Chapter 5

A FrameNet approach to deep semantics for MWEs

🄳 Voula Giouli[a], Vera Pilitsidou[b] & Hephestion Christopoulos[b]

[a]Institute for Language and Speech Processing, ATHENA Research Center, Greece [b]National and Kapodistrian University of Athens, Greece

We present work aimed at enhancing a semantic lexical resource for Modern Greek with multiword expressions and at manually annotating a corpus with semantic roles with a view to supporting the lexical encoding with corpus evidence. The research was conducted within a larger initiative to construct a Greek FrameNet and corresponding corpus. The ultimate purpose was to provide a shallow semantic representation for multiword lexical units that is similar to the semantic representation of single-word predicates. We focus on both verbal and nominal multiword predicates. Specifically, we address the following questions: (a) what discrepancies seem to be prevalent between single- and multiword entries that are classified under the same frame (in terms of the realisation of Frame Elements), and (b) how to encode these discrepancies.

1 Introduction

Multiword expressions (MWEs) are word combinations that present morphological, lexical, syntactic, semantic, and pragmatic idiosyncrasies (Gross 1982, Baldwin & Kim 2010). In terms of meaning, they do not abide by the semantic interpretation rules of the language by which the meanings of phrases can be constructed out of the meanings of their constituents. In this respect, they appear on a continuum of compositionality: some expressions are analyzable (in that one can "analyze" their constituents in order to understand their meaning), whereas

others are partially analyzable or ultimately non-analyzable at all (Nunberg et al. 1994). The mismatch between their phrasal structure and their deep semantics renders them "a pain in the neck for Natural Language Processing" (Sag et al. 2002). In that regard, the community has been spending considerable effort to model them in a way that facilitates their robust treatment with a view to various applications. However, most MWE-specific lexical resources focus only on the representation of their lexical, morphological, and syntactic properties. Similarly, although several annotated corpora have been developed with the view to training and evaluating algorithms for MWE discovery and classification, little work has been devoted to their semantic representation in corpora with respect to developing applications that require deep semantics. Through our work, we seek to bridge this gap by providing a semantic representation for MWEs in a frame-based lexical resource for Modern Greek.

The chapter is structured as follows: in Section 2 we present the rationale, main objectives, and scope of our work; Section 3 gives an account of the theoretical framework within which our work is placed, as well as previous work on MWEs and their representation in large lexical resources and corpora. Section 4 outlines the methodological principles adopted for creating a frame-based lexical resource for Modern Greek and for treating MWEs. The MWEs that belong to the grammatical categories of noun and verb and their treatment within frames are presented in Section 5. In Section 6, we discuss our findings from the annotation we performed focusing on the discrepancies between single and multiword predicates. Finally, in Section 7, we outline our conclusions and plans for future research.

2 Main objectives

In this chapter, we present work aimed at (i) enhancing a semantic lexical resource for Modern Greek with nominal and verbal MWEs and (ii) manually annotating a corpus with attestations of the lexical units to the end of supporting the lexical encoding with further corpus evidence. The research was conducted within a larger initiative to construct a Greek FrameNet (FN-el) and corresponding corpus (Giouli et al. 2020, Pilitsidou & Giouli 2020). The main objective is to provide a semantic representation for MWEs in a way that is comparable to the one provided for single-word predicates. The goal was to develop a lexical resource coupled with corpus annotation that also treats complex predicates of various kinds; the resource will be useful for numerous Natural Language Processing (NLP) applications. Therefore, to better account for the deep semantics of

complex predicates, we wanted to define their argument structure and provide their lexical-semantic descriptions within the theoretical framework of frame semantics. Our dataset comprises a list of nominal and verbal MWEs extracted from corpora and existing resources in Modern Greek. In the paper, we give an account of their encoding by assigning them to a frame and defining their arguments along with the semantic roles they assume. The construction of the lexicon is based on corpus evidence and the performed annotation.

Finally, in our study, we address two questions: (a) What discrepancies seem to be prevalent between single- and multiword lexical units that are classified under the same frame in terms of Frame Elements assignment and syntactic realization? and (b) How are these discrepancies reflected in the encoding of MWEs and single-word predicates? In other words, what are the discrepancies between, for instance, the single word lexical unit (el) αποφασίζω *apofasizo* 'to decide' and the MWE (el) παίρνω απόφαση *perno apofasi* (lit. 'take decision') 'to decide' in terms of the Frame Elements that are realized? We will demonstrate that the differences between synonymous single- and multiword predicates involve not only variations in the syntactic realization of their (core and non-core) Frame Elements but also in the number of Frame Elements realized. Overall, analyzing these discrepancies might provide insights into how the choice between using a single word predicate and a MWE can influence the syntactic and semantic structure of a sentence, thereby impacting the realization of Frame Elements.

3 Theoretical framework and previous work

Our work draws upon the theory of *Frame Semantics* (Fillmore 1976, 1977, 1982, 1985) as well as the principles and methodologies established by pioneering research in lexical resources, that is inspired by the theory. Frame Semantics is an approach that does not rely on relations like hyperonymy and homonymy, but rather, draws upon the whole of human experience in order to organise the lexicon of any given language. This cognitive approach to the representation of meaning is based on the assumption that, in order to comprehend the meaning of any given utterance, one has to draw on their own experience and knowledge, thus evoking certain schemata. The theory focuses on the continuity that exists between language and experience (Petruck 1997). In this context, words gain their meaning within a semantic *frame*. A semantic frame schematises an event or a relation, encompassing a system of interconnected meanings. Understanding any one meaning within the frame necessitates grasping all the others. Thus, when any element of this frame is evoked in text or discussion, all other elements become accessible automatically (Fillmore 1982).

Based on Fillmore's theory, the Berkeley FrameNet (BFN, Baker et al. 1998) is a general-purpose lexical semantic resource for English, and it is the earliest and most complete attempt to organise and categorise lexical units in a lexicon based on frames. Frames are seen, thus, as conceptual structures describing specific types of objects, events, or states along with their components, the so-called *Frame Elements* (FEs) of the frame (Baker et al. 1998, Ruppenhofer et al. 2016), whereas the words that evoke a semantic frame are the *Lexical Units* (LUs) of that frame and are unique pairings of a word form and a meaning. Polysemous words typically evoke different frames. LUs pertain to the grammatical categories of verb, noun, adjective, or adverb. In other words, BFN provides a semantic representation that uses frames (or scenes) as its core, and LUs are ultimately organised around frames. Each frame is defined via a gloss that roughly describes the scene represented and a set of FEs; the latter are usually referred to in the gloss. FEs correspond to *semantic roles* specifically defined within each frame and provide finer distinctions of meaning compared to standard semantic roles. The resulting frame annotation scheme is therefore fine-grained. For each frame, the core FEs are generally assumed as central to the meaning conveyed by the frame. Frames are then populated with lexical units (LUs) – both single- and multiword ones. BFN is therefore a means for the semantic representation of LUs within frames regardless of the grammatical category they belong to (noun, adjective, verb, adverb). A set of typed frame-to-frame relations are used to link frames to one another, giving BFN a net-like structure, and – to some extent – a hierarchical organisation. Figure 1 depicts the frame Lending, its FEs – both core (i.e., LENDER, BORROWER, and THEME) and non-core (i.e., DURATION, TIME, PURPOSE, etc) – and the LUs that evoke the frame. A definition of the frame is provided as well as definitions for all the FEs.

Besides English, various FrameNets have been developed for other languages, for example, Japanese (Ohara et al. 2003, Saito et al. 2008), Chinese (You & Liu 2005), German (Erk et al. 2003, Boas 2002), Brazilian Portuguese (Salomão 2009, Timponi Torrent & Ellsworth 2013), Spanish (Subirats 2009), Italian (Lenci et al. 2010), Swedish (Borin et al. 2010), French (Candito et al. 2014), Hebrew (Hayoun & Elhadad 2016), Korean (Kim et al. 2016), Finnish (Lindén et al. 2017), and Modern Greek (Giouli et al. 2020, Pilitsidou & Giouli 2020). In this context, a rather recent initiative, namely, the Global FrameNet Shared Task (Timponi Torrent et al. 2018) seeks to investigate whether frames are universal – and to what extent – and whether BFN can cover the needs of most languages.

Similar to the general-purpose frame-based resources, other domain-specific ones have been implemented depicting language for specific purposes. For example, the language of sports and football has been modeled within the frame

Definition	The **Lender** gives the **Theme** to the **Borrower** with the expectation that the **Borrower** will return the **Theme** to the **Lender** after a **Duration** of time. This frame differs from the Borrowing frame in that this frame profiles the Lender in active sentences, whereas the Borrowing frame profiles the Borrower.
Example	**I** <u>**lent**</u> **my girlfriend my car for the weekend**
FEs - Core	
Borrower [Borr]	The person or institution who receives the **Theme** from the **Lender** for a **Duration**.
Lender [Lend]	The person or institution who gives the **Theme** to the **Borrower** for a **Duration**.
Theme [Th]	The object that is transferred from the **Lender** to the **Borrower** for a **Duration**.
FEs - Non-Core	
Duration [Dur]	The amout of time in which the **Borrower** has possession of the **Theme**.
Manner [Man] **Semantic Type:** Manner	The way in which the **Lender** lends the **Theme**.
Place [Pla] **Semantic Type:** Location	The location in which the Lender lends the **Theme** to the **Borrower**.
Purpose [Purp]	The aim of the **Lender** which they believe will be accomplished by lending the **Theme** to the **Borrower**.
Time [Time]	The time when the lending event occurs
Frame-frame Relations	Inherits from: Giving Perspective on: Temporary_transfer_scenario
Lexical Units	*lend.v, loan.n, loan.v*

Figure 1: The frame `Lending` in BFN

semantics paradigm in the so-called Kicktionary database (Schmidt 2009), as well as the Copa-2014 FrameNet Brasil, a frame-based trilingual electronic dictionary covering the domains of Football, Tourism, and the World Cup in three languages, namely, English, Spanish and Brazilian Portuguese (Timponi Torrent et al. 2014); similarly, the BioFrameNet database is a lexical resource built around frames in the domain of molecular biology (Dolbey et al. 2006), whereas frameNets tailored to model the legal (Venturi et al. 2009), financial (Pilitsidou & Giouli 2020) or aviation (Ostroški Anić & Brač 2022) domains have also been developed for languages other than English. Going further, FrameNets that are capable of taking other semiotic modes as input data, for example pictures, and videos have recently been implemented (Timponi Torrent et al. 2022).

