, I examined the minimal pairs of Spanish deverbal adjectives with "-dor/-nte" like "secador" y "secante" to explain the different usages of two types of deverbal adjective that are considered highly similar. More concretely, I focused on the nouns modified by each adjective of the minimal pairs. According to the analysis conducted, the two types of adjective modify nouns of different semantic types. The analysis was conducted with the Sketch-diff function of Sketch Engine, which shows the typical nouns modified by each of the two adjectives analyzed. The main proposal of this study is that the crucial semantic features are "±volitional" and "±causative"; with regard to the adjectives with "-dor," at least one of the two values of the nouns modified should be positive, and in the case of the adjectives with "-nte," both features cannot be positive simultaneously.

**Paper #63.** Friday, March 4. 11:00-11:30. ADUANA

**Jesús Romero-Barranco and Soluna Salles-Bernal.** "Native and borrowed negative prefix competition in early modern English medical writing"

**Abstract:**

Early Modern English is characterized by an extraordinary lexical growth motivated by two main linguistic processes, borrowing and word-formation (Nevalainen, 1999, p. 332). Thus, native and foreign forms competed to achieve a position in the lexicon, not only increasing the English vocabulary but also modifying the traditional derivational system of the language by the introduction of foreign affixes (Lutzky, 2004; Cowie, 2002; Baugh & Cable, 1993). Of both, prefixation grew proportionately more than suffixation during the early Modern English period, probably because prefixes have fewer word-class restrictions and they do not change the word-class of the base (Nevalainen, 1999, p. 355). However, scarce research has been done on the development of prefixation in this period, apart from the studies of McConchie (2000) and Mladenovic (2014). All this considered, the present paper pursues the following objectives: 1) to analyse the use of the native un- and the borrowed in-, non- and -dis in early English scientific prose; 2) to assess their across text types; and 3) to classify the bases to which the prefixes are attached (native or foreign). The data comes from the Málaga Corpus of Early English Scientific Prose and the Corpus of Early English Medical Writing.

**Paper #64.** Thursday, March 3. 13:00-13:30. ADUANA

**Javier Calle Martín.** "English Intensifiers in Competition: this/that in American English"

**Abstract:**

Intensifiers are constantly immersed in a process of linguistic renovation to develop new and efficient mechanisms for the expression of intensity. The history of English provides many examples where an old intensifier is progressively replaced by a more effective expression of degree (Méndez Naya 2010: 191-213; Calle-Martín 2014: 399-419). The boosters this/that and so share the same semantic connotations for the expression of a high degree of intensity, with the meaning of 'to this/that extent, so much, so', becoming a typical case of intensifiers in competition. The present paper has been conceived with the following objectives: (a) to trace the development of these intensifiers in British and American English; (b) to analyse the frequency of the construction from a variationist perspective, both across registers and text types; (c) to cast light on the lexical semantic structure of the right-hand collocates in terms of their mode of construal and their attitudinal features. The source of evidence for this study comes from the Corpus of Historical American English (COHA) and the Corpus of Contemporary American English (COCA).

**Paper #65.** Friday, March 4. 11:30-12:00. RETIRO

**Mikhail Kruzhkov.** "Supracorpora databases as corpus-based superstructure"

**Abstract:**

This report presents a new type on corpus-based information resource: supracorpora databases (SCDBs). SCDBs enhance functionality of modern linguistic corpora by supporting customizable annotation of multi-word linguistic items. This is similar to query result categorization function available in some corpora but SCDBs take it further by allowing multifaceted annotation of linguistic items. More importantly, SCDBs allow researchers to create annotated translation correspondences (monoequivalences, MEs) in parallel translation corpora. Aggregation of searchable MEs in a SCDB represents a unique information resource that facilitates creation of new explicit knowledge about cross-linguistic correspondences and translation models. The work reported here was supported by Russian Foundation for Humanities, Grant No. 15-04-00507.

**Paper #66.** Friday, March 4. 10:30-11:00. ADUANA

**Tamara García-Vidal.** "A corpus- and variationist-based approach to the comparative and superlative degree late ME and EModE English adjectives"

**Abstract:**

Scholarly interest in the English adjective comparative system has grown recently in corpus linguistic research. The number of aspects in the evolution of the English adjective comparison has either been considered controversial or uncompleted; since some authors consider that they come from an already existing native resource while others state that they entered into the synthetic English language by external influences. The present corpus-based study offers a better understanding of the evolution of periphrastic comparatives from 1420 to 1710 in English. A variationist investigation was undertaken by analysing synthetic and analytic comparative forms in historical documents contextually belonging to different text-types, collected from the Helsinki Corpus. The results show how the periphrastic form of adjectives was at its peak during the last part of the Middle English period and first part of the Early Modern English one in more formal texts. However, this analytic form of adjectives was steadily outnumbered by inflectional adjectives at the end of the Early Modern English period. In addition, according to the data obtained in this study, it seems that standardisation and prescriptivism reinforced the social downgrading of periphrastic forms, once they were established among the speakers from the higher social positions.

**Paper #68.** Wednesday, March 2. 19:00-19:30. RETIRO

**Luis Miguel Puente Castelo and Leida María Mónaco.** "It is proper subferviently, to inquire into the nature of experimental chemiftry: Difficulties to harmonize disciplinary particularities and compilation criteria during the selection of samples for CEChET"

**Abstract:**

The aim of this paper is to discuss the process of compilation of the beta version of CEChET, the subcorpus on Chemistry in the Coruña Corpus of English Scientific Writing, and to reflect on the difficulties faced during this process and how they were overcome. The paper will examine the historical context of science and, particularly, chemistry in the period, and how this affects the process of compilation. Attention will also be paid to the compilation criteria used in the whole Coruña Corpus, including those regarding the adequacy of authors and samples (Moskovich 2012), and how these criteria have been applied to the compilation of CEChET in order to make it representative of the disciplinary practices of the time. Finally, the corpus will be briefly described on account of a series of parameters: their topics, the size of the samples, their temporal distribution throughout the period, as well as the geographical origin and sex of the authors represented in it.

**Paper #69.** Wednesday, March 2. 19:00-19:30. LIMONAR

**Martina Werner and Claudia Resch.** "Seeking law of the jungle: nominalized infinitives in the Austrian Baroque Corpus"

**Abstract:**

The Austrian Baroque Corpus (ABaC:us) is a digital collection of German Baroque texts of mostly first known editions and rare specimens from different monastic and public libraries. The freely available and STTS-tagged corpus (<https://acdh.oew.ac.at/abacus/>) was compiled at the Austrian Academy of Sciences (2010-2015). The STTS-tagger, developed for present day German (PDG), was adapted for the Baroque texts through several formal and manual revision processes. In our talk, we present results from a corpus-based study on nominalized infinitives (NIs) which are special morphosyntactically – especially as regards restrictions in pluralizability, definiteness, and reflexivity. While some of these linguistic restrictions have been neutralized diachronically, others can still be attested in PDG. The descriptive generalizations from ABaC:us and other corpus data (DDD on Old High German, MHDBDB on Middle High German; DWDS on PDG) suggest a diachronic emergence of a new nominalization pattern of deverbal nouns in terms of a grammaticalization.

**Paper #70.** Wednesday, March 2. 18:00-18:30. RETIRO

**Maria Iakovou, Olga Dima, Irianna Vasiliadi-Linardaki, Sofia-Nefeli Kitrou, Flora Vlachou, Marina Koutsoubou, Tatiana Katsina, Stavrialea Perrea, Froso Pappa, Xristina Kostakou and Maria Kavvadia.** "SEPAME2: A new learner resource for the Greek as a L2"

**Abstract:**

SEPAME2 meaning Learner Longitudinal Corpus for Greek as a Second Language (L2) is the first systematic attempt for Greek as L2 at the collection of developmental data, written and oral. The corpus data are derived by different L1 background adult learners of all language levels (A1-C2) studying Greek at the Modern Greek Language School of the Athens University. Learners' productions are the outcome of different types of graded tasks that learners have to perform in relation to their language level and their cognitive abilities. Notably, different feedback types are applied, the results of which are evaluated in relation to the learners' developmental progress. These three features, longitudinal dimension, duality of mode in language production and personalized feedback, provide the innovative framework on which our project is built. Our presentation will focus on the task design criteria related to our learners' profiles as well as their level performances.

**Paper #71.** Thursday, March 3. 16:00-16:30. RETIRO

**Oleksandr Kapranov.** "Corpus Analysis of Discourse Markers in Corporate Reports Involving Climate Change"

**Abstract:**

This conference presentation involves a corpus analysis of discourse markers in corporate discourse on climate change. Following Louwse and Mitchell (2003), discourse markers in business communication in general and in the genre of corporate annual reports in particular are regarded as cohesive devices which facilitate coherence relations, marking transition points within and between sentences (Taboada, 2006). In this study, the corpus consists of climate change and environment sections of annual reports by BP, Exxon, Shell and Statoil between 2010 to 2015. Second, the corpus is analysed quantitatively in software program WordSmith (Scott, 2012). Corpus investigation involves lexical measures yielded by WordSmith. Data analysis indicates that discourse markers do not exhibit significant changes in terms of their frequency in the corpus between 2010 and 2015. Whilst discourse markers identified in corporate discourse on climate change are not significant, they tend to be characterised by a certain degree of variability. Further analysis of these findings have been carried out within the framework of Rhetorical Structure Theory, where textual coherence and organization are conceptualised as relations between the constitutive parts of the text. These findings will be presented in detail and further discussed at the conference.

**Paper #72.** Thursday, March 3. 17:00-17:30. CALETA

**María Alcántud-Díaz.** "Corpus-based tasks: using a novel to teach language and literature"

**Abstract:**

The present talk illustrates how corpus-based activities derived from an exploration of a novel using corpus tools can be used to teach English language and literature. The analysis previous to the actual elaboration of tasks followed a corpus stylistic approach in the analysis of the first two books of the successful historical-fiction trilogy *All Souls* by Deborah Harkness. The use of corpus linguistics tools facilitated the finding of useful information (different semantic fields related to geography and history) as well as the identification of lexical phrases, patterns or extended units of meaning. Once this information was obtained, the authors resorted to the principles of collaborative learning in order to design activities (jigsaw, a video production, a peer review questionnaire and peer-evaluation) based on the data and the results obtained using CL tools.

**Paper #73.** Friday, March 4. 10:30-11:00. RETIRO

**Ana M<sup>a</sup> Fernández Soneira, Inmaculada C. Báez Montero and Eva Freijeiro Ocampo.** "CORALSE: design of a corpus of Spanish Sign Language"

**Abstract:**

The approval of the law for the recognition of Sign Languages and its subsequent development (together with the laws enacted by the regional governments and the work of universities and institutions such as CNLSE) has changed the landscape of the research activity carried out in the field of SL in Spain. In spite of these social advances, a corpus of Spanish Sign Language (LSE) has not yet been compiled. The average Sign Language corpus is traditionally composed of collections of annotated or tagged videos that contain written material aligned with the main Sign Language data. The compiling project presented here, CORALSE, proposes: 1) to collect a representative number of samples of language use; 2) to tag and transcribe the collected samples and build an online corpus; 3) to advance in the description of the

grammar of LSE; 4) to provide the scientific background needed for the development of materials for educational purposes; and 5) to advance in the development of different types of LSE.

