

AELINCO 2015 BOOK OF ABSTRACTS

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

Organizing Committee

Pedro A. Fuertes Olivera (Chair)

Esther Álvarez de la Fuente

Raquel Fernández Fuertes

Pilar Garcés García

Belén López Arroyo

Marta Niño Amo

Isabel Pizarro Sánchez

Ana Sáez Hidalgo

Ángeles Sastre Ruano

Marisol Velasco Sacristán

Student Helpers

Tamara Gómez Carrero

Noelia Recio

Silvia Sánchez Calderón

Scientific Committee

Francisco Alonso Almeida (Spain)

Theo Bothma (South Africa)

Pascual Cantos Gómez (Spain)

Gloria Corpas Pastor (Spain)

Raquel Criado Sánchez (Spain)

Danie Prinsloo (South Africa)

Teresa Fanego Lema (Spain)

María Luz Gil Salom (Spain)

María de los Ángeles Gómez González (Spain)

Sylviane Granger (Belgium)

Andrew Hardie (UK)

Ulrich Heid (Germany)

Julia Lavid López (Spain)

María José López Couso (Spain)

Juana I. Marín Arrese (Spain)

Antonio Moreno Sandoval (Spain)

Isabel Moskowich-Spiegel Fandiño (Spain)

José Luis Oncins (Spain)

Javier Pérez Guerra (Spain)

Emilio Ridruejo Alonso (Spain)

Jesús Romero Trillo (Spain)

Chelo Vargas Sierra (Spain)

And the members of the AELINCO Board

Aquilino Sánchez Pérez (Spain)

M^a Luisa Carrió Pastor (Spain)

Miguel Fuster Márquez (Spain)

Antonio Moreno Ortiz (Spain)

Sponsors

Financial support for this event has been received from

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

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

Institutional Sponsors

Universidad de Valladolid:

Departamento de Filología Inglesa

Research Unit: International Centre for Lexicography

Ministerio de Economía y Competitividad



Although in terms of frequency this is not the syntactic function PPs 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.

After introducing some basic notions, these structures are analyzed focusing on their presence in both spoken and oral texts within the ICE-GB. This is a one-million-word corpus which is both morphologically tagged and syntactically parsed. Moreover, it was compiled in the nineties and consists of both spoken (60%) and written material (40%).

The ICECUP (ICE Corpus Utility Program) software retrieved 3307 examples from 3223 sentences. These instances were then filtered since some of them were later classified as “noise” (in some cases the PPs were performing other functions either at phrase or at clause level and in others the element acting as the complement of the preposition was not a NP). For these reasons the final subcorpus consists of 1332 examples.

67.49% of these instances (899) are found in oral texts whereas 32.51% of them (433) belong to written texts. All these examples have been classified into different groups and subgroups corresponding to the different text categories available in this corpus (Nelson, Wallis and Aarts, 2002: 307-8). The results are presented in charts by means of both figures and percentages and different conclusions are later drawn based on the analysis of these charts.

Thus, for example, it can be noticed that, although it was expected that the amount of structures under study would be higher in spoken than in written texts because of the structure of the corpus itself, the relative frequency (which takes into account the relationship between the number of examples and the number of words) proves so, too: 0.1410% in spoken texts as opposed to 0.1022% in written texts, with an average of 0.1255% in the whole corpus. Moreover, there are more examples in dialogues (581) than in monologues (318) and in printed texts (332) than in non-printed ones (101).

This information proves especially relevant for non-native speakers of English since it allows them to become aware of the differences between speaking and writing. According to the evidence, some units are used more often in spoken language than in written English. Therefore, when producing any kind of text, students will feel more confident for they will be able to choose the appropriate structures bearing in mind these issues.

(23)

Casas-Pedrosa, Antonio Vicente (Universidad de Jaén, Spain): The economy principle and English predicative prepositional phrases

PANEL: CORPUS-BASED GRAMMATICAL STUDIES

This paper is aimed at analysing the relationship between the economy principle and English Predicative Prepositional Phrases (henceforth, PPPs). These can be defined as those phrases which are headed by a preposition whose Complement (C) is a Noun Phrase (NP) and which perform the function of Subject Complement (Cs). Such is the case of “under arrest” in [1]:

[1] The vessel remained under arrest from September the twenty-sixth until October the nineteenth (ICE-GB:S2A-065 #18:1:A).

As for the economy principle and the principle of the least effort, Vicentini (2003) studied the origin of these theoretical notions. Different examples obtained from the BNC and ICE-GB corpora and from various dictionaries confirm the hypothesis according to which the selection of certain PPPs will allow speakers to convey a given meaning by means of a

more reduced number of words. Thus, the PPPs “in clover” and “in hand” are defined as “to have enough money to be able to live a very comfortable life” (Turnbull, 20108: 278) and “receiving attention and being dealt with” (Crowther, 19955: 537), respectively:

[2] "As I was saying," Patrick Milligan continued, once his youngest was out of the house, "if the best came to the best, and your sister married the old codger, we could be in clover" (BNC:EEW 2057).

[3] In fact the repairs were already in hand <,,> (ICE-GB:S1B-069 #163:1:B).

These sentences clearly show that PPPs which are formally simple can express complex ideas. In fact, “in clover” and “in hand” illustrate the smallest structure of a PP, just consisting of a preposition and a NP as its C.

However, on some occasions certain PPPs are chosen to avoid redundant structures such as “be being”. In fact, the use of “at issue” and “under construction” in [4] and [5] prevents speakers from saying “may be being dealt with” and “which will be being built”, respectively:

[4] Again, the meaning of ‘necessary’ may be at issue but the important factor is that the presumption can be and, in many cases, probably will be cancelled out by express terms (BNC:HxD 175).

[5] One of the major features is a timber-framed house which will be under construction throughout the show, allowing visitors to see the various elements and skills involved (BNC:A16 61).

Furthermore, there are other reasons why PPPs are selected in certain communicative contexts. Thus, “in the club” is defined as “pregnant” (Rundell, 20072: 273; Turnbull, 20108: 279) and “off your chump” as “crazy” (Rundell, 20072: 240), but these PPPs also convey some other subtle nuances. For that reason they are labelled as “British informal old-fashioned” (Rundell, 20072: 273 and 240, respectively). Therefore, it can be concluded that these are counter-examples since in some cases speakers will opt for more complex structures (“pregnant” and “crazy” are one-word adjectives, whereas the PPs “in the club” and “off your chump”, on the contrary, consist of 3).

To this last group of examples belong some PPPs which can be classified as euphemisms. Rees (2006: v) defined them as follows: “[...] the word or phrase has the specific function of concealing something of the nature and meaning of what it describes”. Such is the case of the PPP “in Abraham’s bosom” in [6], which could be replaced by the adjective “dead”:

(24)

Choudhary, Prakash (National Institute of Technology Manipur, India), Nain, Neeta & Ahmed, Mushtaq (Malaviya National Institute of Technology Jaipur, India): A Linguistic Structure to Develop and Annotate Urdu Corpus for Multidisciplinary Research on Urdu Handwritten Documents

PANEL: CORPUS DESIGN, COMPILATION AND TYPES

In this proposal, we are describing a methodology for building an Urdu corpus CALAM (Cursive and Language Adaptive Methodology) including a large volume of 1200 handwritten text image file. A language independent structure has been designed to annotate handwritten Urdu script image for higher level at lines, words, components level with a XML standard to provide a ground-truth of each image at different four levels of annotation in a standard encoding UNICODE UTF-8. For capturing maximum variations in Urdu words and balanced the corpus, data collection is distributed within 6 categories and