The theory of Frame Semantics has been further utilised for the formulation of the Frame-based Terminology (FBT) theory (Faber 2011, 2015) and for the concomitant creation of frame-based terminological databases, like Ecolexicon (Faber & Buendía Castro 2014). Being a cognitive approach to terminology that is based on frame-like representations in the form of conceptual templates underlying the knowledge encoded in specialised texts, FBT directly connects specialised knowledge with Cognitive Linguistics and Semantics (Faber 2015). Specialised language concepts cannot be activated in isolation unless they are part of a larger structure or event. Our knowledge about a concept initially gives us the context or the event in which the concept retains its meaning. In this approach, frames are viewed as situated knowledge structures and are linguistically reflected in the lexical relations that arise from terminographic definitions. Concepts within a thematic field are thus inter-connected with each other based on the events of the field and the frames evoked. These frames are the context in which FBT specifies the semantic, syntactic, and pragmatic behavior of specialised language units. Consequently, instead of being described as static entities out of context, concept representations are treated as dynamic entities within the relevant context (Faber 2011).

Our work builds on the theory of Frame Semantics, Frame-based Terminology, and prior work on BFN, to create a lexical resource that incorporates LUs and frames that belong to language for general purposes (LGP) as well as to language for specific purposes (LSP). To elaborate, we have dealt so far with the grammatical categories of verbs and nouns. Both single and multiword entries have been included in the resource. It is worth mentioning that the majority of the MWE nouns in this work belong to LSP, in other words, they are terms, that is, lexical items characterised by their reference to a scientific field and constitute the (specialised) vocabulary of that field (Sager 1990).

3.1 MWEs in lexical resources

Two types of lexical resources may be identified with respect to MWEs: MWE-dedicated, that is, resources that have been developed with a primary focus on modeling MWEs, and MWE-aware ones that take MWEs into account in addition to other lexical units. Most MWE-dedicated lexical resources are primarily focused on the encoding of their lexical, morphological, and syntactic idiosyncrasies. Recommendations for representing MWEs in mono- and multilingual computational lexica (Calzolari et al. 2002, Copestake et al. 2002) aim at creating a shared model that is suitable for representing MWEs across different languages – yet, they focus mainly on the syntactic and semantic properties of support verbs and noun compounds and their proper encoding thereof. Similarly, Villavicencio et al. (2004) discuss the requirements for the efficient representation of English idioms and verb-particle constructions (VPCs) in lexica by means of augmenting existing single-word dictionaries with specific tables.

In this regard, within the Lexicon-Grammar framework (Gross 1975), French verbal MWEs were classified in the so-called Lexicon-Grammar tables (Gross 1982), where their syntactic and distributional properties and selectional restrictions were represented formally. In this approach, the surface structure of a verbal MWE is represented as a Part-of-Speech sequence of constituents, either continuous or not. The labels N, A, Adv, and PP are used to denote non-lexicalised constituents headed by a Noun, Adjective, Adverb, or Preposition respectively. Lexicalised elements are denoted as *C*. Modification, possible alternations, and distributional properties are encoded as binary properties within the Lexicon-Grammar tables. Along the same lines, similar lexical resources based on the same formal principles and linguistic criteria have been created for verbal idiomatic expressions in other languages, including Greek (Fotopoulou 1993, Mini 2009). The same approach has been adopted for the representation of adverbial MWEs in French by Laporte & Voyatzi (2008) and nominal MWEs in Greek by Anastasiadis-Symeonidis (1986).

Over the years, MWE-specific lexicons of various types have provided elaborate linguistic information for morphological, structural, and lexical properties of MWEs including variation and internal modification of MWEs. Shudo et al. (2011) report on the representation of Japanese MWEs in a comprehensive dictionary that provides detailed descriptions of their syntactic structure (dependencies), internal modification, and functional information. Similarly, Zaninello & Nissim (2010) propose a representation of MWEs in Italian based on their morphosyntactic properties and lexico-semantic information acquired semi-automatically from corpora. Odijk (2013) reports on the successful experiments and semi-automatic

expansion of DuELME (Grégoire 2010), a lexical database for Dutch MWEs; in the database, MWEs are classified in the so-called equivalence classes based on their syntactic structure, seen as syntactic patterns that occur frequently in a dependency parsed corpus of Dutch.

Recently, MWE-aware lexical resources provide elaborate representations of the structure of MWEs (cf. Leseva et al. 2024, Markantonatou et al. 2024 [this volume]) by making use of the Universal Dependencies formalism (Nivre et al. 2016). Similarly, the notion of the catena provides a mechanism for representing the structure of MWEs (cf. Osenova & Simov 2024 [this volume]). All these representations are aimed at the development of reliable gold standards to aid the task of MWE identification in running text.

In contrast, semantic MWE-aware lexicons, for example, WordNet (Fellbaum 1998), Verbnet (Kipper et al. 2008), SAID (Kuiper et al. 2003), and WikiMwe (Hartmann et al. 2012) give an account of various types of MWEs – yet they are solely focused on their semantic representation, overlooking other aspects. More recently, VerbAtlas (Di Fabio et al. 2019), a large-scale handcrafted lexical-semantic resource aimed at bringing together all verbal synonym sets from WordNet into semantically coherent frames, also treats verb-particle constructions (i.e., *take off*) as well as fully lexicalised idiomatic expressions (i.e, *kick* one's *heels*, *take a* firm *stand*, etc.), one of its main contributions being the definition of a set of explicit and cross-frame semantic roles that are linked to the selectional preferences of the verbal predicates.

Moreover, Fotopoulou et al. (2014) propose a model for encoding MWEs of all grammatical categories (noun, verb, adjective, and adverb) providing information on their syntactic structure, morphological and grammatical idiosyncrasies, variation, as well as information about their degree of fixedness. In addition, they provide lexical semantic relations (i.e., synonymy, antonymy, part-hole) giving an account of idiomatic expressions that also bear a literal meaning. To further account for the properties of Greek verbal MWEs, Markantonatou et al. (2019) have developed an infrastructure that accounts for the variability attested and the need for maximal generalisation.

3.2 MWEs in corpora and the corpus-lexicon interface

Besides lexical resources, the modeling of MWEs (i.e., their variations, internal modification, etc.) has also been attempted in both MWE-dedicated and MWE-aware corpora. Notably, the PARSEME initiative features corpora in more than 26 languages from different families that bear annotations for verbal MWEs (VMWEs) facilitating their discovery and identification in running text (Savary

et al. 2017, Ramisch et al. 2018, 2020, Savary et al. 2023). The annotation is performed based on guidelines that are as universal as possible, but which still allow for language-specific categories and tests. The DiMSUM 2016 shared task for joint identification and supersense tagging of nominal and verbal MWEs (Schneider et al. 2016) developed training and test data in English (tweets, service reviews, and TED talk transcriptions). Similarly, a MWE-related dataset in English, Portuguese, and Galician was released within the SemEval-2022 Task 2 (Tayyar Madabushi et al. 2022) on multilingual idiomaticity detection: the task was aimed at identifying whether a sentence contains an idiomatic expression, and at representing potentially idiomatic expressions in context based on semantic text similarity.

Other attempts at MWE semantic annotation in corpora include the annotation of MWEs in the Proposition Bank (PropBank), one of the earliest attempts to develop semantically annotated corpora (Palmer et al. 2005). Support verb constructions and idiomatic expressions in PropBank were later assigned one or more semantic role(s) depending on their meaning (Bonial et al. 2014a,b). Support verb constructions in PropBank were treated in two consecutive annotation iterations: initially, the light verbs were annotated as appropriate by selecting (or creating) the relevant support verb roleset; annotation proper was then performed on the predicative noun. However, one of the main drawbacks of PropBank is that the roleset used is too generic, thus leading to inconsistencies in labelling.

In between the corpus and the lexicon, Giouli (2023) proposes a model for representing the semantics of VMWEs by (a) taking into account their inherent idiosyncrasies: lexical, syntactic, and semantic, and (b) linking lexicon entries with their occurrences in a corpus that bears rich linguistic annotations (including Semantic Role Labelling). The model is claimed to entail a holistic approach to VMWE representation.

By default, BFN is placed in the lexicon-corpus and syntax-semantics interface. Therefore, it accounts for the semantics of lexical entries also considering context within frames. This holds true for single and multiword entries. Lexicalised noun-noun compounds (i.e., *wheel chair*.n), verb-particle constructions (i.e., *help out*.v), as well as idiomatic expressions (i.e., *aid and abet*.v, and *cook* someone's *goose*.v) are treated on their own as LUs, that pertain to the grammatical categories of noun or verb. For example, the verbal MWEs *aid and abet*.v and *help out*.v are both assigned to the frame `Assistance`, and their FEs along with their syntactic realisation are attested as shown in Table 1.

While BFN includes MWEs in the database, it does not analyze them internally. However, sentences in BFN bear a multi-layer annotation: Frame Element, Gram-

Table 1: Encoding of the MWE LU help out.v in BFN

Frame Element	syntactic realisation	n. of occurences
BENEFITED-PARTY	NP.Obj	3
FOCAL-ENTITY	PP(of).Dep	1
GOAL	DNI	2
HELPER	NP.subj	3

matical Function, and Phrase Type, and thus constitute clear examples of basic combinatorial possibilities (valence patterns) for each target LU. In this regard, all BFN annotations are constellations of triples that make up the FE realisation for each annotated sentence, each consisting of a FE or semantic role that is relevant to the frame itself (i.e., Agent, Experiencer, Cogniser, etc.), a grammatical function (i.e., Subject, Object) and a phrase type (i.e., Noun Phrase (NP), Verb Phrase (VP), Prepositional Phrase (PP), etc.). As a result, the syntactic realisation of the FEs is revealed via the annotation performed on the LUs and their FEs. This annotation provides us with a description of the syntactic valence properties of LUs, that is, the syntagmatic types that co-occur in the syntactic locality of the lexical item plus the grammatical functions they assume, as shown in (1):

(1) [All these commissions$_{\text{HELPER}}$] *helped* [me$_{\text{BENEFITED-PARTY}}$] *out* [of the pains$_{\text{FOCAL-ENTITY}}$]
[All these commissions.NP-SUBJ] *helped* [me.NP-OBJ] *out* [of the pains.PP]

Building on the dichotomy between the syntactic and semantic heads of expressions, only relatively recently has BFN given an account of the representation of support verb constructions in the database (Petruck & Ellsworth 2016). In this approach, the semantically empty support verb is assigned the tag *Supp*, whereas both frame assignment and annotation are performed with the predicative noun as the target as shown in (2).

(2) [Horatio$_{\text{PROTAGONIST}}$] *took*$^{\text{Supp}}$ a *dirty nap*. (Petruck & Ellsworth 2016)

FrameNets for other languages, for example, German, also treat MWEs of various types including support or light verb constructions, idioms, and metaphors (Burchardt et al. 2009). Finally, Borin (2021) discusses the inclusion of MWEs in the Swedish FrameNet++, also elaborating on the description of MWEs from a broad typological point of view. In this study, we elaborate on the idiosyncrasies of MWEs and the issues raised during annotation.

