



Copyright: III Congreso Internacional de Lingüística de Corpus  
CILC 2011.

Departamento de Lingüística Aplicada. Edificio 4P

Avinguda Tarongers, s/n 46022 Valencia

Tel. +963877530

Fax. +96377539

E-mail: [cilc2011@upvnet.upv.es](mailto:cilc2011@upvnet.upv.es)

Web: [www.cilc2011.upv.es](http://www.cilc2011.upv.es)

## Contenidos

Bienvvenida / Welcome .....	3
Comité organizador / Organizing Committee .....	5
Paneles y directores / Panels and directors .....	6
Junta directiva de AELINCO / AELINCO Executive Board .....	7
Agradecimientos / Acknowledgment .....	8
Planos de situación / How to get to the congress.....	9
Situación de las aulas y salas de presentación / Room Plan .....	12
Información para los congresistas / Information for congress attendees ....	13
Programa del congreso / Conference Programme.....	15
Sesiones paralelas / Parallel sessions.....	21
Conferencias plenarias / Plenary sessions.....	39
Resúmenes de ponencias / Abstracts.....	41

## BIENVENIDA CILC 2011- VALENCIA

El Comité Organizador del III Congreso Internacional de la Asociación Española de Lingüística de Corpus (AELINCO) les da la bienvenida y agradece tanto a los ponentes como a los asistentes su participación en este encuentro anual de la Asociación. La Universidad Politécnica de Valencia, así como el departamento de Lingüística Aplicada han acogido este congreso con entusiasmo, recibiendo apoyo y ánimo para esta iniciativa.

Estamos muy satisfechos por la gran cantidad de propuestas recibidas tanto nacionales como internacionales, lo cual evidencia la importancia de este tipo de debates anuales sobre aspectos específicos de la Lingüística. En línea con el objetivo específico de AELINCO y de los congresos anteriores, la tercera edición del Congreso Internacional de Lingüística de Corpus se centra en la difusión de investigaciones desarrolladas en el marco de la Lingüística de Corpus y da cabida a estudios sobre distintos aspectos y aplicaciones del lenguaje natural o las lenguas particulares basados en el análisis de córpora mediante las herramientas ofrecidas por las tecnologías de la información y de las comunicaciones (TICs).

Esperamos que esta edición del congreso de la Asociación sea del agrado de todos los participantes, que disfruten del intercambio de investigaciones y proyectos que se presentan en los nueve paneles temáticos del congreso, así como de las ponencias plenarias.

Por supuesto, todo nuestro esfuerzo no obtendría su fruto sin la valiosa ayuda de todos aquellos que han participado en la organización del congreso y de las entidades financiadoras de este evento.

Benvinguts i benvingudes a València!

Esperamos que disfrutéis de vuestra estancia en Valencia.

EL COMITÉ ORGANIZADOR

CILC 2011

## WELCOME CILC 2011- VALENCIA

The organizing committee of the III International Congress of the Spanish Association of Corpus Linguistics (AELINCO) wishes you a warm welcome to Valencia and would like to thank all the speakers and attendees who are taking part in this annual meeting of the Association. Both the Universidad Politécnica de Valencia, and the Department of Applied Linguistics were enthusiastic about holding the congress here, and we are grateful for the support and encouragement given in order to bring this about.

We are delighted with the number of proposals submitted from within Spain itself, and from all over the world, which shows how relevant these annual events are in order to promote discussion and to reflect on specific aspects of studies in Linguistics. In line with the specific aims of AELINCO and previous Conferences, the third edition of the International Conference on Corpus Linguistics focuses on the dissemination of research conducted within the framework of Corpus Linguistics, including different aspects of natural language processing and corpus analysis using the different tools which have been developed in the field of Information and Communication Technologies (ICTs) for the study of specific languages and genre.

We sincerely hope that the present edition of the AELINCO congress will be a success, and that the participants enjoy having the opportunity to exchange ideas and inform each other about different research projects in the nine thematic panels and the plenary sessions.

Lastly, we would like to thank all those who have participated in the organization of the congress and the different sponsors, without whose help and finance the event would not have been possible.