**Paper #75.** Friday, March 4. 12:00-12:30. LIMONAR

**Miguel Ángel Candel-Mora and Olga Torrest-Hostench.** "Corpus-based statistical machine translation of Big data: orality features in consumer generated content"

**Abstract:**

Big data analysis has attracted considerable attention in recent years due to its numerous research opportunities and business applications in almost every field of knowledge. Similarly, Web 2.0 and the active participation of users has favoured one of the largest sources of big data susceptible of research from the point of view of natural language processing, computational linguistics and machine translation. Text data analysis appears as a challenging approach to explore features of machine translation that could not be addressed otherwise. The objective of this study is to determine the behaviour of corpus-based statistical machine translation engines in the translation of these orality features from the viewpoint of post-editing strategies: from a corpus of 1,000 reviews originally written in English and its contrast with its Spanish MT output, this work attempts to identify the unique features of orality of this new digital genre and how SMT handles them.

**Paper #76.** Friday, March 4. 11:30-12:00. CALETA

**Mayowa Akinlotan.** "The effects of complexity and proficiency on determiner number (dis)agreement in the Nigerian English noun phrase"

**Abstract:**

The paper investigates the extent to which the grammatical number (dis)agreement hypothesis in the New Englishes (Platt et al 1984, Gorlach 1998, Mesthrie et al 2008) manifests in the determiner system of the Nigerian English, and how the variables of proficiency, text type, register, structural complexity, and determiner form influence scenario found. Applying principle of accountability (Labov 1972, Tagliamonte 2012), together with test statistic on data drawn from the Nigeria-ICE, it is found that grammatical number in the Nigerian determiner system is more likely to agree (98%) than to disagree (2%); a scenario mainly influenced, at varying significant degree, by complexity and proficiency. The 2% further shows that the irregularity is more likely to form with the use of quantifier or demonstrative than with the use of the indefinite article. We argue that this scenario suggests a manifestation of fossilisation by transferring from the syntactically unique determiner systems of the local Nigerian languages to Nigerian English.

**Paper #77.** Wednesday, March 2. 16:00-16:30. CALETA

**James Elwood, Katerina Petchko and Gavin O'Neill.** "Graduate Writing in English at a Multi-Polar Institute in Japan: A Corpus Analysis"

**Abstract:**

This presentation will address academic writing produced by 1150 international students over a 10-year period at an English language-medium, multi-polar graduate institute in Japan. The student body includes students from myriad educational backgrounds with a concomitant range of English proficiencies, and to instruct such a diverse group, keen awareness of the students' writing proficiency and shortcomings facilitates completion of their respective programs. Based on corpus analysis, we will first explore lexical profiles and the use of academic lexis (Coxhead, 2000) across the three types of corpora. Second, we will examine verb usage, an important yet problematic area for these graduate students. The study utilized three corpora: the seven-million-word Graduate Policy Studies Corpus of Written English incorporating the students' final policy papers, smaller discipline-specific corpora culled from leading journals, and general-use corpora. The presentation will conclude with implications from these results and suggestions for further uses of locally-created corpora for very heterogeneous student populations.

**Paper #79.** Thursday, March 3. 18:00-18:30. LIMONAR

**Marlén Izquierdo and María Pérez Blanco.** "The herbal tea promotional text in English and Spanish: a contrastive rhetoric study"

**Abstract:**

This study aims to contrast herbal tea promotional texts in English and in Spanish at the rhetorical level. Making use of the move-step method proposed by Swales (1990), a comparable corpus of 150 texts per language was rhetorically tagged. Even though the texts are fairly short, a fairly complex macrostructure of a textual genre heretofore never studied was observed. As expected of any genre, such a macrostructure is rather flexible (Swales 1990). Thereby the analysis carried out reveals that some moves -and steps- seem to be more central or necessary than others, whose typicality is lower. In addition, the study brings to the fore what moves and steps are preferred in each language. A

preliminary look at the lexico-grammar of the genre suggests its dominant evaluative function. Results will be applied in the development of a writing tool.

**Paper #80.** Friday, March 4. 13:00-13:30. LIMONAR

**María Teresa Ortego Antón, Elena Jiménez García and Purificación Fernández-Nistal.** "The Design and Compilation of a Corpus to Analyse Undergraduate Students' Communication in Higher Education Scenarios through Social Media"

**Abstract:**

This paper seeks to contribute to the recent academic study of linguistic behaviour of undergraduate students in social networking sites used as academic and educative communication tools. In order to get an approach to this new core form of communication, in the first stage we describe the methodology used to design a Spanish monolingual corpus. The corpus is composed of samples taken from the communicative output among undergraduate translation students in social networking sites. Consequently, we set the parameters which give balance and representativeness to the sample and we describe the corpus features, such as the population, the delimitation of social networking sites, the size and the subject area coverage. Finally, we provide a detailed description of the methodology used to analyse errors produced when undergraduate students communicate through social networking sites.

**Paper #81.** Thursday, March 3. 19:00-19:30. LIMONAR

**María Teresa Ortego.** "An Approach to the Strategies Used to Translate Social Media Terms from English into Spanish"

**Abstract:**

In recent decades, the number of new concepts and terms has arisen rapidly in computing and, more precisely, in social media. Additionally, expert knowledge, which used to be exclusive for experts, also interests middlebrow language users as a result of the democratisation of education and the media broadcasting. Moreover, in the current globalised world, the need to transfer scientific knowledge to other languages such as Spanish arises, since it is produced or spread mainly in English. Taking into account the need of more studies which deal with the translation of specialised vocabulary, in this paper we try to get an approach to the main strategies used when translating social media vocabulary from English into Spanish. First, we compiled a comparable corpus composed of texts about social media published in the main British and Spanish newspapers. Then, we selected the most frequent terms and analysed how the interlinguistic transfer was made. From the results of the analysis we will get the most common strategies used when translating social media terms from English into Spanish.

**Paper #82.** Wednesday, March 2. 12:00-12:30. CALETA

**Veronica Colwell and Hugo Sanjurjo.** "Corpus-based applications in the EFL for the workplace: Getting the recipe right"

**Abstract:**

This presentation reports on how recently designed EN-ES bi-writing apps that serve to support small Spanish businesses which struggle to meet the challenges posed by English as a lingua franca in the market place, along with the custom-made comparable corpora from which these apps are built, can serve as highly relevant learning resources for improving editing skills in the L1 Spanish EFL undergraduate class for professional and for academic purposes. The report focuses on one such app, specialized by genre, the BiTeX bi-writing app, which helps produce recipes in English and in Spanish. Results suggest that such relevant workplace resources concerned with English for international communication are particularly well suited as they create real-life opportunities for consciousness-raising and language awareness-raising while promoting ongoing development in a range of skills and competences pertinent to the professional future of such learners.

**Paper #83.** Thursday, March 3. 13:30-14:00. LIMONAR

**María Teresa Ortego and Purificación Fernández-Nistal.** "Errors produced by English Interference in the Academic Communication among Spanish Students through Social Media: a Case Study"

**Abstract:**

The use of social networking sites in the academic community has become very popular in recent years, especially among undergraduate students. Hence, new forms of interaction between students have arisen. Provided the importance of these new communication scenarios in higher education and the amount of errors produced in Spanish communication due to the interference of English language, in this paper we set the parameters to analyse the errors made by Spanish undergraduate students when communicating in academic scenarios through social media. We focus on some Spanglish cases. The results might provide us with data to create new tools to improve academic communication through social media.

**Paper #84.** Wednesday, March 2. 17:00-17:30. ADUANA



**Rosa Rabadan, Hugo Sanjurjo and Veronica Colwell.** "Bi-Texting your food: Helping the gastro industry reach the global market"

**Abstract:**

The technologization of cross-linguistic communication and the expansion of the learning of foreign languages has helped create new, non-linguist users. Corpus-based applications offer a way of responding to these new challenges. This presentation focuses on the design and building process of the BiTeX bi-writing app. The starting point is a custom-made, rhetorically and POS annotated, En-Es comparable corpus of recipes containing 135,912 words in the En and 145,449 in the Es subcorpus respectively. The BiText prototype has been developed using MongoDB, Express, Node.js and jQuery, providing a fast and efficient architecture for data-intensive treatment for real time applications. BiTeX aims at helping improve international communication in the restaurant and catering community, as well as boosting collateral business niches including recipe books, tourist-oriented websites, etc.

**Paper #85.** Wednesday, March 2. 15:50-16:00. LIMONAR

**Natalia Judith Laso, Emilia Castaño and Isabel Verdguer.** "Migration law metaphors in an ESP learner corpus"

**Abstract:**

Metaphorical expressions are pervasive in language (Lakoff & Johnson 1980; Lakoff & Johnson 1999; Kövecses 2005). This is not surprising since human beings usually conceptualize abstract concepts (target domain) in terms of concrete or more structured domains (source domain) through mappings that take the form TARGET DOMAIN IS SOURCE DOMAIN. Thus, metaphors are not seen as mere stylistic devices but as cognitive templates that are the surface realization of cross-domain mappings (Lakoff 1993). This paper reports on a qualitative corpus-based study on EFL learners' metaphorical production. The analysis of a learner corpus of legal English, which included essays written by undergraduates, showed that learners use metaphorical language and that the metaphorical expressions identified are mainly based on the following mappings: MIGRATION IS A NATURAL FORCE, STATES ARE PEOPLE, LAWS ARE WEAPONS OR REMEDIES, as stated in the literature (O'Brien 2003; Biria 2012; Charteris-Black 2006; Winter 2009; Cunningham-Parmeter 2011). Bearing in mind that legal discourse should be characterized by a neutral tone, it is important that learners take a critical approach when using certain metaphorical expressions which may have negative connotations. The explicit teaching of conceptual metaphors and their linguistic expressions may contribute to raising learners' metaphorical awareness.

**Paper #87.** Friday, March 4. 11:30-12:00. ADUANA

**Soluna Salles.** "Punctuation patterns in a 17<sup>th</sup> medical manuscript: corpus-based study of G.U.L. MS 303, Treatise on the Diseases of Women"

**Abstract:**

The regularization of punctuation and the orthographic system is pointed as one of the distinctive features of early modern English (Lass, 1999, p. 8). In the case of handwritten material, punctuation depended greatly on the scribe's preference and it was considered to be rather unpredictable. However, recent studies have demonstrated that punctuation followed particular patterns depending on the typology of the text (Calle-Martín, 2004; Calle-Martín & Miranda-García, 2005; Esteban-Segura, 2009; Marqués-Aguado, 2009, 2011; Alonso-Almeida & Ortega-Barrera, 2014). This paper seeks to research the patterns and functions of punctuation marks occurring in a 17th century medical manuscript entitled Treatise on the Diseases of Women (Glasgow University Library, MS Hunter 303) following the theoretical framework proposed by Lucas (1971) and Görlach (1992). Additionally, punctuation uses are considered within the different genres occurring in the manuscript. We postulate that the process of vernacularisation could have influenced the development of a stable and ordered punctuation system in the early modern English scientific prose. Data have been obtained from the Málaga Corpus of Early Modern English Scientific Prose.

