

# Greek within the Global FrameNet Initiative: Challenges and Conclusions so far

**Voula Giouli<sup>1,2</sup>, Vera Pilitsidou<sup>1</sup>, Hephæstion Christopoulos<sup>1</sup>**

<sup>1</sup>National and Kapodistrian University of Athens, <sup>2</sup>Institute for Language & Speech Processing, ATHENA RC

Address 1, Dragatsaniou 4, 105 59 Athens, Greece

Address 2, Artemidos 6 & Epidavrou, Maroussi, Greece

voula@athenarc.gr, verapilitsidou@gmail.com, hchristo@turkmas.uoa.gr

## Abstract

Large coverage lexical resources that bear deep linguistic information have always been considered useful for many natural language processing (NLP) applications including Machine Translation (MT). In this respect, Frame-based resources have been developed for many languages following Frame Semantics and the Berkeley FrameNet project. However, to a great extent, all those efforts have been kept fragmented. Consequentially, the Global FrameNet initiative has been conceived of as a joint effort to bring together FrameNets in different languages. The proposed paper is aimed at describing ongoing work towards developing the Greek (EL) counterpart of the Global FrameNet and our efforts to contribute to the Shared Annotation Task. In the paper, we will elaborate on the annotation methodology employed, the current status and progress made so far, as well as the problems raised during annotation.

**Keywords:** frame semantics, FrameNet, corpus annotation, lexical resource, Greek

## 1. Introduction

Over the last decades, a number of frame-based lexical resources have been developed based on the Berkeley FrameNet project (Baker et al., 1998) for languages other than English. In this context, the challenge has always been the alignment of frames across languages. In this paper, we describe on-going work carried out in the framework of preparing the Greek component of the Global FrameNet (FrameNet-EL). The purpose of our work is two-fold: (a) participation in the Multi-lingual FrameNet shared task, and (b) the development of language resources, i.e., a general-purpose lexical resource and an annotated corpus that will be applicable for a number of applications. The paper is organized as follows: In section (2), we provide the background and objectives of our work; our corpus data is described in section (3). The methodology adopted towards developing the language resources is presented in section (4), whereas, some preliminary results as well as issues and problematic cases that we faced throughout the various stages of our work so far are presented in section (5). Finally, our conclusions and prospects for future research are outlined in section (6).

## 2. Background and Objectives

According to Charles J. Fillmore’s Frame Semantics (Fillmore, 1977, 1982, 1985), there is continuity between language and experience (Petrucci, 1996). In this context, words gain their meaning in a semantic frame which can be an event or a relation. The term “semantic frame” or “frame” refers to any system of meanings which are connected in a way that, to understand any one of these meanings, we must be able to understand the whole structure to which it belongs (Fillmore, 1982: 111). Fillmore calls the elements of such a structure “Frame Elements” (FEs) and the words that evoke the semantic frames “Lexical Units” of the frame (LUs).

FrameNet, the lexical database for the English language for general purposes (Baker et al., 1998), was developed at the University of Berkeley in California based on the

mentioned theory. Over the years, a number of frame-based language resources have been developed for various languages (FrameNet Brazil (Salomão, 2009), Spanish FrameNet (Subirats, 2009) and Japanese FrameNet (Ohara, 2009), and the Swedish FrameNet++ (Ahlberg et al., 2014), inter alia). In this context, the Global FrameNet project (Torrent et al., 2018) has evolved, in order to examine, for example, to what extent the semantic frames developed for English are appropriate for other languages, whether some frames are universal and whether there are certain semantic domains in which frames tend to vary more across languages, or whether there are regular patterns of differences based on language families, regional groupings, etc.

As far as Modern Greek (MG) is concerned, there has been previous work in language for specific purposes and in language for general purposes – yet these studies remain fragmented and limited in scope. In fact, an initial attempt to build a frame semantics lexical resource for MG is reported in Gotsouli et al. (2007). However, this initial work was conceived of as the preliminary phase of a pilot project for the development of the basic infrastructure and design of the actual resource. Later, Dalpanagioti (2012) followed a frame-driven approach to the bilingual lexicographic process for creating a bilingual lexical database of motion verbs for EL and EN. Another attempt was made by Pilitsidou (2018), who used the FrameNet and Frame Semantics approach to create a domain-specific bilingual terminological database in EL and EN for the financial domain based on corpus evidence; the outcome of this work is a bilingual lexical resource in electronic format consisting of financial terms (LUs) of EL and EN, which are described and defined through the semantic frames that they evoke and the semantic relations, as well as a fully annotated corpus in various levels.

This paper reports on our ongoing participation in the Shared Annotation Task and the contribution to the overall objectives collaborating with teams from other languages towards developing a database of alignments of frames and FEs across languages. Therefore, the paper is

aimed at describing the progress made so far as well as the various issues and challenges we faced while working on the EL component of the Global FrameNet project. Effort has also been made to detect and categorize the differences spotted between the MG and English language. In the long run, our objective is to create a frame-based lexical resource for the MG language and to integrate it into existing semantic lexica.

