



## **A Guide to IDION**

IDION is a lexicographic environment for the documentation of Greek idioms. The technical term “Multi Word Expressions (MWE)” (Πολυλεκτικές Εκφράσεις (ΠΛΕ)) is often used for this type of linguistic structure.

MWEs may either be nominal (e.g. «παιδική χαρά» = childish happiness = ‘playground’) or verb (e.g. «έκανα μαύρα μάτια να δω κάποιον» = I made black eyes to see someone = ‘I have not met someone for a long time’).

This is a brief guide to the usage of IDION. A detailed description of the lexicographic environment is given, along with instructions on how to edit its contents.

## Installation Instructions

### IDION installation

IDION application does not need to be installed.

Decompress the file [multiword-editor-app-1.0.0-SNAPSHOT.zip](#).

Then click on the file [mweditor](#).

### IDION execution

The application is executed by double-clicking the [mweditor.exe](#) in the file [bin](#).

### IDION updates

The application can be updated via the menu as follows:

Select [Help](#).

Then select [Check for Updates](#).

If there are updates available, click on [Next](#).

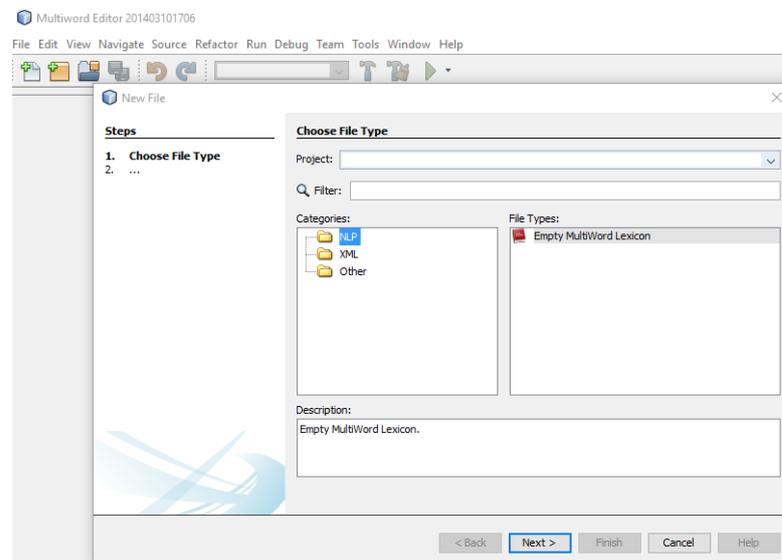
### Creation of a new lexicon

A “lexicon” is a collection of lemmas. Currently, IDION contains the lexicon “VERB\_MWEs” with nearly 300 verbal MWEs. Lexica are just files for IDION. To create a new lexicon, in other words to open a new file, the following steps are followed:

From [File](#) select [New File](#).

In the emerging window select [NLP](#) and the type [Empty MultiWord Lexicon](#). Then press [Next](#). In the emerging window name your file and specify the file location where it will be saved.

Finally, press [Finish](#).



Picture 1. Creation of a new lexicon.

### Opening a lexicon

From [File](#) select [Open File](#)

Choose the file you wish to edit/read.

### Saving a lexicon

From [File](#) select [Save](#) or [Save As](#)

### Adding a new lemma

To add a new lemma:

Right-click on the list in the left side of the screen.  
In the emerging menu select **Add**.

## General Remarks

1. Expressions are introduced in the **lemma form**: “κάνω την πάπια” and NOT in the form “Ο Χ κάνει την πάπια” or any other form. If the expression has a fixed subject and a variable pronominal object then the first person is selected for the object: “με κόβει η λόρδα”,
2. Synonymous MWEs with minimally different fixed constituents are encoded as individual lexical entries.

- (1) a. κόβω κάποιον λωρίδες  
cut somebody.ACC stripes.ACC  
b. κόβω κάποιον φέτες  
cut somebody.ACC slices.ACC  
“to kill somebody”

- (2) a. Πίνω το πικρό ποτήρι  
Drink the bitter.ACC glass.ACC  
“I experience a sad situation”  
b. Καταπίνω το πικρό ποτήρι  
Swallow the bitter.ACC glass.ACC  
“I experience a sad situation”

MWEs (1a) and (1b) are synonymous and have minimally different fixed constituents. Since the fixed parts are essentially different, we will consider them distinct MWEs and for each of the two MWEs a distinct entry will be encoded. **Encodings of the form “κόβω κάποιον λωρίδες/φέτες” are not allowed.** The synonymy relation has to be encoded with the RELATIONS tab (see point 2 below). Similarly, (2a) and (2b) are synonymous MWEs with different verbs and are encoded as individual lexical entries.

