CADIS – A Corpus for the Analysis of Identity Traits in Academic Discourse

Maurizio Gotti (Università di Bergamo)

Abstract

This paper focuses on a research project meant to investigate identity traits present in a corpus of texts written in English, drawn from academic journals pertaining to various disciplinary contexts. The study is part of a more general project focusing on the relationship between socioculturally-oriented identity-constructing factors and textual variation in different branches of English specialised discourse. The project takes into account the internationalisation of specialised discourse in English, not only in Anglophone countries but wherever institutional and professional settings evolve in a way that transcends the linguistic, cultural and conceptual standards of their local communities. In this sense, we can refer to a gradual 'globalisation' or 'hybridisation' of discursive practices that first appear in English-speaking environments but significantly affect also other languages, which are subject to standardising pressures in their semantic, textual, sociopragmatic and even lexico-grammatical construction.

For an in-depth analysis of variation in intercultural communication, our research unit is considering a variety of texts produced by scholars and academic institutions in various parts of the world, in order to identify textual variants due to the use of English as a first language, a second language, or a lingua franca within the scientific community. Our investigation will rely on a specially compiled corpus of texts for academic communication: CADIS. The corpus will also comprise some Italian texts for comparative purposes. Our approach is not limited, however, to linguistic evidence but is supplemented – wherever possible – with information gathered directly from the interactants.

This paper will concentrate on the criteria followed in the construction of such a specialised digital corpus and will discuss aspects pertaining to the collection and classification of texts according to their genre, the gathering of ethnographic data on the communicative events and actors involved, and a reconstruction of the general and specific sociolinguistic context of such texts. All these factors are expected to contribute to the definition of identity variants and their evaluation and interpretation is being carried out in the light of the latest research insights.