From another perspective, one of our objectives is to examine whether the alignment of a Greek FrameNet with similar resources is feasible and whether the aligned lexica can be utilized for the translation process. Through the efforts for the creation of FN in other languages, the fact that frames are to an extent universal can be proven. As the lexical resource FrameNet can be machine-readable, it has the potential to be a very useful approach for assisting translation. In fact, FN’s semantic organization makes it cross-lingual, as different societies are highly likely to recognize the same types of events (Tantos et. al., 2015: 168). A database like FN offers a very useful tool for distinguishing subtle, yet crucial, differentiations in meaning, in a way that differs from other lexical resources, thus rendering it a very promising tool for the translation process.

### 3. Corpus Description

According to the global guidelines, annotation at this stage was performed on the transcribed TED Talk “Do schools kill creativity?” (Robinson 2006) and the subtitles provided for a number of languages; we report here on the annotation of the Greek counterpart of the TED talk. The EL corpus comprises 251 sentences and 3012 tokens. We pre-processed the raw text at various levels of linguistic analysis (Part-of-Speech tagging and lemmatization, syntactic annotation) using UDPipe annotation platform (Straka & Starková, 2017).

### 4. Annotation Methodology

The task of annotation was viewed as a two-stage procedure: (a) creation of the LUs (or lexical annotation), and (b) annotation of the corpus using the LUs already created and extending or modifying them where needed. Both tasks, that is, LU creation and corpus annotation, were performed by two annotators via the dedicated MLFN WebTool (Matos & Torrent, 2016). Blind annotation of FEs, GFs and PTs was performed separately by each annotator. At planned intervals, comparisons of the annotated data revealed discrepancies which were extensively discussed and resolved so as to reach a shared understanding of the task at hand and produce an initial version of the annotated text that has been thoroughly checked for mistakes or inconsistencies. At this stage, difficult or ambiguous cases were identified and accounted for. A step-by-step description of the procedure followed is provided below.

#### 4.1. LU Creation

The LUs to be annotated in the corpus were initially created in order to make the annotation process easier; in this respect, the approach we adopted was purely lexicographic in the sense that we first extracted all the lemmas from the EL text and then assigned them a frame

on the basis of their semantics. As expected, this agnostic procedure yielded different LUs for polysemous lemmas. In these cases, word sense discrimination was aided by existing reference works (monolingual and bilingual dictionaries) and corpus evidence in order to decide about the number (and types) of senses. At the next stage, selection of the appropriate frame each LU evokes was challenging. Following the global guidelines provided by the shared task organizers (Torrent et al, 2018), we adopted the frames as defined in the 1.7 release of the Berkeley FrameNet data (BFN 1.7). Since we were not allowed to make any changes, we tried to identify the frame each LU evokes through extensive search in the BFN 1.7 in order to locate the most appropriate one taking also into account its best translational equivalent(s) in English.

In case no translational equivalent of a Greek LU in English has been created yet in the BFN 1.7, the selection of the appropriate frame was performed by annotators using the following decision tree:

- Firstly, option (A) was to search in the BFN 1.7 data for a morphologically related LU that belonged to a different grammatical category (Part-of-Speech) – yet, it retained the meaning of the word to be annotated. In case an LU was spotted, we adopted the frame assigned to it. For example, the LU *αλήθεια.n* (truth) was created under the frame *Correctness*, based on the adjective *αληθής.a* and its translational equivalent *true.a*.
- If search (A) failed, we proceeded to option (B), that is, we checked if we could locate a synonymous word. In case a synonym was found in the BFN 1.7 data, we adopted its frame, as in the case of the adjective *ταλαντούχος.a* (talented); since its translational equivalent *talented.a* was not listed in the BFN 1.7, its synonymous word *skilled.a* led us adopt the frame *Expertise*.
- In both cases, (A) and (B), we also checked that the frame selected was a perfect fit, that is, it was actually evoked by the meaning of the LU, in that the latter (a) matches the underlying meaning of the frame and (b) features at least its core FEs.
- If both (A) and (B) failed, then we tried option (C), which entailed searching the list of existing frames and trying to identify a frame that would be the nearest match. In this case, we were expected to report the reason the frame was not considered as a perfect fit by selecting the most appropriate one from a list provided in the annotation tool: (a) different perspective, (b) different causative alternation, (c) different inchoative alternation, (d) different stative alternation, (e) too specific, (f) too generic, (g) different entailment, (h) different coreness status, (i) missing FE and (j) other.
- Finally, in case we were unable to locate an appropriate frame, we left the LU under consideration aside making a note for future reference. This is the case of the LU *βασίζομαι.v*

(to be based on), for which the frames of the morphologically related lemmas could not be adopted and no synonymous word could be found.

It should be pointed out, however, that during the annotation process we found out that certain LUs could be assigned to more than one frames, as for example the LU *δημιουργικότητα.n* (creativity); using option (A) we located the LU *creation.n* that evokes the frame *Intentionally\_create*. At the same time, an extensive search in the frames showed that the frame *Mental\_property* is also applicable.

Finally, following common lexicographic practices, for each LU, we provided sense description, in the form of a short lexicographic gloss in English.