4 Methodology

In this section we present the methodology we adopted for building our frame-based lexical resource, outlining the different steps taken in the development process. It should be noted that the approach taken to FrameNet development is not uniform: teams have adopted various methodologies, ranging from manual construction entirely from scratch (in a way that is similar to the lexicographic process followed in BFN) to projecting translations from BFN to the target language, and even to semi-automatically grouping LUs for creating frames using data-driven techniques. In all these cases, the question raised is whether the frames defined in BFN for the English language are generally applicable to other languages as well, given the cultural differences entailed, as well as the idiosyncrasies and grammatical peculiarities of each language, and how and to what extent mappings from one FrameNet to another are feasible. From another perspective, there are three approaches to frame development (Ruppenhofer et al. 2016, Candito et al. 2014, Virk et al. 2021), namely, the lexicographic frame-to-frame strategy, the corpus-based lemma-to-lemma approach, and the full-text strategy. The lexicographic frame-to-frame strategy is aimed at documenting the range of syntactic and semantic combinatorial possibilities of words in each of their senses. Thus, annotation is performed on selected sentences of the corpus, that is, sentences that best record the valences of words. In this approach, annotation is relative to one lexical unit per sentence: the target. In general, we select sentences for annotation where, with the exception of subjects, all frame elements are realised locally by constituents that are part of the maximal phrase headed by the target word. The frame-by-frame strategy enforces coherence of annotations within a frame (Candito et al. 2014). By contrast, in the full-text annotation mode, all content words, that is, words bearing a lexical meaning, are treated as targets, and annotation is directed toward their dependents. In between the two strategies, the lemma-by-lemma annotation mode is focused on lemmas – possibly polysemous ones – rather than frames, and the annotation of these lemmas within different frames.

Although BFN was constructed as a general framework for applying semantic annotations on textual data cross-linguistically, certain frames need to be adapted to fit other languages. To this end, prior to annotation proper, a pilot annotation phase was carried out (Giouli et al. 2020) in which translations from BFN were projected to the Greek data. As shown in Table 2, in most cases, the BFN frames were applicable to the Greek data. However, we could not account for 12.3% of LUs, due to either a *frame shift* (i.e., a frame change) or a missing frame (i.e., a frame that is not provided for English). Researchers working on other languages also report frame shifts (Yong et al. 2022). To avoid shortcom-

ings and gaps, we opted for constructing the Greek FrameNet manually from scratch instead of projecting annotations.

Table 2: From BFN to FN-el: appropriateness of BFN to Greek.

	number	percent
perfect fit	549	87.70%
non perfect fit	54	8.63%
missing frame	23	3.67%
total	626	100.00%

After a closer inspection of the data, the following reasons for frame shifts were identified (in order of occurrence):[1]

- *Too specific*: the LU requires a frame more generic than the one available in the original database;

- *Too generic*: the LU requires a frame more specific than the one available in the original database;

- *Different causative alternation*: the LU requires a causative interpretation that is not present in the original frame, which may be either inchoative or stative;

- *Different inchoative alternation*: the LU requires an inchoative interpretation that is missing in the original frame, which may be either causative or stative;

- *Missing FE*: the original frame lacks a FE that is required in the target frame;

- *Extra FE*: there is a FE in the original frame that is not required in the target frame;

- *Different perspective*: the LU was proved to impose a perspective that is different from the one in the original frame;

- *Different stative alternation*: the LU requires a stative interpretation that is not present in the original frame, which may be either causative or inchoative;

---

[1]These tags were to a great extent adopted from Global FrameNet annotation.

- *Different entailment*: the LU has different entailments from the ones fore-seen by the original frame;

- *Different coreness status*: some non-core FE should be core in the target language.

Within the FN-el project, we adopted a modular approach to lexicon development, in the sense that predicates pertaining to a pre-defined set of semantic classes (namely, emotion, cognition, communication) or domains (finance, health) were selected and accounted for, thus opting for a domain-by-domain strategy.[2] More precisely, micro-projects were run towards treating predicates that pertain to each semantic class and/or domain. In this regard, we adopted the lemma-to-lemma strategy followed by a frame-to-frame one; multiple iterations of this procedure were conducted.

The task was organised as a four-stage procedure: (a) corpus creation and LU selection; (b) frame schematisation based on the syntactic and semantic properties of the selected LUs; (c) corpus annotation with a view to confirming or rejecting our initial intuitive decisions; and (d) frame validation and adjudication, where appropriate, and their extension with new LUs. More precisely, custom-made corpora of newswire texts, as well as corpora with a high term ratio that pertain to specialised domains were created to identify and extract words pertaining to the grammatical categories of noun and verb – also coupled with statistical information. An effort was made to extract the MWEs (verbal and nominal) from the corpora. N-grams were then extracted using SketchEngine (Kilgarriff et al. 2014), whereas terms were extracted semi-automatically using AntConc (Anthony 2005).[3]

After sense discrimination for polysemous words, meaningful groupings of word-sense pairings were performed – initially based solely on dictionary definitions. Frames were then constructed and populated with LUs; polysemous words fall under different frames, depending on their meaning within a given context. Each frame was further enhanced via the definition of the schema evoked and schematised via its FEs (core and non-core). Stipulating FEs was perhaps the most challenging aspect of the work. Note that core FEs grant a frame its uniqueness. Moreover, relations between frames were defined, the most important being *Inheritance*, *Perspective-on*, *Using*, *Subframe*, and *Precedes*.

---

[2]This is the approach taken to the French FrameNet construction (Candito et al. 2014) and is assumed to enforce the coherence of frame delimitations.

[3]Available online: http://www.laurenceanthony.net/.

This procedure for lexicon building is seen as the bottom-up part of the hybrid methodology we adopted: from corpora and lexical units to the definition of frames. The bottom-up approach to lexicon creation process was then complemented with a top-down one, according to which the frames were then populated with new LUs, that is, single- and multiword entries that are synonymous to the existing ones. The two approaches are complementary and were initiated in cycles during the project.

5 The treatment of MWEs in FN-el

Currently, our FN-el database contains c. 2,500 LUs organised around 62 frames. Of these, a total of 561 LUs are terms in the domain of finance, their termhood being determined based on specific criteria; we ended up with 39 frames (9 scenes) for the domain of finance. The remaining LUs are treated under frames in the semantic classes of activity, cognition, communication, and emotion. Numerical data regarding the current status of FN-el is depicted in Table 3.

Table 3: LUs in FN-el: numerical data.

	single	multiword	total
nouns	823	205	1028
verbs	671	572	1243
adjectives	127	32	159
adverbs	84	3	87
total	1705	812	2517

Each frame contains a definition of the scenario (gloss), the FEs (both core and non-core) along with the LUs that populate it. LUs that pertain to the grammatical categories of noun and verb have been extensively treated so far; both single and multiword lexical units are included in the resource and encoded as appropriate. An example of a frame in FN-el, namely, the `Agreement-or-Disagreement` one is presented in Figure 2. As shown, the gloss (definition) showcases the FEs of the frame – both core and non-core ones; FEs are also coupled with glosses. The LUs that evoke the frame are also provided. In our resource, we retain the respective terminology: names of frames, FEs, frame-to-frame relations, and glosses are all in English. In effect, using English as metadata ultimately facilitates the alignment of FN-el to BFN.

Definition	**COGNISER-1** and **COGNISER-2** hold a positive or negative opinion with respect to a **TOPIC**. **COGNISER-1** and **COGNISER-2** may also be referred to in the text collectively as **COGNISERS**. **COGNISER-1** and **COGNISER-2** may also appear as **OPINION_HOLDER-1** and **OPINION_HOLDER-2**. The **INTENSITY** of the opinion expressed, the **MANNER** of the expression and the **REASON** may also be expressed.
FEs - Core	
COGNISER	The **COGNISER** holds an opinion about a particular **TOPIC**; this opinion is seen in comparison to the opinion of another **COGNISER**.
TOPIC	A phenomenon or state or affairs that the **COGNISER** is considering with respect to their opinion.
OPINION_HOLDER	The **OPINION_HOLDER** holds a particular opinion, or point of view, which may be portrayed as being about a particular **TOPIC**.
FEs - Non-Core	
INTENSITY	Any description of the degree to which **COGNISER-1** and **COGNISER-2** hold an opinion or point of view about a particular **TOPIC**.
MANNER [Man] **Semantic Type:** Manner	Any description of how **COGNISER-1** and **COGNISER-2** hold and express the same or different opinion about a particular **TOPIC**.
REASON	Typically, the rationale or motivation behind the opinion held by the **COGNISERS**. In can be realised as a PP-ως constituent.
Lexical Units	*διαφωνία.n diafonia 'disagreement', διαφωνώ.v diafono 'disagree', δίνω τα χέρια.vid dino ta cheria (lit. 'give the hands) 'to agree', κάνω συμφωνία.lvc kano simfonia (lit. 'make agreement') 'to agree' συμφωνία.n simfonia 'agreement', σύμφωνος.adj simfonos 'congruent', συμφωνώ.v simfono 'to agree'*

Figure 2: The `Agreement-or-disagreement` frame in FN-el

MWEs that are listed as LUs in a frame appear in their *canonical form*: for nominal MWEs (NMWEs), that is, MWEs headed by a noun, the canonical form entails that the head noun is in the nominative case, singular number. A VMWE in its canonical form is a verbal phrase whose head verb is in a lemma form and whose other lexicalised components depend either on the verb or on another lexicalised component; non-lexicalised elements and open slots are not included in the canonical form. Since lexicon building is based on pre-processed data, we are no longer interested in the representation of the internal structure of the MWEs and their syntactic variations; these are depicted via the annotated instances that are included as examples in the database. We will elaborate on the treatment of MWEs and the representation of their valences in Sections 5.1 and 5.2.

5.1 Nominal MWEs

So far, 205 NMWEs have been included as LUs in the database and were assigned a frame based on their meaning. Currently, a large portion of the NMWEs encoded in FN-el are terms pertaining to the specialised language of finance and banking (133 LUs out of 205). The NMWEs for the financial domain were extracted semi-automatically from domain corpora using the methodology presented in Section 4. However, since these LUs belong to LSP, we had to diverge from BFN's frames in many ways described below. In terms of their structure, the NMWEs included in FN-el are constructions that have been extensively discussed in the literature on Modern Greek, namely, Adjective Noun (A N), Adjective Adjective Noun (A A N), Noun Noun (N N), and Noun Noun in the genitive (N Ngen) sequences (Anastasiadis-Symeonidis 1986, Ralli 2007, Gavriilidou 2013). In this regard, the NMWE in (3) is an A N construction headed by the N, whereas, the NMWE in (4) falls in the category of N Ngen constructions, where the second, non-head constituent is assigned the genitive case. The NMWE in (5) is an A A N continuous structure, where the third constituent, the noun, functions as the head, while (6) is an example of a N N structure, with its first constituent being the head.