### Caution:

Differences among fixed constituents of the MWE concerning diminutives, specific adjectives, a definite/indefinite article, alternative prepositions, different form of adverb/preposition, additional constituents that modify the conveyed meaning (ανάβω **όλα** τα λαμπάκια κάποιου, πίνω το πικρό ποτήρι **ως τον πάτο**) are encoded as alternative forms (see also the FORMS tab).

3. **Distinct meanings of the same MWE form are encoded in separate entries.** (3a) means that someone was imprisoned, while (3b) that someone has lost money and is in debt.

- (3) a. Ήθελε φακελάκια αυτός ο γιατρός και (ΕΠΙΤΕΛΟΥΣ) τον βάλανε μέσα!!!<sup>1</sup>  
Wanted bribe this the doctor and (AT LAST) him put-V.3PL.PAST in  
“This doctor wanted a bribe and (AT LAST) he was imprisoned”

- b. Το σπίτι που έβαλε μέσα χοντρά τη Σάρα Τζέσικα Πάρκερ...<sup>2</sup>  
The house-NOM which put-V.3SG.PAST in a\_lot the Sarah-

<sup>1</sup> <http://www.athensmagazine.gr/portal/athenstalk/51087>

<sup>2</sup> [www.womenonly.gr/gallery.asp?catid=37522&subid=2&pubid](http://www.womenonly.gr/gallery.asp?catid=37522&subid=2&pubid)

ACC Jessica Parker

“This house costed Sarah Jessica Parker a lot of money, and now she is in debt”

4. Semantic relations among MWEs are encoded with the RELATIONS tab.

Relations among MWEs are encoded only in the RELATIONS tab. **Relations should be encoded carefully.** For instance, “τα παίρνω στο κρανίο” (‘I get very angry’) is not synonym with the MWE “κόβω τα φτερά κάποιου” (‘I dishearten someone’).

## **IDION Tabs**

IDION is organized in 7 sections (tabs). These are the following:

1. PREVIEW tab
2. GENERAL tab
3. FORMS tab
4. USE tab
5. DIAGNOSTICS tab
6. CORPUS tab
7. RELATIONS tab

Each tab is described in detail below.

## PREVIEW TAB:

The tab PREVIEW is **auto-generated**. It offers an overview of the information about the entry (Picture 2).



Picture 2. The PREVIEW tab.

## GENERAL TAB:

The meaning of the MWE is encoded in the GENERAL tab . There is also space to add some general comments/notes.

**English translation** of the examined MWE should also be provided. Translation in other languages can be provided.



Picture 3. Example of a translated MWE.

1. A free NP is denoted with the pronouns κάποιος, κάποια, κάποιο, κάτι,..., using the abbreviations κπ (sb)/κτ (sth). A PP with a free object is represented as P plus free NP (3a).

- (3) a. βγάζω την γλώσσα μου σε κπ  
Pull-V.1SG.PRES the-DET tongue-ACC my-POSS in-PREP  
someone-ACC  
“I tease someone”  
b. βγάζω την μάσκα κάποιου  
draw-V.1SG.PRES the-DET mask-ACC someone’s-GEN  
“I unmask/reveal someone”

2. Participation to the dative genitive alternation is not marked on the lemma.

## FORMS TAB:

Exhaustive morphological and syntactic description of the MWE is encoded with this tab. Encoding is theory neutral and is aimed to serve as a basis for any type of parser. To this end, morphological tags are standardized (PAROLE [http://nlp.ilsp.gr/nlp/tagset\\_examples/tagset\\_en/](http://nlp.ilsp.gr/nlp/tagset_examples/tagset_en/)). Similarly, the encoded syntactic relations are kept theory independent, therefore minimal, and include:

- Information about free constituents
  - Phrasal information for constituents that are realized with full phrasal structures
  - Lexical information for constituents that are realized with weak pronouns
- binding and control relations
- delineation of fixed/semi-fixed strings

A tabular arrangement with five columns is used together with controlled vocabularies that help to minimize the number of errors. The label 'tokens' is used to cover both lexical and phrasal parts of a MWE. Each token occupies a line in the tabular format. New tokens can be added and existing ones can be removed with the use of the corresponding buttons.

