

# IDION: VMWE documenting web-environment addressed to human user and NLP. Application to Greek and Pomak VMWEs.

Stella Markantonatou, Kyriaki Salveridou, Nicolaos Valeontis,  
Dimitris Karamatskos, George Pavlidis  
Institute for Language and Speech Processing/Athena RC

*Relevant UniDive working groups:* WG2

## 1 Introduction

We present the web user interface of the verb multiword expressions (hereinafter: VMWEs) resource IDION that, like other state of the art VMWE resources (Losnegard et al., 2016) is addressed to the human user (layman, linguist, language teacher/student) and to NLP. The interface for the annotator has been described in Markantonatou et al. (2019). IDION has been used to document about 2500 Greek and 160 Pomak VMWEs (Pomak is an endangered South East Slavic language variety (Karahóga et al., 2022)) in the framework of the Philotis project. Pomak VMWEs have been collected with field work and Greek VMWEs from various sources, including fieldwork and (Samaridi and Markantonatou, 2014). IDION's contents are available under a CC-BY-NC license.

## 2 Starting a search for a VMWE

In IDION, a VMWE lemma form contains (i) the lexicalised components (Savary et al., 2018, 94), namely the components with a fixed form and the components with a fixed lemma but not fixed form; the latter are in the form that best approximates the lemma convention, e.g., if the verb appears in the second and third persons of all numbers, in the VMWE lemma form it is in the second person singular and, (iii) place-holders for the free dependants of the VMWE, such as free NPs and free possessive pronouns. The VMWE lemma form represents the possibly fixed order of the VMWE components and otherwise, it keeps the lexicalised components together. A VMWE may have more than one lemma forms, the so-called *variants* due to orthographic variations, usage of diminutives, optional words such as appositions and adverbs, and the Dative Genitive alternation (here used as a cover term for the Dative Shift alternation (1) and the inalienable possession genitive and ethic genitive (2) alternation).

A fuzzy matching facility retrieves (lists of) VMWEs by searching their VMWE lemma forms

(variants are included). For each VMWE, its definition in its language and translations in other languages by professional translators are offered.

When a VMWE is selected, a set of tabs pop up; the first two lead to the variants of the lemma form (if any exist) and their English gloss. The remaining tabs are described below.

## 3 Tab: Usage examples

The tab "Corpus" provides access to usage examples. For each usage example a set of translations may be available, its source, the date when the example was obtained, whether the example is a grammatical utterance in its language, whether it is a case of gradation (mitigation, intensification) as well as its analysis in the Universal Dependencies formalism (hereinafter: UD).<sup>1</sup> Book references, URLs and individuals are among the possible sources of examples. Ungrammatical examples are sometimes provided, e.g., an example of the ungrammatical cliticisation of a lexicalised object NP (not all of them do); such examples are addressed to the human user and to NLP and draw on the intuitions of the annotators.

## 4 Tab: UD analysis of the lemma form

The tab "UD analysis" gives access to the graphical version of the UD analysis of the variants of the lemma form. The CoNLL-U version of the UD analysis can be viewed and downloaded through a dedicated button.<sup>2</sup> The UD analysis offers detailed structural information about the VMWE.

## 5 Tab: Syntactic flexibility of the VMWE

The tab "Flexibility" leads to a set of diagnostics of syntactic flexibility each one represented by a button: morphological flexibility of the verb (certain VMWEs may not occur in all persons, numbers and tenses), whether a word or a phrase can be inserted among the parts of a VMWE, word order

<sup>1</sup><https://universaldependencies.org/>

<sup>2</sup><https://universaldependencies.org/format.html>

permutations, cliticisation of lexicalised parts, usage of the VMWE in the passive voice, and Dative Genitive alternations (1),(2). The buttons of the diagnostics lead to usage examples exemplifying the corresponding property of the VMWE or to ungrammatical examples if the particular VMWE fails the flexibility diagnostic (since usage examples are retrieved from the web, failure in retrieving usages that satisfy a diagnostic is considered as evidence of inflexibility of the VMWE).

## 6 Semantic relations

VMWEs are documented for semantic relations: synonyms, opposites, causativity and gradation. This special feature of IDION has led to interesting observations about the semantic organisation of the VMWE field in Greek and, to some extent, in Pomak. Since IDION is not a general language lexical database, such as FrameNet<sup>3</sup> or Vallex<sup>4</sup>, a formal modeling of the interaction of semantics and subcategorisation properties/valency has not been included in its design.

### 6.1 Tab: Synonyms

Sets of synonymous VMWEs are dynamically generated at request with an implementation of the assumed transitive property of synonymy. This implementation infers that two expressions are synonymous even when their documentation does not contain this statement. The approach requires a large number of VMWEs in order to return really interesting results. The 2500 Greek VMWEs are arranged into about 770 synonyms sets, two thirds of which have more than two members.

A comment on the adopted treatment of synonymy is due. It is well-known that synonymy cannot be considered "the linguistic equality relation" because, in this way, synonymous would be the words or phrases that would be able to substitute each other in any context and such words or phrases hardly exist in any language. On the other hand, our everyday linguistic practice seems to consider synonymy a fact, e.g., when we explain the meaning of a word or a phrase using the language to which they belong (Hüllen, 2004, 38). It seems that humans can function well with a notion of close "semantic proximity" and this relation can order linguistic entities in distinguish-

able domains (Hüllen, 2004, 39). Transitivity is among the properties of the "semantic proximity relation" that helps defining such domains. Clearly this approach requires constant maintenance of the documentation by (near-)native speakers.