## 4.2. Corpus Annotation

After the LUs had been created, annotation proper was performed. At this stage, each sentence in the corpus was annotated at the following layers: (a) Frame and Frame Element (FE) layer, (b) Grammatical Function (GF) layer, and (c) Phrase Type (PT) annotation<sup>1</sup>.

The major challenge was the identification of the correct LU already created (see section 4.1). Again, we also had to tackle polysemous lemmas by selecting the most appropriate LU to annotate. As a matter of fact, sense discrimination was often a challenge and fine distinctions between closely related frames made it difficult to spot the difference between them. In these cases, context was always helpful, especially in cases where the distinctions might be extremely fine. For example, the polysemous lemma *πηγαίνω.v* (to go) evokes more than one frames. However, as shown in example (1), given the context, the *Participation* frame has been selected based on the context – instead of the *Motion* one:

- (1) “Δεν πάτε συχνά σε πάρτι”  
Den pate sichna se parti  
Not go2.pr often to parties  
“You don’t go to parties often”

LUs were then populated with information regarding the FEs found in the corpus and their realizations. At the next level, each FE was assigned the grammatical function it assumes in the sentence. The following grammatical relations are foreseen: Noun Subject (Nsubj), Object (Obj), Indirect Object (Iobj), Clause Subject (Csubj), Clause Complement (Ccomp), Xcomp, Head, Dep, Nmod, Appositive, Ext. These relations were adopted from Universal Dependencies (UD)<sup>2</sup>. Subsequently, Phrase

<sup>1</sup> Two more layers are also foreseen by the Shared Task organizers, namely, Other and Sentence. The layer Other involves annotation of relative pronouns and their antecedents, whereas, the Sentence layer features tags applicable to the whole sentence, and may include notes such as the existence of a metaphor, or how prototypical the sentence is. For the time being, we did not perform any annotations at these layers.

<sup>2</sup> <https://universaldependencies.org/u/dep/index.html>

Types (PTs) chosen from UD tags were used to tag the realizations of FEs in the corpus. In the next section we will elaborate further on the results obtained, focusing on the creation of LUs and their annotation in the corpus.

In Figure 1, two examples of annotated sentences from the corpus are presented.

Είχαμε ΓΕΜΙΣΕΙ [GOAL το μέρος] [THEME με ατζέντηδες που φορούσαν μπλουζάκια] (Filling) (Implied AGENT: We)

Είχαμε γεμίσει το μέρος με ατζέντηδες [WEARER που] ΦΟΡΟΥΣΑΝ [CLOTHING μπλουζάκια] (Wearing)

Ithame gemisi to meros me agentides pou forousan blouzakia

Had1.pl filled the.acc place.acc with agents.acc who wore3.pl T-shirts.acc

“We had filled the place with agents wearing T-shirts”

[ADDRESSEE Σας] ΟΔΗΓΗΣΑΝ, πιθανότατα, [CONTENT μακριά από κάποια πράγματα στο σχολείο] όταν ήσασταν παιδιά, πράγματα που σας άρεσαν, [MEANS με τη δικαιολογία ότι δεν θα βρίσκατε ποτέ δουλειά κάνοντας αυτά], σωστά; (Talking\_into) (Implied SPEAKER: They)

Σας οδήγησαν, πιθανότατα, μακριά από κάποια πράγματα στο σχολείο όταν ήσασταν ΠΑΙΔΙΑ, πράγματα που σας άρεσαν, με τη δικαιολογία ότι δεν θα βρίσκατε ποτέ δουλειά κάνοντας αυτά, σωστά; (People\_by\_age)

Σας οδήγησαν, πιθανότατα, μακριά από κάποια πράγματα στο σχολείο όταν ήσασταν παιδιά, [CONTENT πράγματα που] [EXPERIENCER σας] ΑΡΕΣΑΝ, με τη δικαιολογία ότι δεν θα βρίσκατε ποτέ δουλειά κάνοντας αυτά, σωστά; (Experiencer\_focused\_emotion)

Σας οδήγησαν, πιθανότατα, μακριά από κάποια πράγματα στο σχολείο όταν ήσασταν παιδιά, πράγματα που σας άρεσαν, με τη ΔΙΚΑΙΟΛΟΓΙΑ [EXPLANATION ότι δεν θα βρίσκατε ποτέ δουλειά κάνοντας αυτά], σωστά; (Justification)

Σας οδήγησαν, πιθανότατα, μακριά από κάποια πράγματα στο σχολείο όταν ήσασταν παιδιά, πράγματα που σας άρεσαν, με τη δικαιολογία ότι δεν θα ΒΡΙΣΚΑΤΕ [TIME ποτέ] [THEME δουλειά] κάνοντας αυτά, σωστά; (Getting) (Implied RECIPIENT: You)

Σας οδήγησαν, πιθανότατα, μακριά από κάποια πράγματα στο σχολείο όταν ήσασταν παιδιά, πράγματα που σας άρεσαν, με τη δικαιολογία ότι δεν θα βρίσκατε ποτέ δουλειά ΚΑΝΟΝΤΑΣ [ACT αυτά], σωστά; (Intentionally\_act)