**Paper #88.** Thursday, March 3. 16:30-17:00. LIMONAR

**Qiuorong Zhao.** "Translation-induced language variation and change: A case study of the impact of English-Chinese translations on the use of the Chinese Preposition Zai"

**Abstract:**

Zai is one of the most frequent prepositions, which has experienced a dramatic change from verb to preposition in modern Chinese. This paper focuses an examination of trends in the use of the prepositional phrase zai in translated Chinese and non-translated native Chinese. A multiple diachronic corpus will be used for this study, consisting of the comparable corpus (1910-1949), an English-Chinese parallel corpus (1930-1949) and a reference corpus, which consists of approximately 7.6 million words, 134 works. The findings show that in both translated Chinese and non-translated native Chinese, the frequency of the prepositional phrase zai increased in the time-span under consideration. The scope of the prepositional phrase became extended and the structure of the prepositional phrase lengthened. Translations from English can reasonably be assumed to

represent a factor in this change, in particular given the favorable social context, which in general must be seen as an important factor in determining the potential of translation to influence native non-translated target language production.

**Paper #89.** Wednesday, March 4. 18:00-18:30. CALETA

**Lina Bikeliene.** "Evaluative language in Lithuanian students' writing. A study of 'interesting' and 'important'"

**Abstract:**

Critical thinking, as one of the competences to be developed at university, requires the ability to evaluate. Though different parts of speech can fall under the category of evaluative language, adjectives are among the most frequently used means to express evaluation. The present study is limited to the use of two positive evaluative adjectives in Lithuanian learner English. It aims to analyse whether the use of 'interesting' and 'important' in Lithuanian students' writing depends on the text genre, language proficiency, educational institution, and whether it differs in NNS and NS writing. The data was obtained from the Lithuanian parts of the Corpus of Academic Learner English and the International Corpus of Learner English, the corpus of first year students' examination essays written at Vilnius University, and comparable native English corpora. The results of the study highlight the different roles the variables under consideration play in the use of the studied adjectives.

**Paper #90.** Thursday, March 3. 17:00-17:30. RETIRO

**Jesús Fernández-Domínguez.** "Right-headed compound verbs in Spanish and the influence of English"

**Abstract:**

Verb compounds with the structure [N + V] are generally considered archaic and infrequent in Contemporary Spanish, but two recent tendencies have been reported against this view. First, Spanish verbs with the structure [N + V] ("bioestimular") are based on previously existing nouns and generate back-formed compound verbs with a Latin/Greek initial element ("bio-"). A second process is one where both elements are native, again involving back-formation, from a Spanish noun or adjective. This happens in cases like "bocabrir" (from "bocabierto"). This paper approaches the structure, formation and re-flourishing of Spanish [N+V] verb compounds in order to evidence the parallels between them and their English counterparts. Data from Spanish and English corpora is used to detect frequencies with the aim of detecting to what extent Spanish [N + V] verb compounding is revitalising, and whether there is structural influence from English in the process.

**Paper #91.** Thursday, March 3. 16:30-17:00. CALETA

**Maggie Charles.** "Checking grammar or hunting for vocabulary? Users report on their independent consultation of self-compiled corpora"

**Abstract:**

This paper investigates the types of searches that users carry out when they consult a self-compiled corpus outside class and for their own purposes. The study presents data from 76 users one year after they took an EAP course in which they built personal corpora of research articles in their own fields. Most respondents (57%) built corpora of between 100,000 and 499,999 words, while a further 17% compiled corpora that were 500,000 words or over. Two types of search are distinguished: 'checking searches' aim to improve lexico-grammatical accuracy, while 'hunting searches' are designed to enhance language and content. Users reported more frequent and more successful checking than hunting searches, with prepositions, articles and collocations all frequently checked. This paper presents further data on the searches reported by independent users and discusses the implications for designing courses to facilitate more successful corpus use.

**Paper #95.** Thursday, March 3. 17:00-17:30. ADUANA

**Evelyn Gandón-Chapela.** "The architecture of ellipsis in English: on lexical and syntactic distance"

**Abstract:**

In this paper, I undertake a corpus-based analysis of the lexical (in number of words) and syntactic distance (in number of clauses) between the antecedent clause(s) and the ellipsis site in examples of Post-Auxiliary Ellipsis (two subtypes: pseudogapping and VP ellipsis) in Modern English, using data from the Penn Parsed Corpus of Modern British English (1700–1914). The analysis has revealed that most instances of pseudogapping occur in contexts where the antecedent is contained within the same sentence, the target of ellipsis appears in a different clause, there are no intervening clauses and, in addition, the lexical distance ranges from 0-10 words. Regarding VP ellipsis, in the vast majority of the examples the lexical distance is lower, i.e. 0-5 words. VP ellipsis differs from pseudogapping with respect to the most frequent type of boundedness established: the antecedent and the ellipsis site may either appear in different sentences or within the same sentence and in a different clause (with no intervening clauses) respectively.

**Paper #96.** Wednesday, March 2. 13:00-13:30. LIMONAR

**Carlos Prado Alonso.** "The 'that is' parenthetical construction and the indirect expression of attitude in academic texts"

**Abstract:**

Parenthetical constructions have been the subject of extensive research from a functional perspective (cf. Blakemore 2005, 2007; or Dehé 2014; Dehé and Kavalova, 2007, among others). This study is a further contribution to this line of research and offers a corpus-based analysis of one type of parenthetical construction namely, the that-is parenthetical construction, as shown in (1). 1. The binomial model may still be used as a baseline for comparative purposes; that is, we suggest to discuss the phenomenon in terms of its departures from the binomial model. This paper presents an in-depth corpus-based analysis of the factors that drive the pragmatic use and distribution of that-is constructions in different academic texts. It usually argued that parenthetical clauses are the result of a stylistic choice or the result of a sort of on-line reformulation and revision (cf. Blakemore, 2005). Beyond that, however, the analysis of the data retrieved from the corpora have shown that, in academic prose, the use of that-is parentheticals is also sensitive to more precise dimensions of linguistic variation. In sum, the paper shows that these parenthetical can also be used as discourse markers through which the expression of attitude by the writer is shown in discourse.

**Paper #99.** Thursday, March 3. 19:00-19:30. ADUANA

**Francisco Gago Jover.** "The Old Spanish Textual Archive (OSTA): diseño y desarrollo de un corpus de textos medievales"

**Abstract:**

The Old Spanish Textual Archive is a corpus of about 35,000,000 words from more than 500 manuscript and printed texts produced in Spain between the thirteenth and the sixteenth centuries. All the texts were transcribed using the same, homogeneous criteria. The corpus interface allows searches of various types: exact words (medieval), wildcards, lemma (modern), part of speech, chronology (original production date and specific production date), thematic tagging, or any combination of these. The paper will discuss the main difficulties faced in the elaboration of the Old Spanish Textual Archive, among them: 1) delimitation of the textual corpus (differentiation between transcription, codex and work), 2) morphological tagging (after extensive research it was decided to use a customized version of Freeling), and 3) thematic categorization of the corpus.

**Paper #100.** Wednesday, March 2. 18:30-19:00. CALETA

**Miharu Fuyuno, Yuko Yamashita, Takeshi Saitoh and Yoshitaka Nakajima.** "Semantic Structure, Speech Units and Facial Movements: Multimodal Corpus Analysis of English Public Speaking"

**Abstract:**

This study examines connections between the semantic structure and speech units, and characteristics of facial movements in EFL learners' public speech. The data were obtained from a multimodal corpus of English public speaking constructed from digital audio and video data of an official English speech contest held in a Japanese high school. Evaluation data of contest judges were also included. For the audio data, speech pauses were extracted with an acoustic analysis software, and the spoken content (text) of each speech unit embedded between two pauses was then annotated. The semantic structures of the speech units were analysed based on segmental chunks of clauses. Motion capturing was applied on video data; forty-two tracking points were set on each speaker's eyes, nose, mouths and face lines. The results indicated: (1) Speakers with higher evaluations showed a similar semantic structure pattern in their speech units. It was also confirmed as similar to that for NSE samples. (2) Horizontal facial movements and the angles of face rotations were extracted from motion capturing. The result is expected to be useful for defining a facial movement model that effectively describes good eye contacts in public speaking.

**Paper #101.** Wednesday, March 2. 19:00-19:30. CALETA

**Shin'Ichiro Ishikawa.** "V-THAT Structures in Asian Learners' L2 English Speeches and Essays"

**Abstract:**

V-THAT structures, in which verbs take a THAT-complement clause in post-predicate position, are said to account for over 80% of all types of THAT-clauses. Biber et al. (1999) analyzed their structural patterns, common controlling verbs, and omission of the complementizer. However, how Asian learners with varied L1 backgrounds and L2 proficiency levels use V-THAT in written and spoken English has not been wholly elucidated. Therefore, using the International Corpus Network of Asian Learners of English (ICNALE), we investigated the use of major V-THAT structures by Asian learners of English (ALE) as well as English native speakers (ENS). Our analyses have suggested that (1) ALE use V-THAT less than ENS both in speeches and essays, though the difference is not significant; (2) Learners in Japan, Korea, and Thailand relatively overuse V-THAT; (3) Both of ALE and ENS frequently use "think," "agree," "say," and "know," and many of

which are mental verbs; (4) Learners with different L1 backgrounds use different types of controlling verbs; (5) THAT-omission rate is approximately 80% and 60% for ALE and ENS respectively; (6) learners' speeches and writings are classified into four major clusters. These findings will shed a new light on English teaching for Asian learners.

**Paper #102.** Thursday, March 3. 11:00-11:30. CALETA

**Cristóbal Lozano and Ana Díaz-Negrillo.** "Using learner corpus methods to explore variability in L2 acquisition: revisiting the Morpheme Order Studies"

**Abstract:**

This paper investigates the Morpheme Order Studies (MOS) phenomenon within Learner Corpus Research. In particular, it uses a new methodological improvement, namely Interlanguage Annotation (ILA), on COREFL (CORpus of English as a Foreign Language), an on-going corpus from EFL Spanish secondary-school students. The study explores morpheme variability across four proficiency CEFR-based levels in COREFL (A1-B2). The general results partially replicate previous MOS results, but diverge in systematic ways. The fine-grained ILA approach defended here and the exploration of case studies reveals key findings that have gone unnoticed in previous research.