Benvinguts i Benvingudes a València!

Welcome, and enjoy your stay in Valencia!

The Organizing Committee

CILC 2011

**COMITÉ ORGANIZADOR**  
***ORGANIZING COMMITTEE***

**COORDINACIÓN**

María Luisa Carrió Pastor

**SECRETARÍA ACADÉMICA**

Ana Botella Trelis

Miguel Ángel Candel Mora

Luz Gil Salom

Penny MacDonald Lightbound

Carmen Soler Monreal

Keith Stuart

**RESÚMENES DE PONENCIAS (orden alfabético del apellido del primer autor)**  
**ABSTRACTS (alphabetic order of surname)**

**Alcantud Díaz, María**

**Panel: 2. Discurso, análisis literario y corpus**

**VIOLENCE IN CHILDREN'S TALES: A SYSTEMIC CORPUS AND CRITICAL DISCOURSE ANALYSIS OF CINDERELLA**

The main aim of this article is to discuss the results achieved after investigating the presence of violence in the brothers Grimm's Cinderella (Tatar 1987,1992,2004); through a corpus-based analysis (Biber 1998) with the intention of finding out what kind of verbal processes predominate in this tale and whether they can be related to violent actions. The tool used for the analysis was WordSmith Tools 5 (Scott, 2010). The study involved first an analysis of frequencies of the lexical units in Cinderella, followed by a comparison of the results obtained in the frequency test to two reference corpora: British national Corpus and Cobuild Concordancer. The analysis was completed with a study of the concordances of some selected words, seeking in detail the context in which they appear. Once the quantitative and qualitative surveys were completed, I then proceeded to analyse the type of verbal processes (Halliday 1994:106-175) extracted from the frequency list. These were classified according to the framework proposed by Downing (2002:111). Thus verbal processes were classified as belonging to six categories: material, mental, verbal, behavioural, existential and relational. After classifying them, these same verbal processes were analysed according to four parameters: who (agent), what (type of action) to whom (affected) and under what circumstances. The results obtained in the frequency and concordance tests of this tale, seemed to indicate that violence is certainly present in Cinderella. The method proved to be a good tool to check whether each character's identity and their social position (power) were somehow related to the infliction of violence. That is, if some characters took the advantage of their predominant position and thus inflicted violence upon other characters. As a general conclusion of the analysis of the results, a tentative proposal could be formulated: that a corpus-based analysis in conjunction with both, a transitivity analysis and a critical discourse analysis, could empirically detect the presence of controversial and polemic topics such as violence in different types of texts. The results could be used as evidence to support a social intervention by means of a linguistic intervention (Graddol and Swann 1989) aimed at decreasing the amount of violent language and situations reproduced in children's tales.

**Alcaraz-Mármol, Gema and Lourdes Cerezo-García**

**Panel: 8. Los corpóra y la adquisición y enseñanza del lenguaje**

**SPECIFIC FREQUENCY AND ITS ROLE IN FOREIGN LANGUAGE VOCABULARY ACQUISITION**

Several studies (Saragi et al. 1978; Hulstijn et al. 1996; Reyes 1999; Waring and Takaki 2003; Pigada and Schmitt 2006; Webb 2007) have highlighted the role of specific frequency – i.e, the number of times a word occurs in a text – when it comes to second language vocabulary acquisition. In fact, especially in non-naturalistic contexts of learning, "individual texts within each corpus can vary from one to another and from the overall frequency list which a corpus produces" (Milton 2009: 25). As stated above, the specific frequency of a word may differ from general frequency. Knowing the number of times a word is to be encountered for acquisition would help designers create reading materials adjusted to the learners' needs. Unfortunately, to date, there is no agreement on the number of occurrences that are necessary for acquisition. What is more, we do not even know whether all words need to be encountered the same number of times. A number of studies have focused on this issue (Horst et al. 1998; Laufer 1998; Nation and Wang 1999; Rott 1999). Scholars have tried to determine, as accurately as possible, the number of times a word needs to occur to enable acquisition. What we find in this respect are various different outcomes, ranging from 5 and 20 occurrences. Yet, most of these works are carried out under artificial or laboratory conditions which may be far from mirroring the authentic learning context. The current study aims to approach the real situation of the classroom. It seeks to define the relationship between specific frequency and vocabulary acquisition within the context of EFL formal instruction. We pursue to answer two research questions:

1) Is there a significant relationship between specific frequency and immediate vocabulary acquisition, regarding receptive and productive knowledge?