Sas odigisan, pithanotata, makria apo kapia pragmata sto scholio otan isastan pedia, pragmata pou sas aresan, me ti

dikaiologia oti den doulia kanontas afta,	tha sosta?	vriskate	pote
You.pl were led some.acc things.acc at the.acc school.acc were2.pl children, things.nom that you.gen were liked, with the.acc excuse.acc that not would find2.pl never job.acc doing them, correct?			
<p>“They most likely led you away from certain things at school when you were children, things you liked, with the excuse that you would never find a job doing those things, correct?”</p>			

Figure 1: Examples of annotations from within the corpus

## 5. Preliminary Results

This being a work in progress, presented below are our initial findings. After completing a certain part of the annotation, the team got together and discussed their findings. As will be shown below, there were several issues, some of which pertain to the nature and structure of the MG language as well as the translation of the text, which, more often than not, was not optimal, while others regard proposed additions or revisions of the FN frames and FE structure. We did encounter several cases of mistranslation or bad wording, some of which made it impossible to assign frames and FEs, while in other cases we had to make some uncommon decisions. Issues such as annotation of multiword expressions and grammatical differences between EN and EL will be discussed more thoroughly below.

### 5.1. LU Creation: Results and Issues

In total, c. 603 LUs were created that evoke c. 250 frames; regarding the verbs of the EL corpus, which are the main focus, more than about 220 frames have been assigned to the 167 unique verbs. In most cases, frame assignment via the EN LUs was a laborious – yet straightforward – task and the BFN 1.7 frames were proved a perfect fit, whereas, in a number of cases, no frame seemed to be a perfect fit. It should be noted that the already existing BFN 1.7 frames worked very well in almost all cases of commonly used phrases and words with a distinct and specific meaning, even in cases of polysemy where the word meanings were quite discrete.

Table 1 provides quantitative data on the frame assignment of the LUs we have taken into account so far. As one can see, the percentage of perfect fits is quite high (87.8%), as opposed to the 8.6% of non-perfect fits and the 3.7% percentage of the cases where no available frame could be found. However, it should be noted that, in order to achieve that satisfactory percentage, we often had to diverge from the frames the BFN assigns to certain LUs or make our own choices in cases of LUs that are not indexed. The main causes for cases of non-perfect fits were different perspective and different entailment, followed by too specific or too general frame, missing FE and different causative alternation.

No of existing LUs	626
No of LUs created	603
Perfect fits	549
Non-perfect fits	54
No frame assigned	23

Table 1: Quantitative results of frame assignment

A recurring issue that led us to sometimes taking unusual initiatives or resulted in the assignment of non-optimal frames are the cases of systematic polysemy, which is a phenomenon encountered across languages and should be considered by the FN team. Quite often, the FN catalogue seems to take into account only a certain shade of the occasional word’s meaning, and does not assign it to other frames that cover its different uses. This also led us to speculate that there might be some important frames missing from the FN catalogue. FN in general sometimes does not seem to distinguish between subtle differentiations in meaning, and there are words that in certain contexts could easily be assigned to a frame, but there are occasions where a perfect fit is impossible to find. For example, LUs such as *πανεπιστήμιο.η* (university) and *σχολείο.η* (school) fall in this category. According to the FN index, these LUs evoke the frame *Locale\_by\_use*, but this is merely one of their meanings. This is a classic case of systematic polysemy, as the words do not only denote the building itself, but also the institution and the activities that take place there.

There were also other cases where the FN frame was more or less a good fit, but we did notice some missing FEs that would be useful in MG and perhaps other languages (sometimes in English, as well) or cases where the description of the frame contained FEs we consider redundant or too specific. For example, the LU *παίζω.ν* (to play) evokes the frame *Competition*; however, a game is not always competitive, and this is not captured in the frame – or any other frame in BFN 1.7.

Another example is *χορεύω.ν* (to dance). FN lists *dance* under *Self\_motion*, but the definition of the frame is: “The SELF\_MOVER, a living being, moves under its own direction along a PATH. Alternatively, or in addition to PATH, an AREA, DIRECTION, SOURCE, or GOAL for the movement may be mentioned.” There is clearly no necessity for a PATH or DIRECTION when someone is dancing. Another possible choice, not mentioned in the English FN, would be *Moving\_in\_place*, which is sometimes true for this specific activity and sometimes not. Perhaps a more suitable frame would be one referring to pastime activities.

A similar case is “Σε λίγο τα πτυχία δεν θα αξίζουν τίποτα.” A not very elegant, but closer to the EL text back translation would be “In a while, the degrees will be worth nothing.” The LU *αξιζώ.ν* (to be worth) could be assigned to the *Deserving* frame. The frame’s definition according to FN is “The existence of a STATE\_OF\_AFFAIRS is sufficient reason for taking an ACTION. The agent who is justified in taking the

suggested ACTION is not part of the immediate scene, however.” Based on the phrase above, we believe that the ACTION is an optional element in the frame, since there is no reference to what ACTION would be taken if the STATE\_OF\_AFFAIRS (the degrees) were sufficient.