### 6.2 Tab: Opposites

VMWEs are documented for opposites, that is VMWEs describing situations that cannot hold simultaneously for the same entities, e.g., one cannot at the same time be denoted by the subject of the VMWE *to roll in money* and the VMWE *to be flat broke*. Opposition in language is a multidimensional, much discussed phenomenon described with rich terminology (Lyons, 1977, 270–287). In IDION, we have chosen the term *opposites* because it seems to denote the general idea described above. Contrary to the treatment of synonymy, only the list of opposites for which a VMWE has been documented is returned.

### 6.3 Tab: Causativity relations

Many Greek VMWEs appear in triples (for similar observations in English see Pulman (1993)) of causative (3), non-causative (4) and stative (5) counterparts or in pairs consisting of two of the above possibilities. Some of the verbs that participate in the causative-inchoative alternation, preserve this property in a VMWE. More frequently, though, the non-verbal lexicalised part of the VMWE forms the triple or the pair with morphologically unrelated verbs (3),(4),(5). Causative, non-causative and stative verbs seem to select each other in these triples/pairs. Indicative pairs are *afino-meno* 'leave, let - stay, remain', *vazo - beno* 'put - get into', *rihno - pefto* 'throw, drop - fall'. Often either the causative or the non-causative VMWE is the frequently used one but it is hardly the case that the one of them is in use and the other one is not. The "stative" counterparts present more gaps.

The tab provides access to all relations of this type for which the VMWE has been documented.

In a separate tab, IDION returns all the sets of synonymous VMWEs as well as all the causative - non causative - stative triples or pairs observed for a set of synonyms.

### 6.4 Tab: Gradation

The tab leads to intensifier and mitigator words, if usage examples of the VMWE containing them

<sup>3</sup><https://framenet.icsi.berkeley.edu/fndrupal/>

<sup>4</sup><https://ufal.mff.cuni.cz/vallex>

were retrieved (for a discussion of VMWE gradation in Greek see [Mexa and Markantonatou \(2020\)](#)).

(EL) Dative Shift

(1) *erikse xilo ston Ari*  
threw wood to.the Aris

DG *tou.DG erikse xilo*  
him threw wood

‘He beat Aris.’

(EL) Ethical dative (inalienable possession)

(2) *bgazi tin pisti*  
pulls.out the faith  
*ton pekton*  
of.the players

DG *tous.DG vgazi tin pisti*  
them pulls.out the faith

‘He makes the life of the players really difficult.’

(EL) Causative member of the triple (3)-(5)

(3) *kapios afini stin apexo kapon*  
smn leaves in.the outside smb

‘someone ignores somebody’

(EL) Non-causative member of the triple (3)-(5)

(4) *kapios meni stin apexo*  
someone stays in.the outside

‘someone reaches the condition of being ignored’

(EL) Stative member of the triple (3)-(5)

(5) *kapios ine stin apexo*  
someone is in.the outside

‘someone is in the condition of being ignored’

## Acknowledgements

This work was supported by “PHILOTIS: State-of-the-art technologies for the recording, analysis and documentation of living languages” (MIS 5047429): “Action for the Support of Regional Excellence” that is funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

## References

Werner Hülsen. 2004. *A history of Roget’s Thesaurus*. New York: Oxford University Press.

Ritvan Jusuf Karahoga, Panagiotis G Krimpas, Vivian Stamou, Vasileios Arampatzakis, Dimitrios Karatskos, Vasileios Sevetlidis, Nikolaos Constantinides, Nikolaos Kokkas, George Pavlidis, and Stella Markantonatou. 2022. Morphologically annotated corpora of pomak. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 179–186.

Gyri Smordal Losnegaard, Federico Sangati, Carla Parra Escartın, Agata Savary, Sascha Bargmann, and Johanna Monti. 2016. PARSEME Survey on MWE Resources. In *9th International Conference on Language Resources and Evaluation (LREC 2016)*.

John Lyons. 1977. *Semantics*, volume 1. Cambridge University Press.

Stella Markantonatou, Panagiotis Minos, George Zakis, Vassiliki Moutzouri, and Maria Chantou. 2019. *IDION: A database for Modern Greek multiword expressions*. In *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, pages 130–134, Florence, Italy. Association for Computational Linguistics.

Magda Mexa and Stella Markantonatou. 2020. *Intensifiers/moderators of verbal multiword expressions in Modern Greek*. In *Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress, 7-9 September 2021, Alexandroupolis, Vol. 1*, pages 163–169, Alexandroupolis. Democritus University of Thrace.

Stephen G. Pulman. 1993. The recognition and interpretation of idioms. In C. Cacciari and P. Tabossi, editors, *Idioms: Processing, Structure, and Interpretation*, Laurence Earlbaum Cognitive Science Monographs, pages 249–270. New Jersey.

Niki Samaridi and Stella Markantonatou. 2014. *Parsing Modern Greek verb Mwes with Lfg/xle grammars*. In *The 10th Workshop on Multiword Expressions (MWE 2014), Workshop at EACL 2014*, pages 33–37, Gothenburg, Sweden. EACL, SIGLEX-WE, PARSEME.

Agata Savary, Marie Candito, Verginica Barbu Mititelu, Eduard Bejcek, Fabienne Cap, Slavomır Ceplo, Silvio Ricardo Cordeiro, Gulsen Eryigit, Voula Giouli, Maarten van Gompel, Yaakov HaCohen-Kerner, Jolanta Kovalevskaite, Simon Krek, Chaya Liebeskind, Johanna Monti, Carla Parra Escartın, Lonneke van der Plas, Behrang QasemiZadeh, Carlos Ramisch, Federico Sangati, Ivelina Stoyanova, and Veronika Vincze. 2018. *PARSEME multilingual corpus of verbal multiword expressions*. In Stella Markantonatou, Carlos Ramisch, Agata Savary, and Veronika Vincze, editors, *Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop*, pages 87–147. Language Science Press., Berlin.