**Carrió Pastor, Maria Luisa and Eva Mestre Mestre**

**Panel: 9. Usos específicos de la Lingüística de Corpus**

THE USE OF CORPUS ANALYSIS TO MANAGE FOREIGN LANGUAGE ACQUISITION IN A BILINGUAL COMMUNITY

Worldwide communication is possible nowadays using English as an international language or lingua franca. English is used in countries with different cultural backgrounds, a fact which affects in the use of pragmatic strategies. On occasions, authors who communicate in a foreign language cannot avoid the use of structures that are more common in their mother tongue (L1). In a monolingual community, language errors could be caused by L1 interference; nevertheless the methodology applied in error analysis and in corpus compilation could vary in a bilingual community. The linguistic status of three languages in contact may not be equal; consequently ideological, linguistic and social factors could influence language acquisition. The main objective of this paper was to find out if the general methodology used for corpora classification is adequate for a corpus of learners with different linguistic background. Furthermore, we analysed if the increasing importance of English as a lingua franca influences students to consider local or national languages less important when developing professional skills. In this article, we used corpus analysis methodology to determine if learners whose mother tongues were Spanish and Catalan varied their errors when learning English. Foreign language acquisition is a universal concept although we consider that the proficiency of some skills could depend on the mother tongue of the learner. In order to analyse the corpora, which included the errors of English texts written by students whose mother tongue was Catalan or Spanish, we conducted an experimental research that included the categories of communicative, grammatical and lexical errors. The results showed that students with different cultural backgrounds produced a dissimilar amount of communicative and lexical errors while both groups produced a similar amount of grammatical errors. As a consequence of this research, we concluded that the methodology used to detect errors should vary depending on the linguistic background of learners.

**Casas Pedrosa, Antonio Vicente**

**Panel: 3. Estudios gramaticales basados en córpora**

MAIN FEATURES OF ENGLISH PREDICATIVE PREPOSITIONAL PHRASES IN ICE-GB

This paper is aimed at identifying which are the main characteristics of those English prepositional phrases which perform the function of subject complement in the British component of ICE. Such is the case of "She first fell in love with Will when she was eighteen, and she adores him still" (ICE-GB:W2F-019 #47:1). After introducing the notions of prepositional phrase and subject complement, these structures will be described from the morphological, syntactic, semantic, lexical, and socio-pragmatic points of view and examples will be provided. Although in terms of frequency this is not the syntactic function prepositional phrases more often perform, they are taken into account because of their complexity and due to the lack of detailed analyses. In most cases they are described as isolated examples and this phenomenon is not considered to be a very productive one. Morphologically speaking, prepositional phrases can be defined as those phrases headed by a preposition which requires another unit following it and acting as its complement. Even though there is a wide range of units that can perform the function of complement of a preposition, attention will only be paid to noun phrases. They can be very simple (consisting of a single noun, as "on fire") or more complex (for instance, "in the pink of health"). From the syntactic point of view, prepositional phrases usually perform the functions of adverbial, postmodifier of noun phrases and complement of adjective and prepositional phrases. Nevertheless, they can also behave as subject and object complements: "That is of no importance" (Quirk et alii, 1985: 732) and "I don't consider myself at risk" (op.cit.: 733). As far as semantics is concerned, when acting as subject and object complements, prepositional phrases convey meanings which are similar to those of adjectives, since they express qualities or characteristics. Thus "on cloud nine" and "in the doldrums" can be replaced by "very happy" and "depressed", respectively. Lexically speaking, some of the examples under analysis are idiomatic, their