(2) “Σε λίγο τα πτυχία δεν θα αξίζουν τίποτα”  
 Se ligo ta pthia den  
 tha axizoun tipota  
 In a little the.pl.nom degrees.nom not  
 will be worth nothing  
 “In a while the degrees will be worthless”

Another issue that caught our attention is that sometimes FN does not seem to distinguish between the procedure that leads to a certain result and the case where an LU denotes being in that certain state from the beginning. At least in EL, there are some LUs, mainly verbs, that could assume both meanings, depending on context. This distinction might appear too fine at first sight, it is however frame-defining, and we did notice the lack of available frames in such cases. This is the case of the LUs *ενώνω.v* (to connect, to join) and *χωρίζομαι.v* (to be separated, to be divided). Here are two similar cases: the LU *ενώνω.v* is not in this example a perfect fit for the *Attaching* frame, as shown in (3), since the frame refers to the process of joining, not the state of being joined.

(3) “Ενώνει τα δύο μισά του εγκεφάλου”  
 Enoni ta dyo misa tou  
 egefalou  
 Connects the.sg.acc two halves.acc the.sg.gen  
 brain.gen  
 “Connects the two halves of the brain”

On the other hand, the LU *χωρίζομαι.v* is not a perfect fit for *Becoming\_separated* as shown in example (4), as the phrase refers to the state of being separated.

(4) “Δεν χωρίζεται σε διαμερίσματα”  
 Den horizete se diamerismata  
 Not divided3.sg.pass into compartments.acc  
 “It is not divided into compartments”

However, not all difficulties we encountered should be attributed to shortages in the FN index. Some issues arise from peculiarities and idiosyncrasies of the MG language, such as the middle voice, which are to be expected, since the BFN we take as a starting point was originally designed for the English language. Generally, passive (or non-active) morphology of some EL verbs is found in reflexive, anti-causative and passive structures. However, in some cases, the passive morphology of some verbs signifies usages/senses besides those entailed by the active morphology (Clairis & Babinotis, 2005). These differences in meaning cannot be accounted for at early stages of processing; as a result, when attempting to assign an EL verb to a frame, the annotation tool treats the active and the middle voice of verbs as a single lemma. As expected, this presents a problem in several cases, as the active and middle forms of a verb might belong to a different frame. For example, the LU *εμφανίζω.v* (to reveal, to present) in active voice needs to be assigned to

different frame as opposed to its middle voice counterpart *εμφανίζομαι.v* (to appear or arrive). The middle voice of the verb can be assigned to frames such as *Arriving* or *Becoming\_visible*, while the active voice *εμφανίζω.v* most certainly does not belong there. Similarly, the LU *ωφελούμαι.v* (to benefit from), which is middle voice in MG, and its active voice *ωφελώ.v* (which means *benefit* as in “These new courses will benefit the students”) need to be indexed under different frames. The frame *Cause\_benefit\_or\_detriment* certainly is not the best fit for the middle voice, while other frames only tangentially relate to the verb’s meaning. This could imply that there’s a missing frame in the FN catalogue, as we could not find one suitable to the middle voice form.

Similarly, non-perfect fits showcase differences in perspective between MG concepts and English ones, as shown with the verb *συνταξιοδοτώ.v*. In MG, the verb has both active and passive morphology, whereas the respective verb in English “to retire” corresponds to the passive voice; moreover, in MG, the verb has more specific connotations, as it means to leave one’s job and get a pension, the pension being the core component of the verb’s meaning.

The verbs *επιτρέπεται* (to be allowed) and *πρόκειται* (a rough translation would be “will” or “be about to”) are analogous to *αρέσω.v* (discussed in 5.2). Regarding the first case, we encounter the phrase “Τα παιδιά χορεύουν όλη την ώρα, αν αυτό τους επιτρέπεται,” which can be back translated as “Children dance all the time, if they are allowed to.” This is the middle voice of *επιτρέπω.v* (to allow), and apart from the problems with frame assignment, there are some peculiarities in its use. A more literal translation of the phrase “αν αυτό τους επιτρέπεται” would be “if this is allowed to them” (consequently, it is not a perfect fit for *Preventing\_or\_letting* or *Deny\_or\_grant\_permission*). A peculiarity of this verb, however, is that the middle voice of the verb appears only in the third person singular or plural, meaning that a certain act is allowed to some entity. Moreover, it should be noted that we couldn’t find a suitable FE in the *Preventing\_or\_letting* frame denoting who is allowed or prevented from doing something. Furthermore, in MG the entity that is allowed to do something is realized as the object of the verb in the genitive case or as a complement of the preposition *σε* (e.g. *σε εμένα*, meaning “to me”), not as the subject, as is the case with the English verb *to be allowed*. The verb’s antonym *απαγορεύεται.v* has the exact same properties as the ones just discussed.

(5) “Αν αυτό τους επιτρέπεται”  
 An afto tous epitrepete  
 If this.sg.nom they.gen is allowed  
 “If they are allowed to.”