**Paper #103.** Wednesday, March 2. 13:00-13:30. CALETA

**Ali Sukru Ozbay and Mustafa Naci Kayaoglu.** "Computerized corpus-based investigation of the use of multi-word combinations and the developmental stages by tertiary level EFL learners"

**Abstract:**

It is an old consensus by now that the EFL learners all around the globe are having various problems which may turn their abilities to produce essays into a challenge in terms of lexical complexity, phraseological richness and syntactic variation. Part of the reason for the conduct of the present study, which is mainly a computerized contrastive learner corpus analysis of BAWE and KTUCALE corpora, is to make an analysis of the general frequency, variation and development of multi-word combinations. For this purpose, the expository and academic essays written by tertiary level EFL students of English were taken as samples for the study. In the selection of the samples for the study, purposive sampling methodology was used and the analysis was done by comparing the language use of the learners in two corpora. Quantitative analysis of the learner corpora findings was followed by a qualitative analysis of the learners' responses to the protocol questions and interviews. The results revealed that KTUCALE turns out to be much less complex in terms of MWC diversity, phraseological variation and density than the reference corpus BAWE. The others are that MWC samples from KTUCALE indicated overuses and underuses.

**Paper #104.** Thurs, March 3. 16:30-17:00. ADUANA

**Iván Tamaredo.** "The Effect of Frequency of Co-occurrence on Subject Pronoun Omission in Indian English and Singapore English"

**Abstract:**

This paper deals with the effect of frequency of co-occurrence on pronoun omission in subject position in Indian English and Singapore English. Frequency has been found to influence the mental representation of linguistic structure: a string of words that co-occur very frequently may be stored in memory as a single unit. Our hypothesis is that, if a pronoun and a form of a verb occur very frequently in combination, they may form a single unit and, as a consequence, the pronoun would no longer be an independent element, which would inhibit its omission. This hypothesis will be tested by means of a corpus study in which the percentage of omission of each pronoun-verb combination will be calculated and compared with its co-occurrence frequency. The results of a pilot study support the hypothesis: the probability of a pronoun being omitted decreases as the co-occurrence frequency of the pronoun and the verb increases.

**Paper #105.** Thursday, March 3. 18:00-18:30. RETIRO

**Ali Sukru Ozbay.** "The Potential of Corpus-Based Collocation Instruction on the Awareness Levels of the Turkish EFL Students in Terms of Reading Performances"

**Abstract:**

The fact that languages consist of prefabricated chunks or multi word combinations important for perceiving and producing language for native speakers make collocation instruction vitally significant for EFL learners in their efforts to understand and produce natural and fluent language. Thus, the purpose of this research is to explore whether corpus based explicit collocation instruction will help the EFL students gain awareness of collocations and recognize collocations in different contexts and finally result in better reading performances in English. The data were obtained through pre-test and post- test scores and group interviews. Tertiary level EFL students (n=86) from the English department of a middle sized university in the Eastern Black Sea region in Turkey participated in the study and the study lasted for a year. The study investigated whether there were any differences between the experimental and the control

groups in terms of gaining awareness of collocations and exhibiting better reading performances after corpus-based explicit collocation instruction. Based on the analyses of students' reading scores, the main findings showed that the experimental group showed significant improvement compared to the control group. Another finding was that corpus based collocation instruction has had positive effects on awareness and reading performances.

**Paper #106.** Thursday, March 3. 18:30-19:00. RETIRO

**Carlos Perrián-Pascual and Eva M. Mestre-Mestre.** "Discovering terms for glossary construction"

**Abstract:**

The goal of this work is to outline the main features of DEXTER (Discovering and EXtracting TERminology), an online multilingual workbench for data mining and terminology management with non-structured text-based corpora. The current version supports the processing of corpora in English, French, Italian and Spanish. One of its most distinctive features is a hybrid approach adopted not only to term extraction but also to term validation. DEXTER takes into account the linguistic and statistical properties of lexical units without POS tagging, thus minimizing language-dependent resources and recognizing domain-specific terms of different grammatical categories. It also allows automatic term recognition with the IATE thesaurus and also the manual validation of term candidates through a user-friendly interface. One of the possibilities of DEXTER is the construction of specialized glossaries. Because standard lexical resources take a snapshot of language in use at the time of compilation, the creation of corpus-driven specialized glossaries does not only provide sufficient quality and quantity of data but also relevant types of context related to the terms of a given field of expertise. In this regard, DEXTER can be a powerful tool to help language teachers discover specialized vocabulary for Language-for-Specific-Purposes courses from small- and medium-sized corpora.

**Paper #109.** Thursday, March 3. 12:30-13:00. ADUANA

**Anton Granvik.** "Variación y cambio sintáctico en el establecimiento de la construcción encapsuladora en el español del siglo XVI: presencia y ausencia de la preposición de e las oraciones completivas de sustantivo (N que vs. N de que)"

**Abstract:**

El objetivo de este trabajo es dar cuenta de una variación sintáctica que acaba en un conocido cambio, es decir, la introducción de la preposición de ante la conjunción que en las oraciones completivas de sustantivo: N que > N de que. El trabajo se basa en los datos del Corpus del español de (Davies 2002-) y del CORDE (RAE). Mediante un análisis colostruccionista (Stefanowitsch & Gries 2003) he identificado los 78 sustantivos encapsuladores más centrales del español, de los cuales nueve parecen ser de central importancia para la temprana expansión de la construcción N (de) que en el siglo XVI, a saber, causa, duda, esperanza, fe, opinión, recelo, señal, sospecha y temor. El trabajo se centra en los nueve sustantivos mencionados y su uso en las variantes N que y N de que, entre los años 1510 y 1690, época en la que la construcción experimenta un auge significativo de su frecuencia de uso.

**Paper #111.** Friday, March 4. 12:00-12:30. CALETA

**Olga Nevzorova, Alfiia Galieva and Dzhavdet Suleymanov.** "Building corpus-based semantic classifications of some Tatar affixes"

**Abstract:**

This study is aimed at exploring the semantic properties of Tatar –lik morpheme, in the framework of theoretical and computational linguistics. Turkic languages have complicated agglutinative morphology, which is a challenge for language processing. This study is devoted to the ambiguous Tatar –lik affix which may be joined to nominal, adjectival, verbal stems and form derivatives of different types depending on contextual environment. The answer to the question of the number of the types of derivatives and word forms produced with –lik affix is not trivial, and different researchers distinguish different types of derivatives. Based on a thorough analysis of Tatar derivatives containing -lik affix we identified some empirical features of these constructs and then performed their automatic classification. We determined main grammatical and semantic features and developed rules for classifying word forms with –lik affix and four classes were distinguished. For our experiments we used data from Tatar National Corpus (<http://corpus.antat.ru>).

**Paper #112.** Wednesday, March 2. 13:30-14:00. CALETA

**Joaquín Gris Roca, Raquel Criado Sánchez, Aquilino Sánchez Pérez and Agustín Romero Medina.** "The compilation and assessment of a corpus of activities for teaching EFL from a cognitive perspective"

**Abstract:**

The study of the explicit and implicit potential of activities used in the FL classroom can shed light on the type of knowledge achieved by learners. Our aim is to uncover how EFL teachers perceive their explicit and implicit potential. In our research, Corpus Linguistics techniques were employed to categorize 667 activities extracted from 16 low and upper-intermediate EFL textbooks. The categorization and labelling of the activities was carried out on the basis of their

explicit and implicit teaching potential: (i) fully or mostly explicit; (ii) fully or mostly implicit; (iii) partially explicit and implicit. Thirty activities (ten from each group) were randomly selected. Thirty EFL secondary school teachers assessed their explicit and implicit teaching potential by means of a scale designed for this purpose. Results seem to confirm the usefulness of Corpus Linguistics for compilations of this type and their usefulness for assessment in this issue.

**Paper #113.** Wednesday, March 2. 13:30-14:00 RETIRO

**Stavri Adams.** "The Type(s) of Vocabulary Used in Written and Oral Discourse in Health Education Settings"

**Abstract:**

The need for health professionals to share a common disciplinary lexis could justify the reason why lecturers, medical consultants and health education departments insist on trainee medical students to familiarise with the vocabulary of the discipline. This study sets out to investigate types of vocabulary used for medical communication in educational settings: (a) the general English, b) the technical and c) the sub-technical vocabulary of the discipline. This is based on the rationale that the acquisition of disciplinary lexis is an integral part of every health practitioner's training. This study relies on data from medical written materials and lectures; two corpora have been compiled to satisfy the purpose of this study: the Medical Lexis Reading Materials (MELREM) corpus and the Medical Lexis Oral Corpus (MELORC). The MELREM corpus consists of 2,097,627 running words while the MELORC corpus consists of transcribed oral lectures delivered to medical students which resulted in a 116, 872 running words corpus. The two corpora were analysed on the basis of the kind(s) of vocabulary used in them. More specifically, the proportion of technical and sub-technical vocabulary appearing in each of the two corpora was analysed and compared.

**Paper #114.** Thursday, March 3. 18:00-18:30. ADUANA

**Antonio Ruiz Tinoco.** "Using Geotagged Tweets in Spanish Linguistic Variation"

**Abstract:**

Geotagged tweets are an excellent source for the study of linguistic variation and used by many researchers. Large quantities of synchronic data can be easily obtained from the Streaming API of Twitter and stored in any type of database, exported to any format, processed and analyzed. In this presentation, I show a method to mine geotagged tweets and analyze some lexical (geosynonyms) and morphosyntactic variation observed in different Spanish-speaking areas. With the same tweet corpus, some phonological features of Spanish dialects, as the intervocalic /d/ deletion (for example *pescao* for *pescado*) can also be studied. Geotagged tweets can be easily shown in maps using standard GIS techniques, using points, polygons and heat maps. I also discuss some disadvantages and some workable solutions to avoid common problems mainly related to noise and ambiguity.

**Paper #115.** Wednesday, March 2. 13:00-13:30. ADUANA

**Mick O'Donnell.** "Automatic EFL Proficiency Assessment via detailed and deep feature extraction"

**Abstract:**

In recent years, much attention has been given to the automatic assessment of learner proficiency by computer analysis of their writing. One of the key problems with much of this work is in relation to the deficit of the input: without giving the detailed information as to the language features used by the learner, the assessment program cannot make accurate assessments. This paper discusses recent work by the author towards increasing the degree of detail and depth of the automatic syntactic tagging of EFL learner texts, not only in terms of more traditional areas of syntax (transitivity patterns, Mood, tense-aspect, Voice, etc.) but also in relation to more meaning-bearing syntactic structures: thematic structures, realisation of modality, quantificational strategies in the NP, nominalisation and clause combining strategies. I argue that only by providing such systems with richer information of this type can automatic assessment function adequately. Results using the annotated corpus to train an automatic assessment tool will also be given.