(3) *κόκκινο δάνειο*
     kokkino danio
     red   loan
     'non-performing loan'

(4) *φόρος εισοδήματος*
     foros isodimatos
     tax   income.GEN
     'income tax'

(5) *καθαρά έντοκα   έσοδα*
     kathara entoka   esoda
     net.PL   interest.bearing.PL earnings.PL
     'net interest income'

(6) *δείκτης DAX*
     diktis   DAX
     index   DAX
     'DAX index'

These [A N] and [N NGEN] sequences are LUs with a non-compositional meaning, in that their meaning is not the product of the meaning of their parts. In this regard, the NMWE depicted in (3) is not a loan colored red, but a non-performing one. They are phrasal, and thus syntactic entities, sharing some features with (morphological) compounds, and are inaccessible for the syntactic operations that phrases normally allow. In that respect, they are continuous structures, in the sense that the order of their constituents is fixed, and no other elements can be inserted in between; in some cases, they do not even allow modification. Therefore, as in other lexicographic projects, one of the most challenging issues while creating the resource has been the recognition of NMWEs based on linguistic criteria, and their inclusion in a frame thereof.

Once they were assigned to a frame, the annotation of running text was performed. We aimed to find the syntactic structures MWEs occur in and the valences of MWEs. We will elaborate on the annotation and the issues raised in Section 6. The output of this annotation reveals the FEs that are specific to the LU at hand in the specific frame as well as their syntactic realisations. An example of the representation of a NMWE is provided in Table 4. Namely the multi-word LU (el) κόκκινο δάνειο.nmwe *kokino danio* (lit. 'red loan') 'non-performing loan' evokes the Lending frame to which it has been assigned as a LU of the grammatical category nmwe. Its definition (gloss) is provided in Greek as a paraphrase: (el) μη εξυπηρετούμενο δάνειο *mi exipiretumeno danio* 'non-performing loan'; it has also been assigned FEs as appropriate along with their realisations attested in the annotated corpus.

As shown in (7), the FE BORROWER is realised either as a NP in the genitive or as a PP headed by the preposition σε *se* 'to' as shown in (7) and (8) respectively. Once the BORROWER is realised as a NP in the genitive, the FE LENDER is instantiated by a PP headed by the preposition από *apo* 'by' as shown in (7); otherwise, it is realised as a NP in the genitive (8):

Table 4: he LU *κόκκινο δάνειο*.nmwe ('non-performing loan') in FN-el.

Frame element	Syntactic realisation	Occurences
BORROWER	NP.Dep	3
BORROWER	PP(σε).Dep	1
LENDER	NP.Dep	1
LENDER	PP(από)	1
AMOUNT	NP.Dep	1
DURATION	PP(για)	1
DURATION	NP.Dep	1
TIME	NP(μέχρι)	2
CAUSE	AJP.Dep	1
CAUSE	N.Dep	1

(7) *κόκκινα δάνεια* [επιχειρήσεων_{BORROWER}] [από την   ETE_{LENDER}]
  kokkina dania epichiriseon   apo tin   ETE
  red.PL loan.PL enterprise.PL.GEN   from the.SG.ACC NBG.SG.ACC
  'non-performing loans to households from NBG'

(8) *κόκκινα δάνεια* [τραπεζών_{LENDER}] [σε επιχειρήσεις_{BORROWER}]
  kokkina dania trapezon   se epichirisis
  red.PL loan.PL bank.PL.GEN   to enterprise.PL.ACC
  'non-performing loans to enterprises from NBG'

Notably, shifts or subtle differences in meaning or differences in perspective between LUs are made evident via their FEs. For example, both the multiword term (el) πιστωτικό γεγονός.nmwe *pistotiko γeγonos* (lit. 'credit event') 'bankruptcy' and its near synonym (el) πτώχευση.n *ptochefsi* 'bankruptcy' evoke the frame Wealth with INSTITUTION and PERSON being defined as core FEs of the frame. However, differences in the realisation of FEs shed light on the nuances of the two near-synonymous LUs; as shown in (9) and (10), the LU (el) πτώχευση accepts both PERSON and INSTITUTION as FEs, whereas the multiword term (el) πιστωτικό γεγονός accepts only INSTITUTION as displayed in (11) and (12).

(9) η *πτώχευση* [της   Thomas Cook_{INSTITUTION}]
  i ptochefsi tis   Thomas Cook
  the bankruptcy the.SG.GEN Thomas Cook
  'the bankruptcy of Thomas Cook'

(10) η   *πτώχευση* [ενός   εκ των   συζύγων$_{\text{PERSON}}$]
      i   ptochefsi enos   ek ton   sizigon
      the bankruptcy one.SG.GEN of the.PL.GEN spouse.PL.GEN
      'the bankruptcy of one of the spouses'

(11) *πιστωτικό γεγονός* [για την   Ελλάδα$_{\text{INSTITUTION}}$]
      pistotiko γeγonos gia tin   Elada
      credit   event   for the.SG.GEN Greece.SG.GEN
      'A credit event for Greece'

(12) * *πιστωτικό γεγονός* [για τον   σύζυγο$_{\text{PERSON}}$]
      pistotiko γeγonos gia ton   sizigo
      credit   event   for the.SG.ACC spouse.SG.ACC
      'A credit event for the spouse'

5.2 Encoding Verbal MWEs

Following the typology and criteria defined in the PARSEME initiative (Savary et al. 2017, Ramisch et al. 2018, 2020, Savary et al. 2023), four types of verbal MWEs have been included in the resource: (a) verbal idiomatic expressions (VIDs), that bear a meaning that cannot be computed based on the meaning of their constituents and the rules used to combine them, for example, (el) βάζω πλώρη *vazo plori* (lit. 'put.PRS.1SG prow.SG.ACC') 'to set forth'; (b) light verb constructions (LVCs), i.e., expressions with a rather transparent meaning that comprise a support or light verb that is semantically empty and a predicative noun or a predicative adjective or a prepositional phrase, for example, (el) δίνω υπόσχεση *dino yposchesi* (lit. 'give.PRS.1SG promise.SG.ACC') 'to promise'; (c) multi-verb constructions (MVCs), that is, expressions with coordinated lexicalised head verbs, for example, (el) απορώ και εξίσταμαι *aporo ke existame* (lit. 'wonder.PRS.1SG and be.very.surprised.PRS.1SG') 'to be very surprised'; and (d) verb-particle constructions (VPCs) comprising a verb and one of the adverbs (el) μπροστά *brosta* 'in front, forward', μπρος *bros* 'in front, forward', *πίσω* piso 'back', *πάνω* pano 'up', *κάτω* kato 'down', *μέσα* mesa 'in', *έξω* exo 'out, outside' in Greek. These adverbs are not morphologically derived from adjectives and exhibit most, if not all, of the properties that particles in other languages have (Giouli et al. 2019). Moreover, they have two distinct functions: as adverbs denoting time or location, they are used as modifiers; combined with prepositions, they form complex prepositions (Holton et al. 1997), as for example (el) μπροστά από *brosta apo* (lit. 'in-front from') 'in front of', (el) μέσα σε *mesa se* (lit. 'in to') 'in', (el) πάνω από *pano apo* (lit. 'over of') 'over', etc. Given their resemblance with VPCs in other languages

in terms of their properties, we decided to retain the latter class for Greek as well, and therefore expressions as the ones depicted in (13) and (14) were classified as VPCs. In terms of their semantics, VPCs were identified as non-compositional in meaning. As previously shown (Savary et al. 2019), these constructions are the most ambiguous. Depending on the context, they can be used literally and have a fully compositional meaning. In that case, they are not VMWEs.

(13) $\pi\acute{\epsilon}\varphi\tau\omega_i$ $\mu\acute{\epsilon}\sigma\alpha$ στις   προβλέψεις   μου$_i$
pefto   mesa stis   provlepsis   mu
fall.PRS.1SG in   to-the.PL.ACC prediction.PL.ACC my.1SG
'to succeed in my predictions'

(14) $\beta\acute{\alpha}\zeta\omega$ $\mu\pi\rho\sigma\varsigma$ τη   μηχανή
vazo   bros   ti   michani
put.PRS.1SG forward the.SG.ACC engine.SG.ACC
'to start the engine'

Once they were selected for inclusion, they were assigned a frame based on their semantics. As mentioned above, we have so far treated VMWEs that belong to the semantic domains of emotion, cognition, and communication – and the respective frames. For example, the LVCs (el) κάνω μάθημα.lvc *kano mathima* (lit. 'make.PRS.1SG lesson.SG.ACC') 'to teach', (el) δίνω μάθημα.lvc *dino mathima* (lit. 'give.PRS.1SG lesson.SG.ACC') 'to teach', (el) δίνω συμβουλή.lvc *dino symvuli* (lit. 'give.PRS.1SG advice.SG.ACC') 'to advice' and (el) δίνω οδηγία.lvc *dino odigia* (lit. 'give.PRS.1SG instruction.SG.ACC') 'to instruct', have been included in the resource within the Transferring-knowledge frame which also includes the single word LUs *διδάσκω*.v *didasko* ('to teach'), *μαθαίνω*.v *matheno* ('to teach'), etc. Variants of the selected VMWEs were included in the database as separate LUs and encoded as appropriate. For example, the LVC (el) παίρνω απόφαση *perno apofasi* (lit. 'take.PRS.1SG decision.SG.ACC') 'to decide' and its variant form (el) λαμβάνω απόφαση *lamvano apofasi* (lit. 'take.PRS.1SG decision.SG.ACC') 'to decide' are both treated as LUs in the Deciding frame; the latter has a formal register.

At the next stage, the arguments of the semantic predicate, that is, the VMWE taken as a whole, were identified and assigned FEs as appropriate. In this respect, we are no longer interested in the internal structure of the VMWE, that is, its fixed or lexicalised elements and the grammatical functions they assume, but rather in the non-fixed ones. Thus, FEs realised as arguments or adjuncts of the VMWE (taken as a whole) were identified and encoded.

6 Corpus annotation

Corpus annotation in BFN and related projects is aimed at documenting the range of syntactic and semantic combinatorial possibilities, or valences, of words in each of their senses. FrameNet annotation is always done relative to one particular lexical unit, the target, which is most often a single-word but can also be a multiword expression such as a phrasal verb (for example, *give in*) or an idiom (e.g., *take into account*). In this respect, the final step in our work was the annotation of selected instances of the MWEs used in context. One consideration, therefore, has been the selection of sentences from the corpus that will serve as ideal examples to annotate. This procedure resulted in the validation of frame definition and assignment and led to revisions and amendments where needed. The annotated corpus currently amounts to ca. 2600 sentences.