Regarding *πρόκειται.v*, here we have an even more noteworthy case. It is a middle voice verb that is used solely in the third person singular, and more often than not preceded by the negative particle *δεν* (not). An instance from our corpus is “δεν πρόκειται να γίνει μουσικός”

(“you will never become a musician”). We assigned the verb to the *Destiny* frame, even though we are not very satisfied with that choice, since the sense of fate is not always implied by this verb. What is interesting is that the core FEs (PROTAGONIST, ROLE, STATE\_OF\_AFFAIRS) are not realized as complements to the verb but as a subjunctive subordinate clause:

(6) “Δεν πρόκειται να γίνεις μουσικός”  
 Den prokrite na ginis mousikos  
 Not will3.sg to become2.sg.sbjv musician  
 “You will never become a musician”

Furthermore, a number of ambiguous cases were identified. For example, the LU *κάνω.v* (to do or to make) was assigned to the frame *Intentionally\_act*; however, this is not always the case. In some cases, there might be a different entailment, since doing something does not always imply intention to do it. In this respect, the frame is too specific.

Despite the variety of frames and LUs the BFN 1.7 offers, we encountered several instances where the FN-assigned frames do not cover all cases of the EL lemmas. One example is the EL noun *ιδέα.n* (idea: “να έχεις πρωτότυπες ιδέες” – “having original ideas”). The FN-assigned frame is *Awareness*. However, the particular meaning of the word is not covered by this frame. A possible candidate could be *Coming\_up\_with*, even though this frame is perhaps a bit too specific. This is a case where not all meanings of a word are covered. Similarly, the LU *ανταποδίδω.v* (to reciprocate or to return) in (7) was listed under the *Request* frame, although in this specific instance it does not mean inviting someone but returning a favour instead. *Reciprocate* is not indexed in FN and *return* is indexed only under different meanings.

(7) “Δεν ανταποδίδουν την πρόσκληση”  
 Den antapodidoun tin prosklisi  
 Not return3.pl the.sg.acc invitation sg.acc  
 “They don’t invite you back”

Another instance was *χάνω.v* (in our case, to miss, but also to lose in other contexts): “Έχασα κάτι;” (“Did I miss something?”). We found it impossible to assign a frame to this meaning of the word. A possible candidate could be *Perception*, but *miss* does not imply modalities like hear or taste.

The LU *βρίσκω.v* (to find, to get) of the phrase “Δεν θα βρίσκατε ποτέ δουλειά” (“You would never find a job”) posed a difficulty as well. The *Getting* frame implies the acquisition of an object or some property and the change of ownership. This is not the case here. On the other hand, the frame *Being\_employed*, which would refer to the whole phrase and not just the verb (see cases of multiword expressions below), is not suitable either, because it does not refer to the process of acquiring a job.

(8) “Δεν θα βρίσκατε ποτέ δουλειά”  
 Den tha vriskate pote doulia  
 Not will find3.pl.pret never job.acc  
 “You would never find a job”

Last but not least, the LU *εισαγωγή.n* (in this case, university admission, but entrance or insertion in general – “είναι μια παρατεταμένη διαδικασία εισαγωγής στο πανεπιστήμιο”/“it is a prolonged university admission procedure”) was also problematic. The word in this context cannot be assigned to *Arriving*, since the frame refers to a literal arrival at a place, but neither to *Success\_or\_failure*, as it is too specific. This is a case of a subtle differentiation in meaning which makes it difficult to find a suitable frame

## 5.2. Corpus Annotation: Results and Issues

In total, 222 out of the 251 sentences of the text were annotated, whereas, the annotation effort amounts to 620 annotation sets for verbs, nouns, adjectives, adverbs and numbers. The distribution of the annotated LUs per Part-of-Speech (POS) is depicted in Table 2.

POS	No
adjective	32
adverb	3
noun	157
num	7
verb	421
total	620

Table 2: Distributions of LUs per POS

It should be noted that MG is a pro-drop language, and consequently several core elements, such as the *AGENT* or the *COGNIZER*, appear to be missing from the sentences, while in reality they can be inferred from the verb form. This is a special case of constructional null instantiation that one can come across very frequently in languages that feature this syntactic characteristic, such as Greek, Spanish or the Slavic languages. For this reason, it is not possible to annotate all core FEs in the corpus, unless the annotation platform is modified as appropriate so as to take the peculiarities of pro-drop languages into account, since the labelling Null does not allow the annotator to define the FE that appears to be missing.

Moreover, the annotation process posed challenges due to the genre of the text; as a result, some sentences were not annotated at all since they present phenomena like ellipsis or pragmatic function. This is especially true for questions like the one presented in (9):

(9) “Τι έγινε;”  
 Ti egine?  
 What became?  
 “What happened?”

This meaning of the LU *γίνομαι.v* occurs only in the third person singular, which is a distinctive quality that should be noted.