**Paper #116.** Thursday, March 3. 13:30-14:00. CALETA

**Ralph Rose.** "Sensitivity of pausing strategies to utterance length: Evidence from a crosslinguistic speech corpus"

**Abstract:**

When speakers encounter production difficulties, they have two primary nonverbal speech strategies open to them: a silent pause (SP) or a filled pause (FP: e.g., uh/um). What triggers the choice between these two is uncertain. One line of exploration concerns syntactic complexity, but what constitutes syntactic complexity also remains uncertain. Factors such as linear distance and clause embeddedness have been considered. The present research is part of a larger research program to investigate these questions and begins with exploration of a crosslinguistic corpus in order to identify high-frequency linguistic configurations that co-occur with SPs and FPs. Preliminary results show that while there is no relation between utterance length and SP rate in first language speech, speakers use fewer SPs when

utterances get longer in English. In contrast, there is no effect of utterance length on FP rate. The position of FPs within an utterance does appear important, however: In both first and second language, the longer the utterance, the later in the utterance the speaker is likely to use a FP. These preliminary results suggest that speakers' pausing strategies are sensitive to linear distance and that choice of pause type is modulated by length of utterance and position within utterance.

**Paper #117.** Wednesday, March 2. 15:30-16:00. RETIRO

**Ralph Rose and Hinako Masuda.** "Construction of a multimodal learner corpus of STEM student language production: A pilot study"

**Abstract:**

While the construction of specialized second language (L2) learner corpora is becoming commonplace (Granger et al 2015), there are few such corpora that include first language (L1) samples for baseline purposes, are focused on science, technology, engineering, and mathematics (STEM) students, and contain production data in multiple modalities. The present research is a pilot project to build such a corpus called the Science and Engineering Learner Corpus (SELCor). Language production was elicited from STEM students via several tasks: a L1 reading aloud task, six L2 speaking tasks varying from reading aloud to topic narrative, and one L2 essay-writing task. The speech was recorded and the writing task was performed on a computer which logged key-presses through a Java application created for this purpose. In this pilot project, 4.25 hours of speech (21,319 words) and 2,074 written words were collected from 15 participants. Results show a positive correlation between the amount of speech produced and the ratio of writing insertions to deletions. But there was a negative correlation between the amount of speech and the insertion-deletion ratio: In other words, participants who spoke less, also edited their writing less. Other results and implications for similar corpus efforts will be described.

**Paper #118.** Friday, March 4. 11:30-12:00. LIMONAR

**Yuanyi Liu and Antonio Moreno Sandoval.** "Asimetría en la traducción terminológica entre el español y el chino: el caso de "癌/cáncer""

**Abstract:**

When translating, a single and unambiguous term in one language can have several equivalents in another. The present paper, based on three corpora, aims to analyze a case of terminological asymmetry in medical domain between Spanish and Chinese and how to translate this phenomenon by using frequency data. We will choose two basic medical terms "癌/cáncer" as an example. While in Chinese, there is only a unique form (single word), in Spanish, we have several multiword options. For this paper, we will focus on two of these options: "cáncer+de+noun" and "cáncer+adjective". We will compare the occurrence and normalized frequencies of these multiword structures in three corpora and observe the preference of usage of each. Then, we will apply the empirical observations to different translation activities.

**Paper #119.** Friday, March 4. 11:00-11:30. CALETA

**Carlos Herrero-Zorita and Antonio Moreno-Sandoval.** "Sentence length and NP complexity of general and medical written academic and media texts. An analysis using a trained syntactic parser"

**Abstract:**

The main objective of this work is to perform a comparative analysis of sentence and main noun phrases complexity in two different types of discourses, written media and academic prose, using a trained syntactic parser (Stanford PCFG Parser). For this purpose, we have selected three written sources: a general media corpus, a medical media subcorpus and a medical academic prose subcorpus. From a total of more than 160000 sentences, we have carefully selected as the study sample a total of 300, which have been morphologically and syntactically annotated. Influenced by other studies related to syntax and statistics, our hypothesis is that NPs from academic prose and written media will contain four or more words, and those belonging to academic prose will be larger than the latter. The NPs studied are those that perform the main functions of the clause: subject, object (direct and indirect), attribute and time expressions. The results show a confirmation of our hypothesis. The academic subcorpus has the longest sentences and more complex NPs than the other texts. On the other hand, written media corpora achieve smaller NPs but their results are quite similar.

**Paper #120.** Thursday, March 3. 16:00-16:30. LIMONAR

**Mario Crespo Miguel and Marta Sanchez-Saus Laserna.** "Graded acceptance in corpus-based English-to-Spanish machine translation evaluation"

**Abstract:**

The main objective of this work is to perform a comparative analysis of sentence and main noun phrases complexity in two different types of discourses, written media and academic prose, using a trained syntactic parser (Stanford PCFG Parser). For this purpose, we have selected three written sources: a general media corpus, a medical media subcorpus and a medical academic prose subcorpus. From a total of more than 160000 sentences, we have carefully selected as the study sample a total of 300, which have been morphologically and syntactically annotated. Influenced by other studies related to syntax and statistics, our hypothesis is that NPs from academic prose and written media will contain four or more words, and those belonging to academic prose will be larger than the latter. The NPs studied are those that perform the main functions of the clause: subject, object (direct and indirect), attribute and time expressions. The results show a confirmation of our hypothesis. The academic subcorpus has the longest sentences and more complex NPs than the other texts. On the other hand, written media corpora achieve smaller NPs but their results are quite similar.

**Paper #121.** Thursday, March 3. 18:30-19:00. CALETA

**Barry Pennock-Speck and Begoña Clavel-Arroita.** "Multimodal Deference Strategies in a Corpus of Synchronous Telecollaboration Interactions"

**Abstract:**

In this talk we will describe a study carried out in the context of the European-funded project TILA (Telecollaboration for Intercultural Language Acquisition). We analyse a corpus of twelve recordings of tandem interactions between Spanish and British secondary school students aged aged between 14 and 15. We focus specifically on Goffman's (1956: 476) notion of deference, "a symbolic means by which appreciation is regularly conveyed to a recipient". There are two main types of deference: "avoidance rituals" and "presentation rituals" (Goffman, 1956: 481). Our results show that participants use body language and non-verbal cues along with verbal language in a systematic way to put their interlocutors at ease, defuse potentially embarrassing moments and also to offer them praise. This research provides empirical evidence of the valuable affordances of synchronous telecollaboration when compared with other kinds of online peer interaction, especially of a written nature.

**Paper #122.** Thursday, March 3. 12:30-13:00. LIMONAR

**Chu-Ching Hsu.** "A Corpus-Based Study of Lexical Variety and Language Variation in Literary Translation"

**Abstract:**

Adopted corpus methodology for analysis and framed of Lefevere's theoretical concept of translating as rewriting, this chronological and diachronic study investigates how the "colloquial idiom" and "dialect" in English literary works are rendered into Chinese language and how lexical variety and language variation are used by the different translators to convey the English colloquial language into Chinese written text for different target readers in different time spans, from 1970 to now. Thus, with these aims, an English-Chinese parallel corpus is compiled and built for this study and the collected corpora are composed of Mark Twain's English work Adventures of Huckleberry Finn, where there are several different dialogues written in colloquial idiom and dialect originally used in Mississippi, and its ten different Chinese renditions published in the past four decades, to carry out further analysis and to explore how individual translators impose their language ideology in their renditions.

**Paper #124.** Wednesday, March 2. 16:30-17:00. CALETA

**Kim Collewaert and An Vande Castele.** "A pilot study on the use of pronouns in oral productions of Spanish as a Second Language"

**Abstract:**

This communication aims at investigating the use of personal pronouns in oral discourse of Spanish language learners. The presentation focuses on the use of different kinds of personal pronouns and the pro-drop phenomenon, namely the existence of a null subject, typical of the Spanish language. The study is based upon a corpus of oral productions of second language learners of Spanish who all are native speakers of Dutch and also learned French and English, which means that for them the pro-drop phenomenon is new. We will investigate which kinds of pronouns are used in which syntactic contexts and indicate in what contexts the use of a pronoun is not required. Next to this, we see an unnecessary repetition of proper names and an over-use of personal pronouns as subjects. This is related to the concept of "over-explicitation" or "overspecification" (cf. Ahrenholz 2005; Arts et al 2010; Chini 2005, Muñoz 2001, Leclercq & Lenart 2013), whereby learners of a second language tend to use more explicit forms than necessary.

**Paper #125.** Friday, March 4. 11:00-11:30. RETIRO

**Alfiia Galieva, Bulat Khakimov and Airat Gatiatullin.** "On the Way to the Relevant Grammatical Tagset for Tatar National Corpus"



**Abstract:**

The development of the metalanguage for annotation is one of the topical issues in modern corpus linguistics. One of the main problems in the development of grammatical tagset for Tatar National Corpus was to identify the inventory level of inflectional categories and to create an optimal metalanguage of description. We discuss the factors that complicate the process of grammatical annotation for Turkic corpora in general. First of all, it is lack of a common metalanguage to describe grammatical categories of Turkic languages. Another problem is the poor differentiation of word-building and form-building in the Turkic languages. Syncretism of meanings in many grammatical categories, lack of «pure» categories in Turkic languages and homonymy of affixes also make challenge for us. Currently, the developed grammatical tagset contains 93 tags. Tags for parts of speech and grammatical categories were created to meet the worldwide standards, primarily the Leipzig glossing rules.

**Paper #127.** Thursday, March 3. 18:00-18:30. CALETA

**An Vande Castele and Kim Collewaert.** “Reflexiones metodológicas sobre el diseño y la compilación de u corpus oral de ELE”

**Abstract:**

As a corpus is a representation of the linguistic reality, it is important to have homogeneous, quantifiable and valid data. This presentation aims at discussing the issue of elaborating a corpus of oral data from Spanish language learners. We hereby do not merely focus on the data collection, but also on the difficulties that arise regarding the experimental design, the selection of the participants, the elaboration of a transcription model and the analysis of data. The discussion will be based upon our own research project, for which we collected oral samples from Spanish language learners of different proficiency levels, which will be analysed cross-sectionally. Furthermore, we will focus on the oral experiment specifically designed for this project, similar to those of Watorek (2004) and Leclercq y Lenart (2013). Next to this, we will also discuss the procedure used for the transcription of the data and finally, a codification system will be elaborated (cf. Baumann y Riester 2010, 2012, 2013).

**Paper #129.** Thursday, March 3. 12:30-13:00. RETIRO

**Ana Albalat Mascarell and Maria Luisa Carrió Pastor.** “The United States of America presidential election debates: Annotation and taxonomy of co-speech gestures in a multimodal corpus”

**Abstract:**

Focusing on the gestures that people produce in co-occurrence with speech during a communicative exchange, this paper provides a useful methodological approach to the annotation and analysis of gestural data using new technologies available for multimodal research. To this end a corpus made up of authentic data retrieved from video recordings of the United States presidential election debates was compiled, in which gestures are employed strategically by candidates in order to reinforce their verbal messages for rhetorical and persuasive purposes. Our research sheds light on different issues regarding the process of compilation and preparation of the corpus, including the use of a gesture annotation scheme that can be extrapolated to other dialogic contexts with similar communicative goals. Design considerations and other major results of this corpus-based approach will also be discussed, along with relevant limitations of this experience specifically grounded on both the human and financial resources available.

