

PHD DISSERTATION ABSTRACT

KAZANA EVANGELIA

**“Creating an electronic dictionary of multi-word verb expressions
by using corpora”**

SUPERVISOR: FOTOPOULOU AGGELIKI

**Linguist- Research Director - Electronic Lexicography and Language
Resources Dpt - Institute for Language & Speech Processing**

By using the term multi-word expressions we cover a wide range of linguistic structures which, in the international literature, are called idioms, frozen/fixed & semi-fixed/semi-frozen expressions, (set) phrases, formulae, collocations, cranberry collocations, phraseologisms, phraseological units.

Any sequence or collection of words that functions as a lexical unit, the meaning of which, most of the times, is not inferred from the combination of the meaning of its components, is considered to be a multi-word expression, and, especially, a stereotyped or a frozen/fixed one.

We could argue that multi-word expressions are divided into two main categories in terms of whether the meaning of these structures is lexicalized or analytical: in semantically analyzable expressions and non-analyzable (lexicalized) expressions.

All categories of multi-word expressions have some syntactic plasticity, which is not necessarily and absolutely related to the degree of lexicalization of the sentences, which means that it is not related to whether and to what extent they are semantically fixed/frozen. There are scales connected to both semantic and structural-grammatical behaviour of these sentences.

At the level of meaning: in some cases, there is transparency in meaning and it can be easily perceived even by non-native speakers of a language, while, in other cases, there is opacity, which means that, if the stereotypical sentence is not known to a non-native speaker, its meaning cannot be understood by him.

At the structural level: there are cases in which fixedness refers to the whole sentence. These are the cases of dicta and proverbs, but also of some verb expressions.

This dictionary will codify the grammatical and semantic properties that characterize multi-word verb expressions, while special emphasis will be given to the coding of the degree of fixedness of expressions which extends from non-analyzable to partially analyzable, the limits of which are not discernible.

Finally, it will be examined whether the fixedness is related to the ease or the difficulty in comprehension and learning expressions, (mainly by foreign students) in order to use the results in teaching and in various applications of Natural Language Processing (NLP).