INDIAN JOURNAL OF APPLIED LINGUISTICS
VOL. 39, NO. 1, JAN-JUN 2013

# MAXIECPC: Theoretical background and Descriptive research on General Statistics, Frequency Words and Keywords*

MARÍA CALZADA PÉREZ
*Jaume I University, Spain*

## ABSTRACT

*The present paper revolves around MaxiECPC, one of the various sub-corpora that make up ECPC (the European Comparable and Parallel Corpora), an electronic archive of speeches delivered at different parliaments (i.e. the European Parliament – EP – the Spanish Congreso de los Diputados – CD – and the British House of Commons – HC) from 1996 to 2009. In particular MaxiECPC comprises speeches from 1996 to 2003 with a total of 53,691,918 words. After a brief theoretical overview corpus linguistics (CL) and Translation Studies (TS), the overall methodological approach of ECPC research group (drawing on Tognini-Bonelli 2001, Sinclair 2003, and Scott & Tribble 2006) is described and implemented in its first steps, including a description of general statistics, frequency words, and keywords. Further research will complement the partial application of the methodology briefly sketched here.*

## 1. THEORETICAL BACKGROUND

In simple terms, an electronic corpus "can be described as a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria" (Bowker & Pearson 2002: 9). Within the ample territory of linguistics, the compilation of corpora eventually resulted in the creation of Corpus Linguistic, which, for McEnery & Wilson (1996: 1) meant "the study of language based on

* The present paper has been produced within the framework of research project *Enlarging and Deepening ECPC and ConcECPC: Theoretical, and Descriptive Studies and New Technological Developments*. This research project has been funded by the Spanish Ministry of Science and Innovation (FFI2008-01610) from 1st January 2009 to 30 June 2012.