meaning being metaphorical. Such is the case of "(be) on tenterhooks", which is defined as follows in OALD6 (1340) as "(to be) very anxious or excited while you are waiting to find out sth or see what will happen". More information is provided as regards its origin: "From tenterhook, a hook which in the past was used to keep material stretched on a drying frame during manufacture". As far as socio-pragmatics is concerned, sometimes these structures are selected because they allow speakers to express the same meaning by means of a lower number of words. This is the case of "in hand", defined as "receiving attention and being dealt with" (OALD5: 537). Moreover many of these structures are labelled as "colloquial", "informal", "old-fashioned", or "slang" in dictionaries. In some cases they can even convey two different meanings, one being neutral and the other, informal; the phrase "on the job" in OALD6 (697), is thus defined as "while doing a particular job" and "(BrE, slang) Having sex".

**Castellón, Irene, German Rigau, Salvador Climent, Marta Coll-Florit and Marina Lloberes**

**Panel: 7. Lingüística computacional basada en corpus**

ANOTACIÓN SEMÁNTICA DEL CORPUS SENSEM

Este trabajo presenta la anotación semántica de los núcleos argumentales de SENSEM (Vázquez y Fernández 2008): sus objetivos, metodología, proceso, criterios y resultados. SENSEM es un banco de datos compuesto por un corpus del español y una base de datos interrelacionados. En su estado previo el corpus estaba etiquetado a nivel sintáctico en su totalidad, y a nivel semántico por lo concerniente a la semántica del núcleo verbal (Alonso et al. 2007). En esta investigación se ha afrontado la anotación semántica de los argumentos, centrándose en la de sus núcleos nominales, con el objetivo final de adquirir las preferencias semánticas de los predicados verbales. Las categorías lexico-semánticas utilizadas para la anotación son las de WordNet 1.6 del español (WNe) (Vossen ed. 1998), habiéndose usado asimismo como base de conocimiento de apoyo el Multilingual Central Repository (Atserias et al. 2004) el cual integra WNe con múltiples ontologías de propósito general. La anotación ha sido realizada por un equipo de 6 lingüistas y ha proporcionado los siguientes resultados:

- La anotación de 23.307 formas correspondientes a 3.693 lemas (82,6% del volumen total del corpus).
- Un conjunto de criterios de anotación, incluyendo instrucciones para anotadores, procedimiento de anotación de nombres propios, soluciones a problemas habituales y, especialmente, criterios para la desambiguación de significados.
- Un análisis en profundidad de la adecuación de WNe para la anotación semántica
- Un conjunto de propuestas para la solución de los problemas derivados de inadecuación de WNe: agrupación de sentidos y operadores especiales de anotación.

La principal característica de SENSEM es su diseño especialmente orientado a la estructura sintáctico-semántica del verbo, lo que se concreta en una constitución representativa y equilibrada de lemas y ocurrencias verbales y una anotación manual, detallada y en profundidad de las unidades verbales. La metodología de anotación utilizada incorpora la experiencia de Agirre et. al (2006) en la creación del corpus anotado del euskera Eusemcor. Se dividió en una fase de preparación técnica —preparación y anotación morfosintáctica del corpus mediante FreeLing (Padró et. al 2010) y adaptación de la interfaz de Eusemcor— y una fase de anotación en forma de secuencia de ciclos de etiquetado y establecimiento de criterios y de acuerdo entre anotadores y árbitros. Esta fase ha implicado el análisis de la adecuación de WNe para la anotación semántica de nombres, profundizándose en el ya iniciado por el grupo en Carrera et al. (2008). Como resultado, esta investigación ha generado instrucciones generales de anotación (e.g. aspectos de estructura léxico-semántica a considerar, fuentes primordiales de consulta), criterios de anotación y soluciones a problemas más frecuentes (e.g. aplicación de categorías MUC a la anotación de nombres propios, tratamiento de significados metafóricos o metonímicos, anotación de unidades multipalabra, de variantes morfológicas...). De forma especial se han definido criterios para la desambiguación de significados de WNe, sin duda el problema fundamental del proceso. El corpus SENSEM está a libre disposición de la comunidad bajo licencia GPL.