Variants of a MWE are encoded with additional forms in the FORMS tab. As variants of a MWE we encode cases where the following occur: a diminutive as an alternative, a different form of a lemma (eg. ανεβάζω κπ στα ουράνια/στα μεσουράνια), an alternation between a definite/indefinite determiner or additional constituents (drawn from a very small set of specified adjectives or PPs) that modify the semantics of the MWE (ανάβω **όλα** τα λαμπάκια κάποιου, πίνω το πικρό ποτήρι **ως τον πάτο**).

### Columns 1 and 2:

Free, semi-fixed and fixed parts of the MWE are encoded in the order that they appear in the MWE. **Binding relations** (picture 4) and **optionality of tokens** (picture 5) are also encoded in this area with the appropriate buttons that are provided under each token.

Tokens				
NP-NOM	Lemma:		WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By	
LEMMA	Lemma:	αφήνω	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By	
LEMMA	Lemma:	στο	WordForm:	στο
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By	
LEMMA	Lemma:	νόδι	WordForm:	νόδι
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By	
PnGe	Lemma:	μου	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By: i: NP-NOM	Controlled By	
NP-ACC-anim	Lemma:		WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By	

Picture 4. Binding relations depicted in the DB.

Tokens				
NP-NOM-anim	Lemma:	WordForm:		
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:	
LEMMA	Lemma: βγάζω	WordForm:	Vb	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:	
LEMMA	Lemma: ο	WordForm: τα	AtDfNePIAc	
<input checked="" type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:	
LEMMA	Lemma: φύχι	WordForm: φύχια	NoCmNePIAc	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:	
PnGe	Lemma: μου	WordForm:	PnPp	
<input checked="" type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By: 1: NP-NOM-anim	Controlled By:	

**Picture 5. Optionality of tokens in a MWE.**

The encoder works simultaneously with Columns 1 and 2. Data are introduced in Column 2 while Column 1 is used to provide the appropriate label from the controlled vocabulary and to indicate optionality where it applies. Column 2 is also used to indicate binding relations where they apply.

*The controlled vocabulary provided in Column 1 is:*

- o **LEMMA**: declinable
- o **WF (WORDFORM)**: non-declinable,
- o **COMPL**: designates the notion **COMPLEMENTIZER** (να, θα, που, ...)
- o **VP**: any controlled but otherwise free VP normally following a fixed complementizer
- o **NP-NOM**
- o **NP-NOM-anim**
- o **NP-NOM-nonanim**
- o **NP-GEN**
- o **NP-GEN-anim**
- o **NP-GEN-nonanim**
- o **NP-ACC**
- o **NP-ACC-anim**
- o **NP-ACC-nonanim**
- o **PnGe**

More precisely ( A-D):

- A. Semi-fixed and fixed parts are treated as individual words and not as phrasal constituents and are all listed in their **lemma** form:

(4) κάνω μαύρα μάτια να σε δω  
Do black eyes to see you  
“I haven’t seen you for a long time”

(4) contains two verbal heads, namely κάνω and the semi-fixed δω. Of them the first is free while the second has to be always in Past Subjunctive form and is controlled by the subject. ‘κάνω’ will be assigned the label LEMMA and in the fourth column it will be assigned only Part of Speech and active voice, which are the only constraints that apply to it (Picture 6).

LEMMA	Lemma: κόνω	WordForm:	VbAv
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By

**Picture 6. Example of a non-restricted verb in a MWE.**

‘δω’ will be assigned the label LEMMA but in the third and fourth column it will be assigned two constraints. In the third column it will be assigned the “controlled by” tag. In the fourth column it will be assigned constraints on morphology: VbMnIdPeAv (Picture 7).

LEMMA	Lemma: βλένω	WordForm: δω	VbMnIdPeAv	Select	WWS Index 2
<input checked="" type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By: 1: NP-NOM-anim		Remove

**Picture 7. Example of a controlled verb in a MWE.**

(5) μένω ταπί και ψύχραμος  
 stay broke and cool  
 “I am broke”

(5) contains the cranberry word ταπί. Ταπί will be assigned the label WF and in the fourth column it will not be assigned any tag at all (Picture 8).