From another perspective, the annotation at the GF and PT levels revealed further discrepancies and non-perfect fits. More precisely, we did notice some differences in the realization of the FEs in EN and EL that are worth pointing out; these could either pertain to the frame assignment itself or to differences between the structure and syntax of FEs within a given frame. An example of a

LU as compared to its translation in EN is depicted in Table 3. Following its EN counterpart, the EL LU *αρέσω.v* (to like) was created under the frame *Experiencer\_focus*. However, this was proven to be a non-perfect fit, and was chosen only because no other frame seemed appropriate. The main difference between the verb *to like* and the EL verb *αρέσω.v* is that in English the EXPERIENCER is always realized as the Subject of the verb; in EL, however, the EXPERIENCER is realized either as the complement of the preposition or as the object complement in genitive case. Moreover, in a more general use/meaning, the verb can be used without a complement at all. Consequently, the CONTENT rather than the EXPERIENCER seem to be the focus of the verb; the EXPERIENCER might not even be present in the sentence, as will be shown in the examples below. The following table depicts the different realizations of the EN and EL verbs *αρέσω.v/to like.v*:

Experiencer focus		
	Realization	
	like.v	αρέσω.v
EXPERIENCER	Ext.NP	Obj.NP
CONTENT	Obj.NP	Nsubj

Table 3: Realization of the LUs *to like.v* and *αρέσω.v*

Usages of the verb *αρέσω* are depicted in examples (10) and (11) below:

(10) “Πράγματα που σας άρεσαν”  
 Pragmata pou sas aresan  
 Things.nom that you.pl.gen liked  
 “Things you liked”

(11) “Ο Γιάννης αρέσει”  
 O Gianis aresi  
 The.sg.nom Gianis.nom is liked  
 “Giannis is liked” (meaning, by people in general – note the absence of the EXPERIENCER)

Finally, we should note again that there were a great many instances of mistranslation or bad wording in the EL text, which made the frame assignment very difficult or even impossible in certain cases. One such case, maybe the most characteristic one, is the phrase, “Αυτοί οι άνθρωποι που βγαίνουν από την κορυφή” (“These people who come out of the top”), which makes no sense and it was impossible to infer what the translator meant by it.

### 5.2.1. Multiword Expressions

Multiword expressions (MWEs) have long been regarded as a “pain in the neck” for NLP and translation alike, due to their idiosyncratic behaviour (Sag et al., 2002). In fact, they are lexical items characterized by lexical, syntactic, semantic, pragmatic or statistical idiosyncrasies. We did encounter such cases in the corpus which form solid semantic unities and cannot be treated on a word-by-word basis. Some cases are debatable; for example, collocations

such as the noun phrases *καθηγητής πανεπιστημίου* (university professor) and *εκπαιδευτικό σύστημα* (educational system) could be either regarded as two distinct words or as a homogenous whole, as, for example, a university professor is a distinct vocation compared to, e.g., *καθηγητής αγγλικών.n* (English teacher), which both use the equivalent of the word professor in a totally different context. However, even if it would perhaps be preferable to assign these phrases to a single frame as a whole, it is quite straightforward to frame them word by word.

But not all cases are that simple. As a matter of fact, a number of idiomatic expressions found in the corpus can only be treated as single predicates. For example, the verbal MWE *δεν μας παίρνει* (we can’t afford to) must be assigned as a whole to the *Capability* frame. However, as it is an idiomatic phrase, a word-by-word translation would be “it doesn’t take us.” Clearly there is no point in assigning the LU *take.v* to the *Taking* frame in this instance. This is also the case with a number of Light Verb constructions.

A fact that should be taken into account should the FN annotation platform make it possible to assign frames to MWEs is that MWEs are often discontinuous, as is often the case in MG. For example, the expression in (12), belonging to the *Attempt\_suasion* frame, consists of fixed discontinuous elements and non-fixed ones:

(12) “Κάνε μου τη χάρη.”  
 Kane mou ti hari  
 Do I.gen the.sg.acc favour.acc  
 “Do me this favour”, also meaning “indulge me”

The expression is “κάνω τη χάρη” (do the favour), and the pronoun can be interposed in between, disrupting the continuity of the phrase.

## 6. Conclusion

We have presented work in progress towards developing the Greek section of the Global FrameNet Shared Task. In an attempt to prove the universal nature of frames, effort has been made to construct a frame-based lexical resource for MG and to annotate an EL corpus based on frames that already exist for the English language. This task has not always been an easy and straightforward one. In the paper we have reported on the progress made so far, and on the issues encountered. Future work is already planned towards enriching the EL data with new corpora and annotations and towards using the resource for aiding the translation process. In particular, a future prospect is to add comparable corpora to the data, in order to extend the lexical resource and avoid any inconsistencies that emerge from mistranslation or wrong wording of the translated corpus. As a matter of fact, the need of adding more frames or more FEs to FN, with which it would be possible to include the differentiated meanings of LUs of the MG language, has emerged, so that, in the future, the

database can be used for the MG language. From another perspective, further work is planned towards making meaningful cross-lingual comparisons.

## 7. Acknowledgments

The authors would like to thank the anonymous reviewers and the editors for their valuable comments to the manuscript that contributed to improving the final version of the paper. The research leading to the results presented here was partially supported by the Translation and Interpreting MA programme of the Faculty of Turkish and Modern Asian Studies of the National and Kapodistrian University of Athens.

