A Linguistic Contribution to the Study of Relationships Between Units of Different Levels:
Menzerath’s Law From an Intra-Systemic Perspective

Peter Grzybek¹, Emmerich Kelih¹, and Ernst Stadlober²

¹ Institut für Slawistik,
Universität Graz, Merangasse 70, A-8010 Graz, Austria
² Institut für Statistik,
Technische Universität Graz, Steyrergasse 17/IV, A-8010 Graz, Austria

Abstract As has been repeatedly shown, the frequencies with which word lengths (WL) and sentence lengths (SL) occur in texts and corpora, are very systematically organized. Furthermore, WL and SL frequencies (and statistical characteristics derived therefrom) are efficient factors in text classification, mainly for the distinction of particular discourse styles (cf. Grzybek et al. 2005; Kelih et al. 2005, 2006; Antić et al. 2006). Recently, the relationship between these two linguistic levels has been specifically studied (Grzybek and Stadlober 2006), referring to the theoretical context of Menzerath’s Law (cf. Altmann 1983), aiming at a theoretical modelling of the relation between units of different levels of a given system (the text, or corpus, in our case). Yet, with regard to the WL – SL relation, it has been predominantly quoted in inter-textually oriented studies, i.e. comparing word and sentence length between texts. However, in systematic re-analyses, only weak evidence has been found on the inter-textual level (Grzybek et al. 2007). The present study focuses on the intratextual level, analyzing the relations between WL and SL within a given text, particularly emphasizing the distinction of discourse types, thus potentially identifying discrimination and classification factors.

Key words: TEXT CLASSIFICATION, DISCOURSE TYPES

References