

Summary of Gabrielė Virbickienė bachelor thesis “The Grammatical Variability of Lithuanian Idioms: Corpus-based Study”

The variability of multiword expressions has proven to be one of the key problems in recent attempts to improve automatic detection of multiword expressions. Idioms are characterised by either lexical or grammatical variability. This thesis will use a corpus-based study in order to analyse the grammatical variability of idioms in Lithuanian language. The main purpose of the thesis is to analyse the extent to which the grammatical variability of idioms is linked to the structure of idioms. The study analyses 427 idioms, which are used in contemporary language: 318 of which can be found both in *Lithuanian Nominal Phrases Dictionary* (corpus-based) and in *The Dictionary of Lithuanian Idioms* (2013) by J. Paulauskas, while the rest 109 idioms are featured in *Lithuanian Nominal Phrases Dictionary* only.

Idioms were analysed based on three types of grammatical variability: changes in morphological form, insertions and changes in word order. The results show that 56 % (242) of the analysed idioms allow variation. 58,8% of idioms allowing variation have changes in morphological form and 4,9% of them have insertions. Meanwhile, 27,6% of idioms allowing variation have changes in both word order and morphological form. All three types of grammatical variability were observed in 4,1% of analysed idioms, while changes in word order and insertions – in 2,9%.

After identifying the structural models of idioms that allow variation, it turned out that all idioms containing a verb as their main unit allow morphological variation, 72,5% of them allow changes in word order and 14% of them allow insertions. Meanwhile, idioms containing a noun as their main unit allow morphological variation in 97% of all cases. 4% of those idioms allow insertions, but they never allow changes in word order.

Idioms that do not allow variation differ in their structure from the ones that do. Only 4,7% of idioms that do not allow variation contain a verb in their structure (compared to 45,8% in idioms that allow variation). The verb in these instances take the form of subjunctive mood or imperative mood. In 23,3 % of idioms that do not allow variation, the nouns, numbers and adjectives or pronouns take the form of instrumental case (compared to 8,8% of variable idioms). Idioms that allow variation contain locative case in twice as many instances when compared to idioms that do not allow variation.

The analysis of the results shows that the variability of the idioms is clearly linked to their structure. Moreover, the idioms that allow variability are different from the ones that do not. These findings are important as corpus-based evidences about morphological variability of Lithuanian idioms and can be used to increase accuracy in automatic detection of multiword expressions.