Annotation was performed on top of textual data that were pre-processed automatically via UDPipe (Straka & Straková 2017) at the levels of lemmatisation, part-of-speech (POS) tagging, and dependency parsing. Annotation on the lexical level was performed manually. Two students annotated selected sentences using the web annotation tool WebAnno (Yimam et al. 2013). Annotation was performed as a two-step procedure taking both verb and noun as targets. At the first stage, MWEs that constitute semantic predicates mapped onto a concept were selected. The selected markables were then annotated at the SemPred layer which is available as a WebAnno built-in module. According to the guidelines set, the markable was assigned a Part-of-Speech tag as appropriate, and the canonical form of the MWE at hand. A second span layer, namely, SemArg, represents slot fillers. The arguments and modifiers of the MWE (taken as a whole) were identified, and the semantic roles they assume were further specified. An instance of the annotation tool is illustrated in Figure 3.
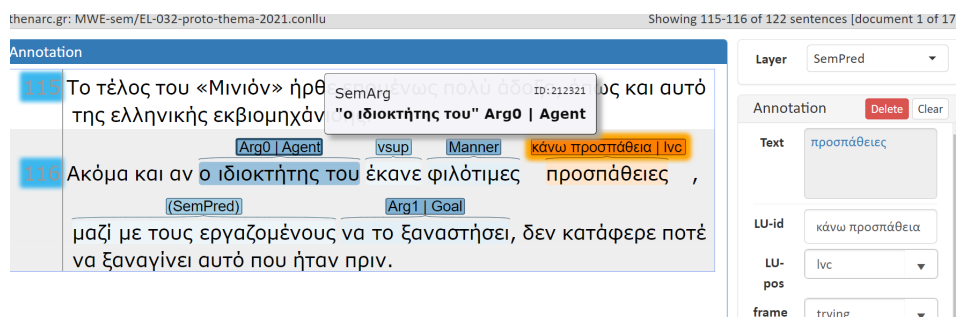

Figure 3: Annotating MWEs in Webanno.

Annotations were carried out independently by the two annotators; however, in order to ensure the highest quality of the dataset created, extended discussions followed each annotation cycle and adjudication of the annotations was performed where needed.

At this point, in order to better account for the properties of MWEs in Greek and their idiosyncrasies, a short description of the Greek language is in order. Modern Greek is a highly inflected language: nouns, adjectives, and certain pronouns show a rich inflectional system that features three grammatical genders (masculine, feminine, neuter), singular and plural numbers, and four cases (nominal, genitive, accusative, and vocative). The verbal inflectional system is equally rich: verbs inflect for person, number, tense, aspect, etc. Moreover, in terms of syntax, Greek is a language with a relatively free order of main constituents in a clause. The basic or unmarked order mainly follows the verb-subject-(object) pattern (Holton et al. 1997: 426); however, other variations are also attested, but these alternatives are appropriate in certain discourse contexts (Holton et al. 1997). This flexibility is due to case marking that signals the function of nominals: subjects are attested in the nominative case, whereas objects are most often in the accusative or in the genitive case; nominal complements of prepositions are also either in the accusative or the genitive. Finally, being a pro-drop language, Greek allows null subjects; the absence of a full or weak subject pronoun is accommodated by verbal morphology.

Following the above, MWEs often occur in various configurations. As a guideline, we tried to select sentences for annotation in which all FEs of the frame are realised by constituents that are part of the maximal phrase headed by the target word, including subjects – if possible. It should be noted that the BFN uses the Constructional Null Instantiation (CNI) tag as a mechanism to model the omitted constituents. Cases of CNI include the omitted subject of imperative sentences, the omitted agent of passive sentences, and of course null subjects, or the PRO-elements; we only adopted the afore-mentioned approach for the null subjects in cases where we had to include such sentences in the corpus.

6.1 Annotation with NMWEs as targets

Annotation with NMWEs as targets was relatively easy, as most NMWEs are continuous structures; modifiers of these NMWEs are realisations of their FEs. For example, the NMWEs (el) φόρος εισοδήματος.nmwe *foros isodimatos* (lit. 'tax income.sg.gen') 'income tax' and (el) τέλη κυκλοφορίας.nmwe *teli kykloforias* (lit. 'tax.pl circulation.sg.gen') 'road tax' which are subsumed under the Tax-payment frame, are annotated as taking the FEs taxpayer and amount, as shown in (15):

(15)   *φόρος εισοδήματος*   [*φυσικών*   *προσώπων*_{TAXPAYER}]
       foros isodimatos   fysikon   prosopon
       tax   income.SG.GEN natural.PL.GEN person.PL.GEN

       [3,7 δισ. ευρώ_{AMOUNT}]
       3.7 disekatomiria evro
       3.7 billion.PL.ACC euro.PL.ACC

       'personal income tax amounting to 3.7 billion euros'

In some cases, NMWEs come in the form of structures with shared heads as nested expressions, raising issues during annotation. As they are encoded as separate LUs in the database, annotation uses the feature *Null* retained for non-lexicalised constituents, and annotation is performed for each MWE separately.

(16)   Τα   [κόκκινα_{TYPE}] *στεγαστικά δάνεια*
       ta   kokina   stegastika dania
       the.PL red.PL   home.PL   loan.PL

       'the non-performing home loans'

When annotation was performed with a verb as targets, occurences of NMWEs were annotated as FEs of the respective frames. As shown in (17), the NMWE (el) κεντρική τράπεζα.nmwe *kentriki trapeza* 'central bank' is realised in the sentence as the BORROWER of the frame Lending in which the LU δανείζω.v *danizo* 'lend' occurs, whereas, the NMWE LU (el) εμπορικές τράπεζες *eborikes trapezes* 'commercial banks' (headed by the preposition από *apo* 'by') is realised as the LENDER.

(17)   [Η   *κεντρική*   *τράπεζα*_{BORROWER}] δανείζεται
       I   kentriki   trapeza   danizete
       The.SG.NOM central.SG.NOM bank.SG.NOM   borrow.PRS.3SG

       [χρήματα_{THEME}] [από τις   *εμπορικές*   *τράπεζες*_{LENDER}]
       chrimata   apo tis   eborikes   trapezes
       money.PL.ACC from the.PL.ACC commercial.PL.ACC bank.PL.ACC

       'The central bank borrows money from the commercial banks'

6.2 Annotation with VMWEs as targets

Annotation of VMWEs proved to be challenging. Only VMWEs in an idiomatic use were taken into account, whereas literal occurrences of MWEs were not annotated. Literal occurrences of MWEs, also referred to as their literal readings

or literal meanings, have received considerable attention equally from the linguistic and the computational communities. In an experiment run for German, Greek, Basque, Polish, and Brazilian Portuguese, Savary et al. (2019) report almost 11.5% of the VMWE occurrences in the Greek corpus to be literal readings of the VMWE surface forms – a phenomenon referred to as the *literal-idiomatic ambiguity*.[4] Other VMWEs were found to be semantically ambiguous (17% of the VMWEs), bearing different meanings based on the context. Usually, VIDs that comprise a verb predicate and the weak form of a personal pronoun are ambiguous, whereas LVCs and VPCs were also found to have more than one sense or usage.

In our database, 31 out of the 671 LUs that are VMWEs (4.77%) are also instances of polysemous entries. Following standard lexicographic practices, the latter were subsumed under different frames based on their meaning. For example, the LVC (el) δίνω απάντηση *dino apantisi* (lit. 'give answer') 'to answer' in (18) has been included in the `Communicating-response` frame; in a broader sense depicted in (19), it also evokes the `Expressing-opinion` one. The two frames are defined via two distinct sets of FEs as shown in Table 5.

Table 5: The `Communicating a response` and `Communicating an opinion` frames.

Frame	Definition	FEs
`Communicating a response`	A Speaker uses language (oral or written) to answer a certain Question that might be asked by an Enquirer. The Manner and Means might be mentioned.	Speaker Enquirer Topic Manner
`Communicating an opinion`	A Speaker or Statement uses language in order to share or make public their Opinion about a certain Topic. Their Strength of Opinion might be present as an adverb.	Speaker Opinion Topic Strength

Once their sense was disambiguated, encoding and annotating them posed no serious problems. Like single-word verb predicates, issues that arise during the annotation of VMWEs of all types are relevant to the granularity of the role-set employed or the specification of the appropriate role. Our approach to MWEs in FN-el is comparable to the approach taken in BFN – especially for the LVCs.

---

[4]For a definition of the literal-idiomatic ambiguity, see (Savary et al. 2019).

Annotation was performed with the semantic head, that is, the predicative noun, as the target as shown in (18) and (19).

(18) Η   υπουργός   *έδωσε*^{Supp}   σαφή   *απάντηση*
     i   ypurgos   edose   safi   apantisi
     The.SG.NOM minister.SG.NOM give.PST.3SG clear.SG.ACC answer.SG.ACC

     στους   μαθητές.
     stus   mathites
     to.the.PL.ACC students.PL.ACC

     'The minister gave clear answers to the students.'

(19) To   κείμενο   *δίνει*^{Supp}   πειστικές   *απαντήσεις*
     To   kimeno   dini   pistikes   apantisis
     The.SG.NOM text.SG.NOM give.PRS.3SG convincing.PL.ACC answer.PL.ACC

     σε αιώνια   προβλήματα.
     se eonia   provlimata
     to eternal.PL.ACC problem.PL.ACC

     'The text provides answers to eternal issues.'

Similarly, VIDs, MVCs, and VPCs were treated as a whole. The major issue we encountered, however, is due to the fact that, unlike NMWEs, VMWEs are highly discontinuous structures leading to issues in annotation, as shown in (20). To overcome this obstacle, layers of annotation provide the dependency graphs that are relative to a sentence. These may be retrieved to account for the structure of the MWE.

(20) Δεν είναι   διαφανής   η   *απόφαση*   που
     Den ine   diafanis   i   apofasi   pou
     Not is.PRS,3SG transparent.SG.NOM the.SG.NOM decision.SG.NOM that

     τελικά *έλαβαν.*
     telika elavan
     finally take.PST.3PL

     'The decision that they finally made was not transparent.'

Discrepancies between the single- and multiword LUs are abundant and need to be identified based on corpus evidence. In the remainder of the section, we will present the mismatches found in our data, which were depicted in the encoding. VMWEs were systematically found to have fewer FEs realised than their single-word counterparts. This is especially true for LVCs as opposed to their single-word counterparts. In most occurrences, the predicative noun is realised

in plural, indicating, thus, an aspectual reading of the LVC, i.e., repetition. In these cases, it is not the verb, but the nominal predicate that triggers the aspectual reading of the whole construction, whereas the verb remains bleached. For example, the multiword LU (el) παίρνω απόφαση.lvc *perno apofasi* (lit. 'take decision') 'to make a decision, to decide' and the single verb (el) αποφασίζω.v *apofasizo* 'to decide' both evoke the Deciding frame defined via the COGNISER and DECISION FEs. In our corpus, the LVC at hand was found to systematically realise only the COGNISER FE in the form of a NP in Subject position (in the nominative case), whereas it consistently lacks the DECISION one, as shown in (21); non-core FEs are usually realised as modifiers of the nominal predicate. By contrast, the FE DECISION is realised only in the single word LU as a to-clause, as depicted in (22).

(21) [Οι   ηγέτες_{COGNISER}] *παίρνουν* [υπεύθυνες_{MANNER}]
  i   igetes   pernun   ypeythines
  the.PL.NOM leader.PL.NOM take.PRS.3PL responsible.PL.ACC
  αποφάσεις.
  apofasis
  decision.PL.ACC
  'the leaders make decisions in a responsible way.'

(22) [Ο   Γιάννης_{COGNISER}] *αποφάσισε* [να φύγει_{DECISION}].
  O   Gianis   apofasise na figi
  The.SG.NOM John.SG.NOM decide.PST.3SG to leave
  'John decided to leave.'