## 8. Bibliographical References

- Ahlberg, M., Borin, L., Dannélls, D., Forsberg, M., Toporowska Gronostaj, M., Friberg Heppin, K., Johansson, R., Kokkinakis, D., Olsson, L.J., and Uppström, J. (2014): Swedish FrameNet++ The Beginning of the End and the End of the Beginning, in Proceedings of the Fifth Swedish Language Technology Conference, Uppsala, 13-14 November 2014.
- Baker C. F., Fillmore C. J., and Lowe J. B. (1998). The Berkeley FrameNet Project. Proceedings of the 17th International Conference on Computational Linguistics – Volume 1 (COLING '98), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA: 86-90.  
<https://dl.acm.org/citation.cfm?doid=980451.980860>
- Clairis, Chr., and Babiniotis, G. (2005). Grammar of the Modern Greek Language (In Greek). Athens: Ellinika Grammata.
- Dalpanagioti, T. (2012). A Frame-driven Approach to Bilingual Lexicography: Lexical Meaning and Usage Patterns in Greek and English Motion Verbs. PhD Thesis. Εθνικό αρχείο διδακτορικών διατριβών.  
<https://www.didaktorika.gr/eadd/handle/10442/27162>
- Fillmore, C. J. (1977). Scenes-and-frames Semantics. Book chapter in: Linguistic Structures Processing. Fundamental Studies in Computer Science, vol. 59. North Holland Publishing: 55-81.
- Fillmore, C. J. (1982). Frame Semantics. Linguistics in the Morning Calm, ed. The Linguistic Society of Korea. Selected papers from SICOL-1981: 111-137.
- Fillmore, C. J. (1885). Frames and the Semantics of Understanding. *Quaderni di semantica*. Vol. 6, no 2: 222-254.
- Gotsoulia, V., Desipri, E., Koutsombogera, M., Prokopidis, P., Papageorgiou, H., and Markopoulos, G. 2007. Towards a Frame Semantics Lexical Resource for Greek. In Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories.
- Matos, E. and Torrent, T. (2016). A Flexible Tool for an Enriched FrameNet. In Proceedings of the 9th International Conference on Construction Grammar (ICCG9), Juiz de Fora, Brazil, october. Federal University of Juiz de Fora (UFJF).
- Ohara, K. H. (2009). Frame-based Contrastive Lexical Semantics in Japanese FrameNet: The Case of Risk and Kakeru: 163–182. Berlin/New York: Mouton de Gruyter.
- Petruck, M. R. L. (1996). Handbook of Pragmatics. Jef Verschueren, Jan-Ola Östman, Jan Blommaert, and Chris Bulcaen (eds.). Philadelphia: John Benjamins.
- Pilitsidou, V. (2018). Terminology: Application of Frame Semantics to the Thematic Field of Finance for the Creation of a Term Base, from a Translation Perspective. (Master's thesis, National and Kapodistrian University of Athens, Athens, Greece). Retrieved from <https://pergamon.lib.uoa.gr/uoa/dl/object/2775104>
- Robinson, K. (2006). Do Schools Kill Creativity? TED Talk. Available at [https://www.ted.com/talks/ken\\_robinson\\_says\\_schools\\_kill\\_creativity](https://www.ted.com/talks/ken_robinson_says_schools_kill_creativity).
- Sag, I., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002). Multiword Expressions: A Pain in the Neck for NLP. In A. F. Gelbukh, ed. 2002. Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing (CICLing '02). Springer-Verlag, London, UK, UK, 1-15.
- Salomão, M. M. M. (2009). Framenet brasil: um trabalho em progresso. *Caleidoscópio* 7(3): 171–182.
- Straka, M. and Starková, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. Association for Computational Linguistics: 8899. Vancouver, Canada. <http://www.aclweb.org/anthology/K/K17/K17-3009.pdf>
- Subirats, C. (2009). Spanish FrameNet: A Frame-semantic Analysis of the Spanish Lexicon: 135–162. Berlin/New York: Mouton de Gruyter.
- Tantos, A., Markantonatou, S., Anastasiadi-Symeonidi, A., and Kyriakopoulou, T. (2015). Computational Linguistics (in Greek). [E-book] Athens: Hellenic Academic Libraries Link.  
<https://repository.kallipos.gr/bitstream/11419/2205/13/nlp2ndEdition.pdf>
- Torrent, T., Elsworth, M., Baker, C., da Silva Matos, E. E. (2018). The Multilingual FrameNet Shared Annotation Task: a Preliminary Report. In The International FrameNet Workshop 2018: Multilingual FrameNets and Constructions.

## 9. Language Resource References

- FrameNet: <https://framenet.icsi.berkeley.edu/fndrupal/>  
Glosbe: <https://el.glosbe.com/>  
Merriam-Webster: <https://www.merriam-webster.com/>  
Oxford Dictionaries: <https://www.oxforddictionaries.com/>  
Portal for the Greek Language: <http://www.greek-language.gr/greekLang/index.html>  
WordReference: <http://www.wordreference.com/>