Abstract

This diploma thesis deals with the search, description, classification, and dictionary comparison of idioms based on data provided by GramatiKat application, which can identify nominal lemmas with anomalous word-form distribution. The aim of the diploma thesis is to determine whether grammatical anomalies in nouns (unusually high frequencies of one or more forms within a morphological paradigm) indicate formally anomalous multi-word lexemes (idioms) and how this relationship varies across different cases. Additionally, it explores the types of idioms that most commonly appear in individual cases of the singular. The analysis, utilizing SYN2015 a SYNv11 corpora, revealed that 28 % of analyzed lemmas are part of one or more idioms. The most common case is the accusative (88 %), while the least common is the vocative (5 %). The analysis also identified various groups of idioms with similar characteristics. For instance, verbal idioms were frequently observed in the dative, contact idioms predominated in the vocative, and grammatical idioms were prevalent in the locative.