**Paper #130.** Thursday, March 3. 18:30-19:00. ADUANA

**Cristina Blanco-García.** “Looking into ephemeral and prototypical conditional subordinators in English: A diachronic account”

**Abstract:**

This paper aims at outlining the diachronic development of some ‘ephemeral’ conditional subordinators, among others *if so be* (that, as), *be it so*, *if case be* (that), *when if* and *conditioned* (that), from their origins in Middle English to their current status in Present-Day English. In order to provide a comprehensive view of the notion of ephemerality, the selected ephemeral connectives will be compared with the prototypical conditional subordinator *if*. The survey is based on data from the Penn-Helsinki Parsed Corpus of Early Modern English and the Penn-Helsinki Parsed Corpus of Modern British English, complemented with data from the Oxford English Dictionary, the Middle English Dictionary and their quotation databases. Additionally, genre-specific specialized corpora such as Middle English Medical Texts, the Corpus of Early English Medical Writing, the Corpus of Religious Prose and The Lampeter Corpus of Early Modern English Tracts will supplement the data from the base line corpora.

**Paper #131.** Thursday, March 3. 18:30-19:00. LIMONAR

**Violeta Kaledaite.** “Existence in a parallel corpus: what a corpus-based analysis reveals about English and Lithuanian”

**Abstract:**

Translations in contrastive studies provide a unique insight into the relationship that exists between different languages. The paper presents the contrastive analysis results of the existential sentence type in English and Lithuanian based on the bidirectional parallel corpus (<http://tekstynas.vdu.lt/>). The size of the corpus is 2,084,159 words. The specific research questions address the frequency and position of the locative element in original English and Lithuanian ESs and their translations into Lithuanian/ English. Next, rendering of the English existential verb to be into Lithuanian is discussed. It is a characteristic feature of Lithuanian ESs to incorporate a verb with its full lexical meaning more often than the verb 'būti' (be). Finally, non-equivalent translations are considered to find out what other constructions are chosen by the translators to relay the meaning of an English ES/ Lithuanian ES. The findings point to constructional and pragmatic differences in expressing existence in the two languages.

**Paper #132.** Thursday, March 3. 12:30-13:00. CALETA

**Awatif Alruwaili.** "Investigating the use of corpus-based approaches for the teaching general verbs in the language classroom"

**Abstract:**

Corpus-based learning has been the subject of increased interest in recent years, in terms of its application in language education to facilitate learners' language acquisition (Bernardini, 2004). However, despite developments in corpus linguistics (CL) and the potential for applying corpus-based learning in language pedagogy, Römer (2006:121) declares that it still has not 'fully arrived'. Moreover, regardless of the wide recognition of corpus benefits, there is still, as she points out, 'hesitation' about using corpus approaches in teaching. Although there is growing interest in the use of corpora in the classroom, relatively little attention has been paid to investigating learners' actual use of corpora in learning collocations. The main aim of this study is to examine the effectiveness of a corpus approach in the classroom through a case study of general verbs (Gvs).

**Paper #133.** Wednesday, March 2. 16:00-16:30. ADUANA

**Lisa Medrouk, Anna Pappa and Jugurtha Hallou.** "Review web pages collector tool for thematic corpus creation"

**Abstract:**

This paper presents an automatic system RevScrap (Reviews Scrapping ) designed for creating large thematic Corpus reviews from multiple Internet resources. We propose a method to extract all the opinions, comments and reviews related to a specific set of seed terms (product, name, service, town, ...) as input in different languages (tests made in French and English) from multiple domains. Our approach uses a search engine (google) and a specific DOM tree algorithm named DOMScrapRev for the detection and the identification of the comments nodes regardless the heterogeneity of the page structure and only keeping the most usefull texts (no redundancy, 1500 car par reviews). The technique is based on the top down tree-parsing algorithm. The results are xml documents named after the request's terms. Our system is aimed to provide available dataset for opinion mining and sentiments analysis.

**Paper #134.** Thursday, March 3. 12:00-12:30. LIMONAR

**Joaquín Primo-Pacheco.** "A Preliminary Corpus Linguistics Approach to Two Gay Television Series"

**Abstract:**

The purpose of this pilot study is to deploy a corpus linguistics methodology drawing on Baker (2004, 2006) in order to carry out a preliminary approach to the spoken language of two gay-themed television series from the United States of America: "Queer as Folk" (Showtime, 2000-2005) and "Looking" (2014-2015). As Baker (2004: 346) puts it, "[c]orpus-based analysis allows researchers to identify more or less objectively widespread patterns of naturally occurring language and rare instances, both of which may be overlooked in a small-scale analysis." Therefore, the ultimate aim of this initial approach is to uncover and identify by means of AntConc (Anthony 2014) any possible linguistic patterns, similarities or dissimilarities occurring between both television series, given the time span of almost ten years between them. Hopefully, a preliminary quantitative corpus linguistics approach shall point out future research directions to be taken from a qualitative discourse perspective, so as to ultimately determine whether or not there has been an evolution between one series and the other as regards use of language, and if any, if such evolution might be said to be positive or negative towards the gay community in terms of representation.

**Paper #135.** Wednesday, March 2. 16:00-16:30. LIMONAR

**Luis Conejero-Magro.** "From Korea to Hawaii: Dystopian Discourse Through Corpus Linguistics"

**Abstract:**

It goes without saying just how complex and difficult it is to render into a foreign language the linguistically marked speech of David Mitchell's *Cloud Atlas* (2004), especially the artificially primitive lexicon used by the Valleysmen. Evidence of this is the ups and downs patent in the version of this lexicon in the Spanish translation of this novel. The

importance of the function of this specific vocabulary is such that the translator's flaw in the recreation of some key terms partially impairs a fuller understanding of Mitchell's dystopia. This is why any improvement, however minimal, in this respect, would be highly beneficial for the Spanish reception of the entire text. Only through a thorough contrastive analysis between Mitchell's original text and that of the Spanish translator's, by means of corpus-based analytical techniques, will we be able to detect the degree of deviation of the Spanish terms from the original ones.

**Paper #136.** Wednesday, March 2. 16:30-17:00. LIMONAR

**Khadija Assulaimi.** "The representation of Saudi Arabia in the American press: A corpus-based critical discourse analysis."

**Abstract:**

By combining tools from Corpus linguistics and Critical Discourse Analysis this study aims to explore and analyse news texts about Saudi Arabia in the American press. It further aims at examining which discursive strategies and linguistic structures are instrumental in constructing the image of Saudi Arabia. The texts are collected from two American newspapers of public record which have a large circulation: the New York Times and the Washington Post. The corpus comprises articles published in three different periods; 1990-1991, 2001-2002, and 2010-2011. This study draws on methodologies from Corpus Linguistics and Critical Discourse Analysis. Moreover, the analysis utilizes the concept of Transitivity, drawn from the ideational function of Halliday's (1994) Systemic Functional Grammar. Furthermore, it employs analytical methods from the discourse historical approach (DHA) (Wodak & Reisigl, 2001). The relevant discursive strategies of the current study include referential/ nomination strategies and predication strategies.

**Paper #138.** Thursday, March 3. 13:00-13:30. RETIRO

**M. Pilar Valverde Ibañez.** "A blog corpus of learner Spanish with automatic article error annotation"

**Abstract:**

Articles are the most frequent word in Spanish, and also one of the most frequent grammatical errors for learners, especially for speakers of languages that do not have articles, like Japanese. Based on our previous study on article error annotation, we developed a set of Constraint Grammar rules to automatically detect and correct a subset of article errors in learner Spanish texts. One of the guiding principles of the system is to favor precision over coverage, so we focus on article usage determined mainly by syntactic, lexical and semantic factors, for which there are some linguistic norms with an acceptable degree of reliability. Specifically, we focus on the language produced by learners of Spanish whose mother tongue is Japanese. Since there is no available learner corpus for this language pair, we constructed a corpus made up of personal web blogs (2,126 texts and 634,767 words). One of the distinctive characteristics of this User-Generated Content corpus is that learners frequently use words or short sentences in their mother tongue while they are writing a text in Spanish. This involves not only a mixture of languages but also a combination of different character types (latin alphabet, hiragana, katakana and kanji).

**Paper #139.** Wednesday, March 2. 12:30-13:00. CALETA

**Alicia Ricart.** " "Patients at risk?", "Suffering from diabetes?" A corpus-driven approach to identify grammatical errors"

**Abstract:**

The present study investigates the misuse of prepositions on ESP students of English when producing written material. It was conducted on purpose of identifying the errors committed by Nursing students, who enrolled in the core subject English (2016) at the University of Valencia. To this end, students were grouped in twos and asked to look for three current articles from English or American on-line newspapers, dating between 2013 and the present time. The articles had to deal with a specific health issue of considerable international relevance. Students had to write two compositions on two different specific medical topics suggested by the teacher. A total 90 essays accounted for the corpus, which evidenced that prepositions were among the most recurrent errors committed by the subjects. The online tool WordSmith Tools 5, allowed me to do a qualitative and quantitative analysis of the most frequently misused prepositions: at, in and on. In this way, the frequency and concordance of the prepositions were analyzed to identify incorrect combinations. As a conclusion, it can be stated that corpus analysis is effective for the identification and avoidance of grammatical errors and it can be also used as the basis for the creation of didactic materials.

**Paper #140.** Wednesday, March 2. 15:30-16:00. ADUANA

**Ivan Garrido-Márquez, Laurent Audibert, Jorge García-Flores, François Lèvy and Adeline Nazarenko.** "A French weblog corpus for new insights on blog post tagging"

**Abstract:**

Blogging is a common source of information nowadays containing, news, knowledge, opinions, and communication between writers, and readers. We present a corpus of labeled posts of blogs in French. The posts were directly labeled and classified by the authors with keywords and categories. This corpus was originally gathered to develop and evaluate automatic keyword tag prediction, and classification methods for blog posts.

**Paper #142.** Wednesday, March 2. 13:30-14:00. ADUANA

**Mehdy Dref and Anna Pappa.** "An interaction approach between services for extracting relevant data from Tweets corpora"

**Abstract:**

We present a system based on the need of special infrastructure adequate to software agents to operate, to compose and make sense from the contents of the Web resources through the development of a multi-agent system oriented services interactions.