WF	Lemma: τανι	WordForm: τανι
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By
		Controlled By

**Picture 8. The cranberry word ταπί as depicted in the FORMS Tab of the DB.**

In the same example (5), the adjective ψύχραμος that is controlled for number and gender by the free NP (the free subject of the MWE) will be assigned the label LEMMA in the second column and in the fourth column will not be assigned any constraint because all constraints are supplied by the control relation.

- B. Prepositional heads that are indeclinable by definition, are considered fixed. They may accept fixed (για ένα κομμάτι ψωμί), semi-fixed (στο νυχάκι NP-GEN-anim) or free nominal complements that are encoded accordingly.
- C. One or more strings of one or more fixed wordforms may be indeclinable within the MWE: ‘μαύρα μάτια’ (4) is an example of a fixed string.
- D. Free nominal parts of the MWE may be phrasal or lexical:
  - a. As lexical are treated the free possessive pronouns that are very often bound either by the subject or the object. They are assigned the lemma ‘μου’. The label attached to them is standard and it is the “bleached” Parole tag that includes only the part of speech label (Pn), the case label (Ge) and the indication ‘weak pronoun’ (We). The number and gender labels will be specified in parsing by binding mechanisms.
  - b. As phrasal are treated:
    - i. Free nominal constituents (often analyzed as free subjects, objects, prepositional complements and genitive NPs modifying nouns) that are labeled for their phrasal

category, case and animosity. The controlled vocabulary for this purpose is:

- **VP**
- **NP-NOM**
- **NP-NOM-anim**
- **NP-NOM-nonanim**
- **NP-GEN**
- **NP-GEN-anim**
- **NP-GEN-nonanim**
- **NP-ACC**
- **NP-ACC-anim**
- **NP-ACC-nonanim**

Labels without specification about animosity must be understood as the disjunctive form of the corresponding labels that specify animosity. Thus NP-ACC must be understood as NP-ACC-anim OR NP-ACC-nonanim. For instance, the MWE “τα βάζω με κάποιον/κάτι” has a free noun phrase (the object of the verb), which will be denoted with the term NP-ACC (Picture 9):

Tokens									
NP-NOM-anim	Lemma:	WordForm:	Select	WWS Index					
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By						Remove
LEMMA	Lemma: ο	WordForm: τα	PnPe03NePIAcWe	Select	WWS Index	1			
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By						Remove
LEMMA	Lemma: βάζω	WordForm:	Vb	Select	WWS Index	1			
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By						Remove
LEMMA	Lemma: με	WordForm: με	AsPpSp	Select	WWS Index				
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By						Remove
NP-ACC	Lemma:	WordForm:	Select	WWS Index					

**Picture 9. Example of the controlled vocabulary of the NP-ACC LABEL.**

- ii. Free verbal constituents: If there are no constraints on a subordinated clause, then the following are encoded: the complementizer that introduces the clause, and then an VP. Picture 10 shows the encoding of the MWE ‘βάζω προς να VP’.

Tokens									
NP-NOM	Lemma:	WordForm:	Select	WWS Index					
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By						Remove
LEMMA	Lemma: βάζω	WordForm:	VbAv	Select	WWS Index				
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By						Remove
LEMMA	Lemma: προς	WordForm: προς	AdXxBa	Select	WWS Index				
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By						Remove
COMPL	Lemma: να	WordForm: να	CjSb	Select	WWS Index				
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By						Remove
VP	Lemma:	WordForm:	VbPe	Select	WWS Index				
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By	Controlled By: 1: NP-NOM						Remove

**Picture 10. Example of a VP in a MWE.**

**Columns 3 and 4:**

Marking of the fixed strings and the constrained semi-fixed strings is done explicitly on these columns. A MWE may contain one (6) or more (7) fixed strings. On Column 3 the fixed word forms are given; their PAROLE tags are specified on the fourth column (Picture 11).

WordForm: στους	AsPpPaMaPIAc	Select
Controlled By		
WordForm: πέντε	Nm	Select
Controlled By		
WordForm: δρόμους	NoCmMaPIAc	Select

**Picture 11. The fixed word forms with their PAROLE tags.**

On Column 5 each fixed string is assigned its own number. We make sure that all the words in the same fixed string are assigned the same number and that all the fixed strings in the MWE are marked. Please notice that no theoretical commitment is made as regards the syntactic function of the fixed strings (Picture 12).

