Chapter 38
A Corpus–Based Approach to Teaching English in Abstracts

Amelia Maria Cava
University of Naples Federico II, Italy

ABSTRACT

In the academic field abstract is the first genre that students need to become familiar with. Abstracts have been investigated on structure and language by Swales (1990), Bhatia (1993, 1997), dos Santos (1996), Creminns (1982) and Day (1989) or by international organizations such as ISO and ANSI. Every academic journal presents its personal guidelines on how to write abstracts properly but generally, undergraduate students do not have a clear idea about how to write them. The main purpose of the present research study is to analyze the language used by authors with a pedagogical intent. The investigation is a corpus-based analysis setting out to highlight certain lexico-grammatical features specifically related to the ‘research-process language’. The data are drawn from a corpus of 6 years of three international journals: The International Journal of Primatology, Mathematics and Computers in Simulation and Proteins: Structure, Function, and Bioinformatics.

INTRODUCTION

In the digital era, Corpus Linguistics has become a viable methodology and is no more a ‘pseudo-procedure’, as defined in the past by many situational linguists (e.g. Widdowson 2003). This new trend is especially favored by the advent of more and more powerful computers and huge quantities of digital information. Corpora, as huge depository of digital data, do belong to the modern scene of language investigation. Corpus Linguistics has been used to test an hypothesis or to shed light on new trend in language. In other words, in corpus-based studies the starting point is clearly defined in advance and the investigation is not likely to discover unexpected events, in this perspective, Tognini-Bonelli suggests that:

[Researchers] adopt a ‘confident’ stand with respect to the relationship between theory and data in that they bring with them models of language and...
description which they believe to be fundamentally adequate, they perceive and analyze the corpus through these categories and sieve the data accordingly. (Tognini-Bonelli 2001, p. 66)

On the other hand, in a corpus-driven approach specific evidence is provided by the corpus itself, the researcher posits himself in a more ‘neutral’ position and the methodological path can be described as “observation leads to hypothesis leads to generalization leads to unification in theoretical statement” (Tognini Bonelli 2001, p. 85).

Many studies, to date, have used computer corpora to examine rhetorical phenomena in language and also new developments in literature. Lately, the relationship between corpora and learning has been examined in a more detailed way by several scholars such as Gavioli 2002, Aston 2001 and 2002, Bernardini 2000, Tribble and Barlow 2001, and Flowerdew 2001 with the main intent to teach students grammar, idioms and phraseology.

**BACKGROUND**

The integration of corpora is viewed as a coherent course design step at university settings (J. Flowerdew, 2001; L. Flowerdew, 2001 and 2002; Curado, 2001). A corpus-based analysis of language tends to play a key role in specialized language organization and methodology, especially in English for Specific Purposes (ESP), as suggested by Flowerdew (2001, p. 71). In agreement with Krishnamurthy (2001, p. 83), two important aspects justify corpus integration in language program: “a corpus can give us accurate statistics” and “a corpus can provide us with a vast number of real examples”.

As a consequence, the need of Information Technology (IT) skills in students is absolutely essential. When students become more and more familiar with the foreign language they are acquiring and their level of curiosity increases in a certain way, linguistic change tends to happen together with the development of strong skills in the area of information technology, however, in this occasion, linguistic confusion may arise in the ESP learner. Students need to use technology as a useful tool for learning. Among several scholars, Thompson (2002) proposes an electronic perspective of Internet and self-access study, based on the combination of CALL (Computer Assisted Language Learning) and corpus-driven language learning. He also refers to the need for setting up corpus instruments in an effective EAP framework, since many language instructors still ignore corpus exploitation possibilities for language teaching and learning. This because some scholars (e.g. Widdowson, 2003: 102) consider corpora as a good source of genuine and potentially interesting materials, but still presenting a fundamental discrepancy between their textual nature and the communicative nature of language learning. This is further aggravated by the fact that many of the widely accessible corpora, open source or commercially distributed corpora, were created as tools for linguistic research and not with pedagogical goals in mind, with the result that their content and design does not necessarily meet pedagogical needs. Nevertheless, small and medium-sized corpora could be the way of meeting shared interests and demands in the academic setting (Tribble, 1998; Scott, 2000; Curado, 2002).

The present paper has focused in particular on a corpus-based analysis and on communicative task integration in the EAP context. Two main goals have been followed: building corpus information in the academic context, and structuring corpus exercises as writing a research article abstract.

Over the last thirty years there has been an increasing awareness of the importance of phraseology in English language description. When referring to the term ‘phraseology’, this is used broadly and refers to what Clear (1993, p. 277) meant as “the recurrent co-occurrence of words”. In other words, as suggested by (Hunston 1995) natural languages tend to have more or less fixed co-occurrence of linguistic elements, words are co-selected rather than being selected separately.