Our method follows the different construction ontology techniques and updates them by extracting new terms and integrate them to the ontology. It is based on the detection phrases via the ontological database DBPedia. The system treats each syntagme extracted from the corpus of messages and verifies whether it is possible to associate them directly to a DBPedia knowledge. In case of failure, these service agents interact with each other in order to find the best possible answer to the problem, by operating directly in the phrase, trying to semantically modify it, until the association with ontological knowledge becomes possible. The advantage of our approach is its modularity : it is both possible to add / modify / delete a service or define a new one, and then influence the outcome product. We could compare the results extracted from a heterogeneous body of messages from the Twitter social network with Tagme method, based mainly on storage and annotation of encyclopaedic corpus.

**Paper #143.** Thursday, March 3. 19:00-19:30. CALETA

**M<sup>re</sup> Isabel Velasco Moreno.** "Do all FL classes pursue the same aims?"

**Abstract:**

Last five decades have witnessed a great interest concerning Foreign Language Acquisition. Although there is a lot of research on Second Language Classrooms (Chaudron, 1988, Cameron, 2003), better ways to learn a foreign language (Ellis and Sinclair, 1989) or learners' uptake (Bailey, 1991), we have checked that most of the studies focus on teacher's performance. Contrary to that, this research pretends to investigate the teaching-learning context itself, considering all members' participation. We have investigated discourse generated in class applying a communicative discourse analysis model based on Sinclair & Coulthard (1975), Tsui (1994) and Poyatos, (1994). It is our main aim to analyze English classes for Spanish young learners in different educative contexts to find out how methodology implemented can affect students' oral output.

**Paper #144.** Friday, March 4. 12:30-13:00. CALETA

**Bulat Khakimov, Ramil Gataullin and Rinat Gilmullin.** "Grammatical Disambiguation in the Tatar National Corpus"

**Abstract:**

This article concerns the issues of grammatical ambiguity in the Tatar National Corpus and the possibilities for automation of the disambiguation process in the corpus. Grammatical ambiguity is widely represented in agglutinative languages like Turkic or Finno-Ugric. In order to build the grammatically disambiguated subcorpus, we have developed a special software module which searches for ambiguous tokens in the corpus, collects statistical information and allows to create and implement the formal disambiguation rules for different ambiguity types. Disambiguation in the corpus is based on the context-oriented classification of ambiguity types which has been carried out on statistical corpus data in the Tatar language for the first time. We can say that we use the corpus as a source of our research and at the same time as a destination for implementing the results. Estimated cumulative effect of disambiguation of the identified frequent ambiguity types in the Tatar National Corpus can be up to 50%.

**Paper #147.** Friday, March 4. 13:00-13:30. ADUANA

**Giuli Shavashvili and Nana Shavtvaladze.** "Corpus-based Approach in Teaching Vocabulary – Data of Georgian Language"

**Abstract:**

The paper aims to underline the importance and topicality of using corpora in language teaching. Besides, the main goal is to show the ways of teaching Georgian vocabulary by corpora considering the specifics and difficulties of Georgian Language. The given research is based on the both formal and semantic approaches. The three different types of Georgian stems are separated: nouns, adjectives and infinitives. Seven different forms with different mainly used and most productive affixes for each stem are discussed in the paper. Besides based on given approach the learner can be able to produce the different parts of speech with these affixes. Explanation and description of each for the student can find out in the corpora. Particular examples and main results are introduced in the research as well.

**Paper #148.** Wednesday, 17:00-17:30. CALETA

**Yuka Ishikawa.** "Frequency and Learning Difficulty in Engineering English Vocabulary: A Corpus-Based Study"

**Abstract:**

Teaching ESP vocabulary is considered essential for ESP program, therefore various wordlists for university students majoring in engineering have been created up to now. These lists are highly reliable for they are created based on corpus analyses. However, using these lists in a classroom often causes certain kind of problems: top frequent words are often too easy for learners but removing frequent words from the wordlist can sometimes bring damage to vocabulary learning: some frequent words can be difficult to learn, while some infrequent words are easy to understand for learners. This research aims to identify the frequent words which learners tend to fail to learn and also identify the infrequent words which learners tend to acquire more quickly. The purpose is to explore the relationship between the frequency and difficulty of engineering English vocabulary and reorder the wordlist for engineering students (Ishikawa, 2011). In this study, we will set these research questions. 1. Which frequent words on the wordlist tend to be more difficult to learn? 2. Which infrequent words on the wordlist tend to be less difficult to learn? 3. Which words on the wordlist tend to have the difficulty coincident with the frequency?

**Paper #149.** Wednesday, March 2. 16:30-17:00. RETIRO

**Irene Doval Reixa.** "Bilingual parallel corpora for linguistic research"

**Abstract:**

This paper discusses the specific needs of linguistic research regarding the bilingual parallel corpora and primarily the conclusions to be drawn for design, compilation and domains covered by a corpus. Initially a brief survey of some significant existing related corpora is performed, and their limitations for linguistic studies are outlined. The different issues that were taken into account for the design of the corpus will be explained. I will explain the criteria used to select the texts such as domain, time of the text production, regional language variety as well as functional variety. After describing the manual preparation process of the texts for the alignment we will explain the manual and automatic annotation procedure: the non linguistic annotations, the metadata, we provide the texts with and the automatically linguistic annotation (lemmatization and POS tagging) that is very useful for any linguistic research. Lastly I will explain what issues need to be dealt with due to copyright restrictions and their consequences for the availability of the corpus.

**Paper #150.** Wednesday, March 2. 12:30-13:00. ADUANA

**Daniel Couto Vale and Adjan Hansen-Ampah.** "The meaning of hashtags matters: detecting hashtags such as #JeSuisKouachi"

**Abstract:**

Right after the Paris attacks in January 2015, a wave of Twitter hashtags such as #JeSuisCharlie showing solidarity with the victims became popular. In parallel to that, there were warning signs of hashtags such as #JeSuisKouachi (the Kouachi brothers were two of the terrorists) potentially showing support for terrorism. Even though these hashtags have similar structures, they have completely different meanings. This poses a challenge for NLP. For this paper, we modelled the contexts of situation and discourse as an information state so as to restrict the meaning potential of Twitter hashtags.

By doing so and by using a new parser that does not count on word separators (<https://github.com/DanielCoutoVale/openccg>), we applied restrictions through compositional semantics to tell the potential meanings of a hashtag given the current situation and discourse.

**Paper #151.** Wednesday, March 2. 12:00-12:30. LIMONAR

**Sergio Maruenda-Bataller.** "Corpus Linguistics and Lexical Pragmatics: A view from the representation of intimate partner violence discourses in written media"

**Abstract:**

Our aim in this paper is to incorporate a discursive perspective to the existing work on the combination of CL and Lexical Pragmatics. Specifically, we present a more flexible account of meaning negotiation which lays emphasis on the multiplicity and instability of sensitive socio-ideological discourses and conceptual representations, taking on board political, religious or ideological issues. Our analysis centres on the conceptual representations built around intimate partner violence in an ad hoc corpus of written media from 2005 to 2014, by exploring the (linguistic) means by which social actors discursively represent the participants, the events and state-of-affairs from the viewpoint of attitudinal meaning (Martin & White 2005: 35, 42). Our aim is, therefore, two-fold: (i) to provide an account of the conceptualization and representation of intimate partner violence in contemporary media discourse over the last few years; thus unveiling expectations about gender, sexuality and power implicit in public discourses on gender violence; and (ii) to attempt to provide a social definition of the concepts analysed in the terms of these discursive struggles, far

beyond traditional lexicography. The latter is part of an under-way project of a socio-ideological dictionary of gender/sexual inequalities.

**Paper #152.** Wednesday, March 2. 12:00-12:30. ADUANA

**María Araceli Losey León.** "Multi-word units' conceptual aspects of representation in a genre text type of a Maritime English specialized domain"

**Abstract:**

This paper addresses computer aided corpus-based terminology work for term extraction (TE) to investigate the conceptual information of variants in a genre text type in the maritime domain. This work is structured as follows. In the first section, we shall be dealing with data preprocessing through corpus segmentation according to topic classification criteria of different variants in the same domain specific genre text type. The second section of our study presents the preliminary list of candidate multi-word terms that will lead us to a contrastive analysis where findings will reveal an intricate web of conceptual links among the genre topic variants involved in this specialized domain.

**Paper #154.** Thursday, March 3. 16:00-16:30. CALETA

**Mónica Bautista and Cristóbal Lozano.** "From experiments to corpus-based evidence: Anaphora resolution in L1 English-L2 Spanish in the CEDEL2 corpus"

**Abstract:**

Previous experimental studies on Anaphora Resolution (AR) in native and non-native Spanish have extensively studied the Position of Antecedent Strategy (PAS), a structural strategy in which the syntactic function of the antecedent determines the form of the anaphor (overt/null pronominal subjects). In this presentation we investigate PAS in an L1 English-L2 Spanish corpus (CEDEL2), as corpora offer more natural and contextually richer data than experiments. A sample of lower- and upper-advanced learners (plus a native control subcorpus) were delicately annotated with a fine-grained tagset. Results reveal that both lower- and upper-advanced learners behave similarly to the natives in terms of the PAS, preferring a null pronoun to co-refer to a subject antecedent. But, importantly, results reveal that overt pronouns do not show a clear antecedent bias, but NPs show a stronger bias towards antecedents in non-subject position, a finding which has been overlooked in previous experimental research.

**Paper #155.** Wednesday, March 2. 15:30-16:00. CALETA

**Francisco Pineda.** "Retrieving communicative functions from the BNC through linguistic patterns: making suggestions"

**Abstract:**

In this study I seek to demonstrate that a corpus-based approach to the study of specific aspects of Second Language Acquisition (SLA) could eventually help reduce the form-function gap. I will try to shed light on the strengths and weaknesses of a methodology, Communicative Language Teaching (CLT), officially adopted by the Spanish Education authorities in 1984. Qualitative and quantitative research designs were adopted to try to demonstrate that structures which are traditionally studied from a formal perspective, can become the starting point to introducing the most common communicative functions connected to them. To accomplish my goals, I will focus on the analysis of more than one thousand examples of the structure *why+don't+you?*, retrieved from the British National Corpus (BNC). The major objective is to quantify the percentage of occurrences in which this structure is used as a suggestion.

**Paper #157.** Wednesday, March 2. 18:00-18:30. ADUANA

**Antonio Moreno Ortiz.** "Extracting Domain-Specific Features for Sentiment Analysis Using Simple NLP Techniques: running shoes reviews"

**Abstract:**

In this paper we define a methodology for extracting product features found in product reviews that uses unsophisticated Natural Language Processing and Corpus Linguistics techniques, such as NP chunking, and regular expressions. A domain-specific corpus consisting of 650 running shoes reviews, totaling about 1.1M words, was compiled and annotated in XML. We focused on noun phrases to extract relevant features. A simple Python-NLTK chunker was created which extracted NPs with reasonable precision and recall. Then, manual qualitative exploration was carried out to identify NPs that did not seem to offer relevant information. After filtering the full list of NPs (over 245,000), the remaining NPs were searched for frequent co-occurring combinations of lexical items and syntactic patterns. This methodology allowed us to extract not only a comprehensive list of product features that are commonly discussed in running shoes reviews, but also domain-specific multi-word expressions.