Notably, certain VIDs bear a meaning that also incorporates one of their elements, most often intensifiers, but also other arguments as well. In this respect, the VPC in (23) incorporates the FE MANNER that is realised as the adjunct (el) σωστά *sosta* 'correctly' assumed by its single word counterpart μαντεύω.v *mantevo* 'to guess'. This is due to the fact that the VMWE (el) πέφτω μέσα.vpc *pefto mesa* (lit. 'fall in') 'to guess correctly', bears a positive reading in contrast to its single-word counterpart (el) μαντεύω.v *madevo* 'to guess' that bears a neutral reading. In these cases, we retain the FE at hand in the frame, but we encode it as being realised only in the single-word predicate based on corpus evidence.

(23) πέφτω μέσα
  pefto   mesa
  fall.PRS.1SG in
  'to guess correctly'

In most cases, the argument structure of complex predicates deviates from the patterns assumed by their single-word counterparts. This is particularly true about VIDs, due to the fact that they generally follow the valence of their syntactic verb head. For example, the single-word verbal predicate (el) εξοργίζω.v *exorgizo* 'to enrage' is an Object Experiencer verb, that is, a verb which assumes the FE EXPERIENCER (i.e. the entity that experiences the denoted emotion event); this FE is realised as a NP in the accusative case and in Object position. The CAUSE of the event is realised as an argument, that functions as the Subject of the verb, as shown in (24) (Giouli 2020). In contrast, in the case of the VID (el) ανεβάζω το αίμα στο κεφάλι *anevazo to ema sto kefali* (lit. 'raise.PRS.1SG the.SG.ACC blood.SG.ACC to-the.SG.ACC head.SG.ACC') 'to enrage', the core FE EXPERIENCER is the non-lexicalised element of the VMWE and is realised as a nominal complement (usually, the weak form of the personal pronoun) in the genitive case, whereas the CAUSE of the emotion is realised as a NP in Subject position, as depicted in (25). The weak pronoun (el) μου *moy* 'my' in the genitive case is due to the valence pattern entailed by the syntactic head of the VMWE; yet, it is annotated as EXPERIENCER.

(24) [O   Γιάννης$_\text{CAUSE}$] [με$_\text{EXPERIENCER}$] εξοργίζει.
     O   Giannis   me   exoryizi
     The.SG.NOM John.SG.NOM me1SG.ACC   enrage.PRS.3SG
     'John makes me furious.'

(25) [O   Γιάννης$_\text{CAUSE}$] [μου$_\text{EXPERIENCER}$] *ανέβασε   το*
     O   Giannis   moy   anevase   to
     The.SG.NOM John.SG.NOM me1SG.GEN   raise.PST.3SG the.SG.ACC
     αίμα   στο   κεφάλι.
     ema   sto   kefali
     blood.SG.ACC to.the.SG.ACC head.SG.ACC
     'John made me furious.'

Similar discrepancies are attested for other types of MWEs, for example, LVCs. Note that whereas the FE THEME is realised as a NP in the single word LU (el) αναφέρω.v *anafero* 'to mention', in (26), the same FE is realised as a PP headed by the preposition (el) σε *se* 'to' in the LVC (el) κάνω μνεία *kano mnia* (lit. 'make.PRS.1SG mention.SG.ACC') 'to mention' as shown in (27). These discrepancies between single- and multiword LUs in the realisation of their FEs have been studied and accounted for in the database based on corpus evidence.

(26) [Οι   Times] *αναφέρουν* [τις αντιδράσεις_THEME].
   I   Times anaferoyn tis antidrasis
   The.PL.NOM Times refer.PRS.3PL the.PL.ACC reaction.PL.ACC
   'The Times refer to the reactions.'

(27) [Οι   Times] *κάνουν* *μνεία*   [στις
   I   Times kanun mnia   stis
   The.PL.NOM Times make.PRS.3PL reference.SG.ACC to.the.PL.ACC
   αντιδράσεις_THEME].
   antidrasis
   reaction.PL.ACC
   'The Times refer to the reactions.'

Finally, syntactic alternations (i.e., passivisation, causative-inchoative alternation, etc.) that are attested for the single-word predicates of a frame are also attested for their VMWE counterparts, yet with different verbs as syntactic heads. This holds true for VIDs and LVCs alike. Indeed, LVCs which comprise the light verbs (el) βγάζω *vgazo* 'to take out' and (el) βγαίνω *vgeno* 'to be taken out' combined with the same predicative noun signal the causative – inchoative alternation and, in most cases, are assumed under the same frame. They predominately differ in the syntactic function of their lexicalised elements; as a result, the difference between the two is also depicted via their FEs and the grammatical function they assume. For example, the LVCs (el) βγάζω συμπέρασμα.lvc *vgazo symperasma* (lit. 'take-out.PRS.1SG conclusion.SG.ACC') 'to conclude' and (el) βγαίνει συμπέρασμα.lvc *vgeni symperasma* (lit. 'is-taken-out.3SG conclusion.SG.NOM') 'it is concluded' enter in the causative-inchoative alternation. In the former, the lexicalised element is the argument in object position (and following the rules of the language, it is realised as a NP in the accusative case); on the contrary, the latter has an argument in subject position as the lexicalised element. They are both assigned in the same Coming-to-Beleive frame, yet different FEs are realised for each one of them, since the two multiword LUs differ in the perspective: for the former, the COGNISER is realised, whereas the latter occurs with the THEME as shown in (28) and (29):

(28) [Οι   πολίτες_COGNISER] *βγάζουν* τα
   I   polites   vgazun   ta
   The.PL.NOM citizen.PL.NOM take.out.PRS.3SG the.PL.ACC
   συμπέρασματά τους.
   simperasmata tus
   conclusion.PL.ACC their3SG
   'Citizens come to a conclusion.'

(29) *Βγαίνει*   το   *συμπέρασμα*   [ότι η χώρα
      vgeni   to   simperasma   oti i   chora
      go.out.PRS.3SG.PRES the.SG.NOM conclusion.SG.NOM that the country

      κινδυνεύει_THEME].
      kindinevi
      is-in-danger

      'It is concluded that the country is in danger.'

7 Conclusions

We have presented work aimed at encoding MWEs that pertain to the grammatical categories of noun and verb to a frame-based lexical resource for Modern Greek. The work reported here is part of a larger initiative to construct a lexical database for Modern Greek with an inventory of language-specific frames around which to organise lexical units along the principles already set by BFN and other frame-based resources. Our MWE exploration has also taken into account multiword terms that pertain to the financial domain besides MWEs from the general language. For each MWE, we wish to provide information with respect to frame membership, valence, and access to a large number of annotated examples. Relations with other LUs (both single- and multiword ones) via the frame-to-frame relations already available in the resource have also been defined. The internal structure of the MWEs and their syntactic variations are depicted by means of the annotation layers that are available as pre-processing of the corpus; the focus is no longer on the representation of the internal structure of MWEs and their lexicalised elements, but on their valences; these are depicted via the annotated instances that are included as examples in the database.

Our contribution is two-fold: on the one hand, we provide an overview of the treatment of various types of MWEs in the Greek FrameNet aimed at mapping form onto meaning; on the other hand, we focus on the discrepancies between MWEs and their single-word counterparts. As we have shown, VMWEs were systematically found to have fewer FEs realised than their single-word counterparts bearing the same meaning. Moreover, in LVCs when the predicative noun is realised in plural an aspectual reading of the LVC is possible, i.e., repetition; this aspectual reading is also due to the missing FEs that denote a change in perspective. In a way, this type of representation allows us to provide the deep semantics of MWEs in a way that is comparable to the treatment of single-word lexical entries. For cases of polysemy and near synonymy, the strong apparatus of frame semantics allows us to explore distinct meanings of MWEs that pertain

to LSP (terms) and general language lexical entries alike via frame assignment and FE definition.

The work on FN-el is still in progress, and encoding is continuously subject to refinements and modification. Future work has already been planned towards enriching FN-el with new frames and LUs, both single and multiword ones. In another line of research, the alignment of FN-el frames with the BFN ones is currently underway. Finally, since this lexical resource provides the representation of the lexical and syntactic properties of the MWEs only via the annotated data, we plan to link FN-el to an existing lexical resource for Modern Greek that bears this information.

Abbreviations

BFN   Berkley FrameNet
FE   Frame element
FN-el   Greek FrameNet
LU   Lexical unit
LVC   Light verb construction
MWE   Multiword expression
NMWE   Nominal multiword expression
NP   Noun phrase
PP   Prepositional phrase
VID   Verbal idiomatic expression
VMWE   Verbal multiword expression
VP   Verb phrase
VPC   Verb-particle construction

Acknowledgements

References

Anastasiadis-Symeonidis, Anna. 1986. *Η νεολογία στην κοινή νεοελληνική ('Neology in Modern Greek')*. Thessaloniki: Aristotle University of Thessaloniki.

Anthony, Laurence. 2005. AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. In *Proceedings of the International Professional Communication Conference, 2005 (IPCC 2005)*, 729–737. IEEE.

Baker, Collin F., Charles J. Fillmore & John B. Lowe. 1998. The Berkeley FrameNet project. In *36th annual meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, vol. 1, 86–90. Montreal: Association for Computational Linguistics.

Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkhya & Fred J. Damerau (eds.), *Handbook of Natural Language Processing*, 267–292. Boca Raton, FL: CRC Press.

Boas, Hans C. 2002. Bilingual FrameNet Dictionaries for Machine Translation. In M. González Rodríguez & C. Paz Suárez Araujo (eds.), *Proceedings of the third international Conference on Language Resources and Evaluation*, 1364–1371. Las Palmas, Spain: European Language Resources Association (ELRA).

Bonial, Claire, Julia Bonn, Kathryn Conger, Jena D. Hwang & Martha Palmer. 2014a. Propbank: Semantics of new predicate types. In *Proceedings of the ninth international conference on Language Resources and Evaluation (LREC'14)*, 3013–3019. Reykjavik, Iceland: European Language Resources Association (ELRA).

Bonial, Claire, Meredith Green, Jenette Preciado & Martha Palmer. 2014b. An approach to *take* multi-word expressions. In *Proceedings of the 10th workshop on multiword expressions (MWE2014)*, 94–98. Gothenburg, Sweden: Association for Computational Linguistics.

Borin, Lars. 2021. Multiword expressions: A tough typological nut for Swedish FrameNet++. In Dana Dannélls, Lars Borin & Karin Friberg Heppin (eds.), *The Swedish FrameNet++: Harmonization, integration, method development, and practical language technology applications*, 221–259. Amsterdam: John Benjamins.

Borin, Lars, Dana Danélls, Markus Forsberg, Dimitrios Kokkinakis & Maria Toporowska Gronostaj. 2010. The past meets the present in Swedish FrameNet++. In *Proceedings of the 14th EURALEX International Congress*, 269–281.

Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó & Manfred Pinkal. 2009. Using FrameNet for the semantic analysis of German: Annotation, representation and automation. In Hans C. Boas (ed.), *Multilingual*

FrameNets in computational lexicography: Methods and applications, 209–244. Berlin, New York: De Gruyter Mouton.