- (6) τάζω λαγούς<sup>1</sup> με<sup>1</sup> πετραχήλια<sup>1</sup>  
 Promise rabbits.ACC with vestments.ACC  
 “To promise something impossible”
- (7) κάνω μαύρα<sup>1</sup> μάτια<sup>1</sup> να<sup>2</sup> δω<sup>2</sup> NP.ACC  
 Do black<sup>1</sup> eyes<sup>1</sup> to<sup>2</sup> see<sup>2</sup> NP.ACC

Preview	General	Forms	Use	Corpus	Diagnostics	Relations
Tokens						
NP-NOM-anim	Lemma:	WordForm:		Select	WWS Index:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:			Remove
LEMMA	Lemma: τάζω	WordForm:	Vb	Select	WWS Index:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:			Remove
LEMMA	Lemma: λαγός	WordForm: λαγούς	NoCmMaPIAc	Select	WWS Index: 1	Remove
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:			Remove
LEMMA	Lemma: με	WordForm: με	AsPpSp	Select	WWS Index: 1	Remove
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:			Remove
LEMMA	Lemma: πετραχήλι	WordForm: πετραχήλια	NoCmNePIAc	Select	WWS Index: 1	Remove

**Picture 12. The coindexed tokens denote that the string “λαγούς με πετραχήλια” is fixed and it should be treated like one token.**

A semi-fixed constituent of the type discussed before (4) or of the type given below (8) is assigned the necessary constraints. (8) contains the cranberry word ταπί that will be a word form with no tag assigned and the adjective ψύχραμος that is controlled for number and gender by the free NP-NOM (the free subject of the verb MWE) therefore no special constraints will be assigned to it.

- (8) Ο Νίκος έμεινε ταπί και ψύχραμος  
 The Nikos stayed broke and cool.ACC  
 “Nikos was broke”

We **introduce a new form** if we have variations of the MWE due to the occurrence of diminutives or slightly different versions of a lemma (ουράνια-μειουράνια), specified adjectives, an alternation of definite/indefinite determiner and an alternation of prepositions (Pictures 13-19).

We **do not introduce a new form** in order to encode **differences in animacy** as explained in **Columns 1 and 2** above.

## 1. Diminutives

- (9) a. Θα έχει βάλει η Ελένη την ουρά της βέβαια.  
 Will have put the Helen<sub>i</sub> the tail her<sub>i</sub> of\_course  
 “Of course Helen has been involved in this”
- b. Θα έχει βάλει η Ελένη την ουρίτσα της βέβαια.  
 Will have put the Helen<sub>i</sub> the small\_tail her<sub>i</sub> of\_course  
 “Of course Helen has been involved in this”

The screenshot shows the IDION software interface. At the top, there are tabs: Preview, General, Forms, Use, Corpus, Diagnostics, Relations. Below the tabs is the 'Tokens' section, which is divided into two panes. Each pane contains a table of tokens with columns for Lemma, WordForm, and Controlled By. The 'Bound' checkbox is checked for the token 'μου' (μου) with a 'Bound By' value of '1: NP-NOM-anim'. There is also an 'Add Toks' button at the bottom right of the top pane.

Picture 13. Diminutives in a MWE.

## 2. Adjectives (small number)

- (10) μου άναψε όλα τα λαμπάκια  
 me.DAT.GEN turn-on.3SG.PAST *all*.ACC the lights.ACC  
 “(he) made me completely furious”
- (11) να δούμε ποιος θα βγάλει το λερό φίδι από τη φωλιά του  
 to see who.NOM will.FUT pull-out.3SG the *filthy*.ACC  
 snake.ACC from the nest.ACC its  
 “let’s see who is going to do all the dirty work for us”
- (12) θέλει να κόψει όλες τις γέφυρες  
 wants to cut *all*.ACC the bridges.ACC  
 “he wants to be completely independent”