**Paper #159.** Wednesday, March 2. 18:30-19:00

**María Teresa Cáceres-Lozano and Sara Hernández Arroyo.** "Diseño de un corpus diatópico de nahuatlismos en el español americano"

**Abstract:**

El examen del léxico regional del español americano actual es una de las vías de conocimiento para analizar la continuidad de las voces amerindias en América. Los datos de los estudios extraídos de textos coloniales del s. XVI manifiestan que los prehispanismos más numerosos son las voces nahua en las tres grandes áreas coloniales: Caribe; Mesoamérica y Sudamérica. El objetivo de esta investigación es elaborar un nuevo corpus diatópico y regional que permitirá entender cómo se difundió el léxico prehispánico en la lengua española. A través del análisis cuantitativo se comprueba que más de 1.300 nahuatlismos actuales se difunden de manera distintas en las referidas zonas americanas. Esta conclusión corrobora en parte lo que se indicaba para los siglos XVI y XVII, por lo que esto abre nuevas investigaciones, de las que se obtengan nuevos ejemplos para este corpus diatópico y dialectal. Los resultados de esta indagación quieren ser una contribución a los aspectos diatópicos de las voces nahuas que funcionaron como préstamos en la configuración del español americano.

**Paper #160.** Thursday, March 3. 19:00-19:30. RETIRO

**Reglindis De Ridder.** "Corpora as reference works for the study of lexical geographical markedness"

**Abstract:**

This case study explores the use of corpora as reference works to study geographical markedness, more specifically, geographically marked Belgian Dutch ('Flemish') and Netherlandic ('Holland') Dutch lexis. With the help of Dutch corpora comprising texts originating from both countries, a systematic lexical analysis was carried out focussing on the national varieties of this pluricentric language, i.e. Belgian and Netherlandic Dutch. Since Netherlandic Dutch words and expressions were until recently not labelled in the dictionaries, alternative reference works had to be found to establish the geographical markedness of a given word or phrase. To this end, the 500-Million-Word Reference Corpus of Contemporary written Dutch (SoNaR) and a purpose-built specialised corpus were used.

## INDEX OF AUTHORS

Adams, Stavri.	#113	Clavel-Arroita, Begoña.	#121
Akinlotan, Mayowa.	#76	Collewaert, Kim.	#124, #127
Albalat Mascarell, Ana.	#129	Colwell, Veronica.	#82, #84
Alcantud-Díaz, María.	#72	Conde-Silvestre, Juan Camilo.	#30
Alruwaili, Awatif.	#132	Conejero-Magro, Luis.	#135
Assulaimi, Khadija.	#136	Couto Vale, Daniel.	#150
Audibert, Laurent.	#140	Crespo, Begoña.	#14
Báez Montero, Inmaculada C.	#73	Crespo Miguel, Mario.	#120
Barrios Sabador, M <sup>a</sup> José.	#52	Criado Peña, Miriam.	#36
Bautista, Mónica.	#154	Criado Sánchez, Raquel.	#112
Bergh, Gunnar.	#48	Dahunsi, Toyese Najeem.	#38, #43
Bikeliene, Lina.	#89	De Ridder, Reglindis.	#160
Blanco-García, Cristina.	#33, #130	Díaz-Negrillo, Ana.	#102
Botella, Ana.	#27	Dima, Olga.	#70
Britain, David.	#54	Donnellan, Mark.	#51
Bürki, Dominique.	#54	Doval Reixa, Irene.	#149
Cáceres-Lozano, M <sup>a</sup> Teresa.	#159	Dref, Mehdy.	#142
Calderón López, M <sup>a</sup> Isabel.	#25	Elwood, James.	#77
Calle, Javier.	#19, #64	Fernández-Domínguez, Jesús.	#90
Callies, Marcus.	#39	Fernández de Molina Ortés, Elena.	#48
Candela-Mora, Miguel Ángel.	#75	Fernández-Nistal, Purificación.	#80, #83
Cao, Feng.	#1	Fernández Soneira, Ana M <sup>a</sup> .	#73
Carmona-Cejudo, Jessica.	#47	Ferri-Miralles, Imma.	#27
Carretero, Marta.	#22	Freijeiro Ocampo, Eva.	#73
Carrió-Pastor, M <sup>a</sup> Luisa.	#44, #129	Fuertes Olivera, Pedro A.	#41
Castaño, Emilia.	#85	Fuster-Márquez, Miguel.	#18



Charles, Maggie.	#91	Fuyuno, Miharū.	#100
Gago Jover, Francisco.	#99	Ishikawa, Shin'Ichiro.	#101
Galieva, Alfiia.	#111, #125	Ishikawa, Yuka.	#148
Gandón-Chapela, Evelyn.	#95	Izquierdo, Marlén.	#79
García-Flores, Jorge.	#140	Jiménez García, Elena.	#80
García-Marchena, Oscar.	#13	Kadooka, Kenichi.	#24
García-Vidal, Tamara.	#30, #66	Kaledaite, Violeta.	#131
Garrido-Márquez, Iván.	#140	Kapranov, Oleksandr.	#1
Gataullin, Ramil.	#144	Katsina, Tatiana.	#70
Gatiatullin, Airat.	#125	Kavvadia, Maria.	#70
Gee, Roger.	#61	Kayaoglu, Mustafa Naci.	#103
Gilmullin, Rinat.	#144	Khakimov, Bulat.	#125, #144
Goethals, Patrick.	#40	Kitrou, Sofia-Nefeli.	#70
Gordo Gómez, Pablo.	#41	Koutsoubou, Marina.	#70
Granath, Solveig.	#28	Kostakou, Xristina.	#70
Granvik, Anton.	#109	Kruzhkov, Mikhail.	#65
Gregori-Signes, Carmen.	#18, #60	Labrador, Belén.	#37
Gris Roca, Joaquín.	#112	Laso, Natalia Judith.	#85
Hallou, Jugurtha.	#133	Lavid, Julia.	#22
Hansen-Ampah, Adjan.	#150	Lee, John.	#45
Hennemann, Anja.	#7	Leonhardt, Tobias.	#54
Hernández-Arroyo, Sara.	#159	Lèvy, François.	#140
Hernández Bayter, Henry.	#17	Liu, Yuanyi.	#118
Hernández-Campoy, Juan Manuel.	#30	Louw, William.	#107
Hernández Lorenzo, Laura.	#53	Losey León, M <sup>a</sup> Araceli.	#152
Herrero-Zorita, Carlos.	#119	Lozano, Cristóbal.	#102, #154
Hu, Guangwei.	#1	Lynch, Sarah.	#54
Hsu, Chu-Ching.	#122	Maruenda-Bataller, Sergio.	#151

Iakovou, Maria.	#70	Maskowich, Isabel.	#14
Masuda, Hinako.	#117	Periñán-Pascual, Carlos.	#10
Matamala, Anna.	#15	Perrea, Stavrialena.	#70
Matsumoto, Kazuko.	#54	Petchko, Katerina.	#77
Mendrouk, Lisa.	#133	Pineda, Francisco.	#155
Merino Ferradá, M <sup>a</sup> del Carmen.	#25	Puente Castelo, Luis Miguel.	#68
Mestre-Mestre, Eva M.	#106	Primo-Pacheco, Joaquín.	#134
Mettler, Laura.	#54	Rabadan, Rosa.	#84
Milojkovic, Marija.	#107, #137	Ramón, Noelia.	#37
Mónaco, Leida Maria.	#68	Resch, Claudia.	#69
Moreno Ortiz, Antonio.	#157	Ricart, Alicia.	#139
Moreno-Sandoval, Antonio.	#118, #119	de los Ríos Rodicio, Ángel.	#41
Moya, Patricio.	#44	Rodríguez Martín, Gustavo Adolfo.	#50
Nakamata, Naoki.	#16	Rodríguez-Puente, Paula.	#32, #33
Nakajima, Yoshitaka.	#100	Romero-Barranco, Jesús.	#19, #63
Nazarenko, Adeline.	#140	Romero Medina, Agustín.	#112
Nevzorova, Olga.	#111	Rose, Ralph.	#116, #117
Nieto Caballero, Guadalupe.	#56, #57	Ruano, Pablo.	#56
Niño Amo, Marta.	#41	Ruiz Tinoco, Antonio.	#114
Noya Gallardo, Carmen.	#25	Saitoh, Takeshi.	#100
O'Donnell, Mick.	#115	Salles Bernal, Soluna.	#63, #87
O'Neill, Gavin.	#77	Sánchez Pérez, Aquilino.	#112
Ortego Antón, M <sup>a</sup> Teresa.	#80, #81, #83	Sánchez-Saus Laserna, Marta.	#120
Ozbay, Ali Sukru.	#103, #105	Sanjurjo, Hugo.	#82, #84
Pappa, Anna.	#133, #142	Sastre Ruano, Ángeles.	#41
Pappa, Froso.	#70	Shabashvili, Giuli.	#147
Patin, Stephane.	#21	Shavtvaladze, Nana.	#147

<b>Peenock-Speck, Barry.</b>	<b>#121</b>	<b>Skorzynska, Hanna.</b>	<b>#55</b>
<b>Pérez Blanco, María.</b>	<b>#79</b>	<b>Stuart, Keith.</b>	<b>#27</b>
<b>Suau Jiménez, Francisca.</b>	<b>#35</b>		
<b>Suleymanov, Dzhavdet.</b>	<b>#111</b>		
<b>Tamaredo, Iván.</b>	<b>#33, #104</b>		
<b>Torigoe, Shintaro.</b>	<b>#42</b>		
<b>Torrest-Hostench, Olga.</b>	<b>#75</b>		
<b>Tsutahara, Ryo.</b>	<b>#62</b>		
<b>Valverde Ibañez, M<sup>a</sup> Pilar.</b>	<b>#138</b>		
<b>Vande Castele, An.</b>	<b>#124, #127</b>		
<b>Vasiliadi-Linardaki, Irianna.</b>	<b>#70</b>		
<b>Velasco Moreno, M<sup>a</sup> Isabel.</b>	<b>#143</b>		
<b>Velasco Sacristán, Marisol.</b>	<b>#41</b>		
<b>Verdaguer, Isabel.</b>	<b>#85</b>		
<b>Villegas, Marta.</b>	<b>#15</b>		
<b>Vlachou, Flora.</b>	<b>#70</b>		
<b>Werner, Martina.</b>	<b>#69</b>		
<b>Williams-Camus, Julia.</b>	<b>#32</b>		
<b>Yamashita, Yuko.</b>	<b>#100</b>		
<b>Zamorano, Juan Rafael.</b>	<b>#22</b>		
<b>Zapata-Barrero, Belén.</b>	<b>#30</b>		
<b>Zarco-Tejada, María Ángeles.</b>	<b>#25</b>		
<b>Zhao, Qiurong.</b>	<b>#88</b>		