Calzolari, Nicoletta, Charles J. Fillmore, Ralph Grishman, Nancy Ide, Alessandro Lenci, Catherine MacLeod & Antonio Zampolli. 2002. Towards best practice for multiword expressions in computational lexicons. In *Proceedings of the third international Conference on Language Resources and Evaluation (LREC'02)*, 1934–1940. Las Palmas, Canary Islands: European Language Resources Association (ELRA).

Candito, Marie, Pascal Amsili, Lucie Barque, Farah Benamara, Gaël de Chalendar, Marianne Djemaa, Pauline Haas, Richard Huyghe, Yvette Yannick Mathieu, Philippe Muller, Benoît Sagot & Laure Vieu. 2014. Developing a French FrameNet: Methodology and first results. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno & Stelios Piperidis Jan Odijk (eds.), *Proceedings of the ninth international Conference on Language Resources and Evaluation (LREC'14)*, 1372–1379. Reykjavik, Iceland: European Language Resources Association (ELRA). https://aclanthology.org/L14-1411/.

Copestake, Ann, Fabre Lambeau, Aline Villavicencio, Francis Bond, Timothy Baldwin, Ivan A. Sag & Dan Flickinger. 2002. Multiword expressions: Linguistic precision and reusability. In *Proceedings of the third international Conference on Language Resources and Evaluation (LREC'02)*. Las Palmas, Canary Islands: European Language Resources Association (ELRA).

Di Fabio, Andrea, Simone Conia & Roberto Navigli. 2019. VerbAtlas: A novel large-scale verbal semantic resource and its application to semantic role labeling. In *Proceedings of the 2019 conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 627–637. Hong Kong, China: Association for Computational Linguistics.

Dolbey, Andrew, Michael Ellsworth & Jan Scheffczyk. 2006. BioFrameNet: A domain-specific FrameNet extension with links to biomedical ontologies. In Olivier Bodenreider (ed.), *Formal biomedical knowledge representation: Proceedings of the second international workshop on Formal Biomedical Knowledge Representation (KR-MED 2006), collocated with the 4th International Conference on Formal Ontology in Information Systems (FOIS-2006)* (CEUR Workshop Proceedings 222). Baltimore: CEUR.

Erk, Katrin, Andrea Kowalski, Sebastian Padó & Manfred Pinkal. 2003. Towards a resource for lexical semantics: A large German corpus with extensive semantic annotation. In *Proceedings of the 41st Annual Meeting of the Association for*

Computational Linguistics (ACL'03), 537–544. Sapporo, Japan: Association for Computational Linguistics. DOI: 10.3115/1075096.1075164.

Faber, Pamela. 2011. The dynamics of specialized knowledge representation: Simulational reconstruction or the perception–action interface. *Terminology* 17(1). 9–29.

Faber, Pamela. 2015. Frames as a framework for terminology. In Hendrik J. Kockaert & Frieda Steurs (eds.), *Handbook of Terminology*, vol. 1. Amsterdam/Philadelphia: John Benjamins.

Faber, Pamela & Miriam Buendía Castro. 2014. EcoLexicon. In Andrea Abel, Chiara Vettori & Natascia Ralli (eds.), *Proceedings of the 16th EURALEX international congress*, 601–607. Bolzano, Italy: EURAC Research.

Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database.* Cambridge, MA: The MIT Press.

Fillmore, Charles J. 1976. Frame Semantics and the nature of language. *Annals of the New York Academy of Sciences* 280. 20–32.

Fillmore, Charles J. 1977. Scenes-and-frames semantics. In Antonio Zampolli (ed.), *Linguistic structures processing: Fundamental studies in computer science*, vol. 59 (Fundamental Studies in Computer Science), 55–81. Amsterdam; New York; Oxford: North Holland.

Fillmore, Charles J. 1982. Frame Semantics. In *Linguistics in the morning calm: Selected Papers from SICOL-1981*, 111–137. Seul, Korea: Hanshin Publishing Company.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di semantica* 6(2). 222–254.

Fotopoulou, Aggeliki. 1993. *Une classification des phrases à compléments figés en grec moderne: étude morphosyntaxique des phrases figées.* Université Paris VIII. (Doctoral dissertation).

Fotopoulou, Aggeliki, Stella Markantonatou & Voula Giouli. 2014. Encoding MWEs in a conceptual lexicon. In *Proceedings of the 10th Workshop on Multiword Expressions (MWE)*, 43–47. Gothenburg, Sweden: Association for Computational Linguistics.

Gavriilidou, Zoe. 2013. NN combinations in Greek. *Journal of Greek Linguistics* 13(1). 5–29.

Giouli, Voula. 2020. *Το σημασιολογικό πεδίο των συναισθημάτων: Ταξινόμηση των ρημάτων της νέας ελληνικής που δηλώνουν συναίσθημα.* ('The semantic field of emotions: A lexicon-grammar account of Greek verbs denoting emotion.* Greek. Athens, Greece: National & Kapodistrian University of Athens. (Doctoral dissertation).

Giouli, Voula. 2023. A model for representing the semantics of MWEs: From lexical semantics to the semantic annotation of complex predicates. *Frontiers in Artificial Intelligence* 6. DOI: 10.3389/frai.2023.802218.

Giouli, Voula, Vassiliki Foufi & Aggeliki Fotopoulou. 2019. Annotating Greek VMWEs in running text: A piece of cake or looking for a needle in a haystack? In Maria Chondrogianni, Simon Courtenage, Geoffrey Horrocks, Amalia Arvaniti & Ianthi Tsimpli (eds.), *13th International Conference on Greek Linguistics*, 125–134. University of Westminster, London, UK.

Giouli, Voula, Vera Pilitsidou & Hephaestion Christopoulos. 2020. Greek within the Global FrameNet Initiative: Challenges and conclusions so far. In *Proceedings of the International FrameNet Workshop 2020: Towards a global, multilingual FrameNet*, 48–55. Marseille, France: European Language Resources Association, (ELRA).

Grégoire, Nicole. 2010. DuELME: a Dutch electronic lexicon of multiword expressions. *Language Resources and Evaluation* 44(1). 23–39.

Gross, Maurice. 1975. *Méthodes en syntaxe: Régime des constructions complétives*. Paris: Hermann.

Gross, Maurice. 1982. Une classification des phrases « figées » du français. *Revue québécoise de linguistique* 11(2). 36–41.

Hartmann, Silvana, György Szarvas & Iryna Gurevych. 2012. Mining multiword terms from Wikipedia. In Maria Teresa Pazienza & Armando Stellato (eds.), *Semi-automatic ontology development: Processes and resources*, 226–258. IGI Global.

Hayoun, Avi & Michael Elhadad. 2016. The Hebrew FrameNet project. In *Proceedings of the tenth international conference on Language Resources and Evaluation (LREC'16)*, 4341–4347. Portorož, Slovenia: European Language Resources Association (ELRA).

Holton, David, Peter Mackridge & Irene Philippaki-Warburton. 1997. *Greek: A comprehensive grammar of the modern language*. London; New York: Routledge.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1. 7–36.

Kim, Jeong-uk, Younggyun Hahm & Key-Sun Choi. 2016. Korean FrameNet expansion based on projection of Japanese FrameNet. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System demonstrations*, 175–179. Osaka, Japan: The COLING 2016 Organizing Committee.

Kipper, Karin, Anna Korhonen, Neville Ryant & Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation* 42(1). 21–40.

Kuiper, Koenraad, Heather McCann, Heidi Quinn, Therese Aitchison & Kees van der Veer. 2003. *SAID*. Tech. rep. Philadelphia. DOI: 10.35111/MSVM-T728.

Laporte, Éric & Stavroula Voyatzi. 2008. An electronic dictionary of French multiword adverbs. In *Proceedings of the LREC workshop towards a shared task for Multiword Expressions (MWE 2008)*, 31–34.

Lenci, Alessandro, Martina Johnson & Gabriella Lapesa. 2010. Building an Italian FrameNet through semi-automatic corpus analysis. In *Proceedings of the seventh international conference on Language Resources and Evaluation (LREC'10)*, 12–19. Valletta, Malta: European Language Resources Association (ELRA).

Leseva, Svetlozara, Verginica Barbu Mititelu, Ivelina Stoyanova & Mihaela Cristescu. 2024. A uniform multilingual approach to the description of multiword expressions. In Voula Giouli & Verginica Barbu Mititelu (eds.), *Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives*, 73–116. Berlin: Language Science Press. DOI: 10.5281/zenodo.10998635.

Lindén, Krister, Heidi Haltia, Juha Luukkonen, Antti Olavi Laine, Henri Roivainen & Niina Väisänen. 2017. FinnFN 1.0: The Finnish frame semantic database. *Nordic Journal of Linguistics* 40(3). 287–311.

Markantonatou, Stella, Nikolaos T. Kokkas, Panagiotis G. Krimpas, Ana O. Chiril, Dimitrios Karamatskos, Nicolaos Valeontis & George Pavlidis. 2024. Description of Pomak within IDION: Challenges in the representation of verb multiword expressions. In Voula Giouli & Verginica Barbu Mititelu (eds.), *Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives*, 39–72. Berlin: Language Science Press. DOI: 10.5281/zenodo.10998633.

Markantonatou, Stella, Panagiotis Minos, George Zakis, Vassiliki Moutzouri & 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) at ACL 2019*, 130–134. Florence. DOI: 10.18653/v1/W19-5115.

Mini, Marianna. 2009. *Linguistic and psycholinguistic study of fixed verbal expressions with fixed subject in Greek: A morphosyntactic analysis, lexicosemantic gradation and processing by elementary school children*. University of Patras. (Doctoral dissertation).

Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty & Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the tenth international Conference on Language Resources and Evaluation (LREC 2016)*, 1659–1666. Portorož, Slovenia: European Language Resources Association (ELRA).

Nunberg, Geoffrey, Ivan A. Sag & Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538.

Odijk, Jan. 2013. Identification and lexical representation of multiword expressions. In Peter Spyns & Jan Odijk (eds.), *Essential speech and language technology for Dutch: Results by the STEVIN programme* (Theory and Applications of Natural Language Processing), 201–217. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-30910-6_12.

Ohara, Kyoko, S. Fujii, Hiroaki Saito, S. Ishizaki, T. Ohori & Ryoko Suzuki. 2003. The Japanese FrameNet project: A preliminary report. In *Proceedings of Pacific Association for Computational Linguistics (PACLING'03)*, 249–254. Halifax, Canada: Pacific Association for Computational Linguistics.

Osenova, Petya & Kiril Simov. 2024. Representation of multiword expressions in the Bulgarian integrated lexicon for language technology. In Voula Giouli & Verginica Barbu Mititelu (eds.), *Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives*, 117–146. Berlin: Language Science Press. DOI: 10.5281/zenodo.10998637.