**Picture 14. The two alternative forms of the MWE “μου άναψε όλα τα λαμπάκια”.**

### 3. Definite/Indefinite article

- (13) a. Οι χειρισμοί του έβαλαν τη χώρα στο μάτι ενός κυκλώνα.<sup>3</sup>  
the handlings.NOM his put.3PL.PAST the  
country in\_ the eye a.INDEF typhoon.GEN  
“His handling has put the country in an unfavorable  
position”
- b. Οι "αγορές" έβαλαν στο μάτι του κυκλώνα τη Γαλλία.<sup>4</sup>  
the “markets”.NOM put.3PL.PAST in\_ the eye  
the.DEF typhoon.GEN the France  
“The Markets have put France in an unfavorable position”

**Picture 15. The MWE’s form with an indefinite article.**

<sup>3</sup> <http://glotta.ntua.gr/courses/Technoglossia/Dissertations/samaridi-thesis.pdf>

<sup>4</sup> <http://www.triklopodia.gr/elas-den-tous-chtipise-to-aftokinito-tou-salma-alla-emis/>

Tokens			
NP-NOM	Lemma:	WordForm:	
<input checked="" type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
LEMMA	Lemma: βάζω	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By: PRO	
NP-ACC	Lemma:	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: στο	WordForm: στο	AsPpPaNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: μίση	WordForm: μίση	NoCmNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: ο	WordForm: του	AtDfNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: κυκλώνας	WordForm: κυκλώνα	NoCmNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	

Picture 16. The MWE's form with a definite article.

#### 4. Alternative prepositions

- (14) βάζω το μαχαίρι στο/ως/μέχρι το κόκαλο  
 put the knife in\_the/until/till the bone  
 “to implement radical solutions/ to investigate a case without  
 fear of whom the investigation might touch”
- (15) παίρνω μπρος/ μπροστά  
 take ahead/ahead.ADV  
 “be activated”

<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
LEMMA	Lemma: βάζω	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: ο	WordForm: το	AtDfNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: μαχαίρι	WordForm: μαχαίρι	NoCmNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: στο	WordForm: στο	AsPpPaNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: κόκαλο	WordForm: κόκαλο	NoCmNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	

Tokens			
NP-NOM	Lemma: Free Subject	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
LEMMA	Lemma: βάζω	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: ο	WordForm: το	AtDfNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: μαχαίρι	WordForm: μαχαίρι	NoCmNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: μέχρι	WordForm: μέχρι	AsPpSp
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: ο	WordForm: το	AtDfNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound Bound By	Controlled By	
WF	Lemma: κόκαλο	WordForm: κόκαλο	NoCmNeSgAc

Picture 17. Alternative forms of the MWE “βάζω το μαχαίρι στο/μέχρι το κόκαλο”.

#### 5. Alternative form of the same lemma

- (16) κατεβαίνω στο δρόμο/ στους δρόμους  
 descend in\_the road/ in\_the roads  
 “go down the street to protest”

Tokens			
NP-NOM-anim	Lemma:	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
LEMMA	Lemma: κατεβαίνω	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
WF	Lemma: στο	WordForm: στο	AsPpPaMaSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
WF	Lemma: όρόμος	WordForm: όρόμο	NoCmMaSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:

Tokens			
NP-NOM	Lemma:	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
LEMMA	Lemma: κατεβαίνω	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
WF	Lemma: στο	WordForm: στους	AsPpPaFeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
WF	Lemma: όρόμος	WordForm: όρόμους	NoCmMaPlAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:

**Picture 18.** Example of a same lemma of a MWE with two possible forms.

## 6. Additional constituent that intensifies the meaning of the MWE

(17) ανάβω **όλα** τα λαμπάκια κτ  
 turn-on all.ADJ.PL.ACC the lights.ACC someone.GEN  
 “To make someone furious”

(18) πίνω το πικρό ποτήρι **ως τον πάτο**  
 drink the bitter.ADJ.SG.ACC glass.ACC to the bottom  
 “to experience something sad”

Tokens			
NP-NOM	Lemma:	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
LEMMA	Lemma: πίνω	WordForm:	
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
WF	Lemma: ο	WordForm: το	AtDfMaSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
WF	Lemma: πικρός	WordForm: πικρό	AjCpNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
WF	Lemma: ποτήρι	WordForm: ποτήρι	NoCmNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By: free subject	Controlled By:
WF	Lemma: ως	WordForm: ως	AsPpSp
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
WF	Lemma: ο	WordForm: το	AtDfNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:
WF	Lemma: τέλος	WordForm: τέλος	NoCmNeSgAc
<input type="checkbox"/> Optional	<input type="checkbox"/> Bound	Bound By:	Controlled By:

**Picture 19.** Example of an additional constituent that intensifies the meaning of the MWE.

## USE TAB:

For each MWE entry, a characteristic example along with the phonetic transcription, PAROLE transliteration and English translation is provided at the USE tab.