Ostroški Anić, Ana & Ivana Brač. 2022. Airframe: Mapping the field of aviation through semantic frames. In Annette Klosa-Kückelhaus, Stefan Engelberg, Christine Möhrs & Petra Storjohann (eds.), *Dictionaries and society: Proceedings of the XX EURALEX international congress*, 334–345. Mannheim: IDS-Verlag.

Palmer, Martha, Daniel Gildea & Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics* 31(1). 71–106.

Petruck, Miriam R. L. 1997. Frame semantics. In Jef Verschueren, Jan-Ola Östman, Jan Blommaert & Chris Bulcaen (eds.), *Handbook of pragmatics*, 1–13. Amsterdam: John Benjamins.

Petruck, Miriam R. L. & Michael Ellsworth. 2016. Representing support verbs in FrameNet. In *Proceedings of the 12th workshop on Multiword Expressions*, 72–77. Berlin: Association for Computational Linguistics.

Pilitsidou, Vera & Voula Giouli. 2020. Frame Semantics in the specialized domain of finance: Building a termbase to aid translation. In Zoe Gavriilidou, Maria Mitsiaki & Asimakis Fliatouras (eds.), *Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress*, vol. 1, 263–271. Alexandroupolis: Democritus University of Thrace.

Ralli, Angela. 2007. *Η Σύνθεση των Λέξεων: Διαγλωσσική, Μορφολογική Προσέγγιση ('Compounding: A cross-lingual, morphological approach')*. Athens: Patakis.

Ramisch, Carlos, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archna Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Güngör, Abdelati Hawwari, Uxoa Iñurrieta, Jolanta Kovalevskaitė, Simon Krek, Timm Lichte, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Behrang QasemiZadeh, Renata Ramisch, Nathan Schneider, Ivelina Stoyanova, Ashwini Vaidya & Abigail Walsh. 2018. Edition 1.1 of the PARSEME shared task on automatic identification of verbal multiword expressions. In Agata Savary, Carlos Ramisch, Jena D. Hwang, Nathan Schneider, Melanie Andresen, Sameer Pradhan & Miriam R. L. Petruck (eds.), *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, 222–240. Santa Fe, NM: Association for Computational Linguistics.

Ramisch, Carlos, Agata Savary, Bruno Guillaume, Jakub Waszczuk, Marie Candito, Ashwini Vaidya, Verginica Barbu Mititelu, Archna Bhatia, Uxoa Iñurrieta, Voula Giouli, Tunga Güngör, Menghan Jiang, Timm Lichte, Chaya Liebeskind, Johanna Monti, Renata Ramisch, Sara Stymne, Abigail Walsh & Hongzhi Xu. 2020. Edition 1.2 of the PARSEME shared task on semi-supervised identification of verbal multiword expressions. In Stella Markantonatou, John Mccrae, Jelena Mitrović, Carole Tiberiu, Carlos Ramisch, Ashwini Vaidya, Petya Osenova & Agata Savary (eds.), *Proceedings of the joint workshop on Multiword Expressions and Electronic Lexicons (MWE-LEX 2020)*, 107–118. Barcelona: Association for Computational Linguistics.

Ruppenhofer, Josef, Michael Ellsworth, Miriam R.L. Petruck, Christopher R. Johnson & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice.* https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In Alexander F. Gelbukh (ed.), *Proceedings of the third international conference on Intelligent Text Processing and Computational Linguistics (CICLing 2002)*, 1–15. Springer.

Sager, Juan C. 1990. *A practical course in terminology processing*. Amsterdam: John Benjamins.

Saito, Hiroaki, Shunta Kuboya, Takaaki Sone, Hayato Tagami & Kyoko Ohara. 2008. The Japanese FrameNet software tools. In *Proceedings of the sixth international conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA).

Salomão, Maria Margarida M. 2009. Framenet Brasil: Um trabalho em progresso. *Caleidoscópio* 7(3). 171–182.

Savary, Agata, Cherifa Ben Khelil, Carlos Ramisch, Voula Giouli, Verginica Barbu Mititelu, Najet Hadj Mohamed, Cvetana Krstev, Chaya Liebeskind, Hongzhi Xu, Sara Stymne, Tunga Güngör, Thomas Pickard, Bruno Guillaume, Eduard Bejček, Archna Bhatia, Marie Candito, Polona Gantar, Uxoa Iñurrieta, Albert Gatt, Jolanta Kovalevskaite, Timm Lichte, Nikola Ljubešić, Johanna Monti, Carla Parra Escartín, Mehrnoush Shamsfard, Ivelina Stoyanova, Veronika Vincze & Abigail Walsh. 2023. PARSEME corpus release 1.3. In *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023)*, 24–35. Dubrovnik, Croatia: Association for Computational Linguistics. https://aclanthology.org/2023.mwe-1.6.

Savary, Agata, Silvio Ricardo Cordeiro, Timm Lichte, Carlos Ramisch, Uxoa Iñurrieta & Voula Giouli. 2019. Literal occurrences of multiword expressions: Rare birds that cause a stir. *The Prague Bulletin of Mathematical Linguistics* 112(1). 5–54.

Savary, Agata, Carlos Ramisch, Silvio Cordeiro, Federico Sangati, Veronika Vincze, Behrang QasemiZadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova & Antoine Doucet. 2017. The PARSEME shared task on automatic identification of verbal multiword expressions. In Stella Markantonatou, Carlos Ramisch, Agata Savary & Veronika Vincze (eds.), *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, 31–47. Valencia, Spain: Association for Computational Linguistics. DOI: 10.18653/v1/W17-1704.

Schmidt, Thomas C. 2009. The Kicktionary: A multilingual lexical resource of football language. In Hans C. Boas (ed.), *Multilingual FrameNets in computational lexicography: Methods and applications*, 101–134. Berlin, New York: De Gruyter Mouton.

Schneider, Nathan, Dirk Hovy, Anders Johannsen & Marine Carpuat. 2016. SemEval-2016 task 10: Detecting minimal semantic units and their meanings (DiMSUM). In *Proceedings of the 10th international workshop on Semantic Evaluation (SemEval-2016)*, 546–559. San Diego, California: Association for Computational Linguistics.

Shudo, Kosho, Akira Kurahone & Toshifumi Tanabe. 2011. A comprehensive dictionary of multiword expressions. In *Proceedings of the 49th Annual Meeting of*

the Association for Computational Linguistics: Human Language Technologies, 161–170. Portland, OR: Association for Computational Linguistics.

Straka, Milan & Jana Straková. 2017. Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 shared task: Multilingual parsing from raw text to universal dependencies*, 88–99. Vancouver, Canada: Association for Computational Linguistics.

Subirats, Carlos. 2009. Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In Hans C. Boas (ed.), *Multilingual FrameNets in Computational Lexicography*, 135–162. Berlin/New York: Mouton de Gruyter.

Tayyar Madabushi, Harish, Edward Gow-Smith, Marcos Garcia, Carolina Scarton, Marco Idiart & Aline Villavicencio. 2022. SemEval-2022 Task 2: Multilingual idiomaticity detection and sentence embedding. In *Proceedings of the 16th international workshop on Semantic Evaluation (SemEval-2022)*, 107–121. Seattle, WA: Association for Computational Linguistics.

Timponi Torrent, Tiago & Michael Ellsworth. 2013. Behind the labels: Criteria for defining analytical categories in FrameNet Brasil. *Veredas* 17. 44–65. https://periodicos.ufjf.br/index.php/veredas/article/view/25403.

Timponi Torrent, Tiago, Michael Ellsworth, Collin Baker & Ely Edison da Silva Matos. 2018. The Multilingual FrameNet shared annotation task: A preliminary report. In *Multilingual FrameNets and constructions, The international FrameNet workshop 2018.*

Timponi Torrent, Tiago, Ely Edison Da Silva Matos, Frederico Belcavello, Marcelo Viridiano, Maucha Andrade Gamonal, Alexandre Diniz Da Costa & Mateus Coutinho Marim. 2022. Representing context in FrameNet: A multidimensional, multimodal approach. *Frontiers in Psychology* 13. 1–20. DOI: 10.3389/fpsyg.2022.838441.

Timponi Torrent, Tiago, Maria Margarida M. Salomão, Fernanda C. A. Campos, Regina M. M. Braga, Ely E. S. Matos, Maucha A. Gamonal, Julia A. Gonçalves, Bruno C. P. Souza, Daniela S. Gomes & Simone R. Peron. 2014. Copa 2014 FrameNet Brasil: a frame-based trilingual electronic dictionary for the Football World Cup. In *Proceedings of COLING 2014, the 25th international conference on Computational Linguistics: System demonstrations*, 10–14. Dublin, Ireland: Dublin City University & Association for Computational Linguistics.

Venturi, Giulia, Alessandro Lenci, Simonetta Montemagni, Eva Maria Vecchi, Maria-Teresa Sagri, Daniela Tiscornia & Tommaso Agnoloni. 2009. Towards a FrameNet resource for the legal domain. In Núria Casellas, Enrico Francesconi, Rinke Hoekstra & Simonetta Montemagni (eds.), *Proceedings of the 3rd workshop on Legal Ontologies and Artificial Intelligence Techniques, held in conjunction with the 2nd workshop on Semantic Processing of Legal Text* (CEUR Work-

shop Proceedings 465), 67–76. Barcelona, Spain. https://ceur-ws.org/Vol-465/paper8.pdf.

Villavicencio, Aline, Ann Copestake, Benjamin Waldron & Fabre Lambeau. 2004. Lexical encoding of MWEs. In *Proceedings of the Workshop on Multiword Expressions: Integrating processing*, 80–87. Barcelona, Spain: Association for Computational Linguistics.

Virk, Shafqat Mumtaz, Dana Dannélls, Lars Borin & Markus Forsberg. 2021. A data-driven semi-automatic framenet development methodology. In *Proceedings of the international conference on Recent Advances in Natural Language Processing (RANLP 2021)*, 1471–1479. Held Online: INCOMA.

Yimam, Seid Muhie, Iryna Gurevych, Richard Eckart de Castilho & Chris Biemann. 2013. WebAnno: A flexible, web-based and visually supported system for distributed annotations. In *Proceedings of the 51st annual meeting of the Association for Computational Linguistics: System demonstrations*, 1–6. Sofia, Bulgaria: Association for Computational Linguistics.

Yong, Zheng Xin, Patrick D. Watson, Tiago Timponi Torrent, Oliver Czulo & Collin Baker. 2022. Frame shift prediction. In *Proceedings of the thirteenth Language Resources and Evaluation Conference (LREC'13)*, 976–986. Marseille, France: European Language Resources Association (ELRA).

You, Liping & Kaiying Liu. 2005. Building Chinese FrameNet database. In *2005 International Conference on Natural Language Processing and Knowledge Engineering*, 301–306. Wuhan, China: IEEE.

Zaninello, Andrea & Malvina Nissim. 2010. Creation of lexical resources for a characterisation of multiword expressions in Italian. In *Proceedings of the seventh international conference on Language Resources and Evaluation (LREC'10)*, 654–661. Valletta, Malta: European Language Resources Association (ELRA).