- A tabular representation of a glossed example that has been verbatim retrieved from corpora or the WEB is given. This table is automatically generated from the tab that follows (picture 20).

<b>Example:</b>	Από	την	άλλη	δεν	μπορώ	να	την	αφήσω	στους	πέντε	δρόμους.
<b>Parole:</b>	From	the	other	not	can	to	her	let	the	five	streets.
<b>Transcript:</b>	Apo	tin	ali	den	boro	na	tin	afiso	stous	pent	dromous.

Translation:

On the other hand, I can't abandon her.

Source:

<http://www.avsite.gr/forum/threads/m%CE%BF%CF%85-%CF%84%CE%BF-%CE%B1%CF%86%CE%B7%CF%83%CE%B1%CE%BD-%CF%83%CF%8...>

**Picture 20. The tabular organisation of the USE tab.**

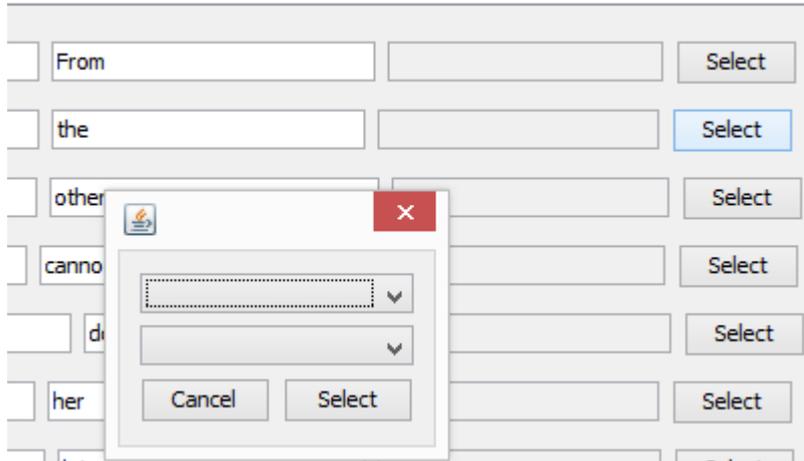
### Caution:

- Examples extracted from the corpus should be **spelt** and **punctuated correctly** e.g. the MWE “*πέρνω για ένα κομμάτι ψωμί*” should to be rewritten as “*παίρνω για ένα κομμάτι ψωμί*”.

Word	Phonetic	prefix	Parole
Από	Apo		From
την	tin		the
άλλη	ali		other
δεν	den		not
μπορώ	boro		can
να	na		to
την	tin		her
αφήσω	afiso		let
στους	stous		the
πέντε	pent		five
δρόμους.	dromous.		streets.

**Picture 21. An example extracted from the corpus.**

- The MWE encoded in the USE tab will be glossed with the PAROLE tagset ([http://nlp.ilsp.gr/nlp/tagset\\_examples/tagset\\_en/](http://nlp.ilsp.gr/nlp/tagset_examples/tagset_en/)). In the fifth column of the USE tab, the Select button allows for an easy and effective definition of every word (Pictures 22-23):



**Picture 22. The glossing button.**

Parole	Tag
From	
the	
other	
not	PtNg
can	VbMnIdPr01SgXxIpAvXx
to	PtSj
her	PnRe03FeSgAcXx
let	VbMnIdXx03SgXxPeAvXx
the	AsPpPaMaPIAc
five	NmCdMaPINmAj
streets.	NoCmMaPIAc

**Picture 23. The glossed tokens of the phrase depicted in picture 21.**

## CORPUS TAB:

Corpus data are introduced by creating new slots with the key **Add**. Corpus data are accessible from the DIAGNOSTICS tab.

The text corpora accommodated in the DB contain **both grammatical and ungrammatical strings** featuring MWEs. Strings are marked for their grammaticality by selecting a label from the appropriate pull down menu at the far right part of the window. The grammatical strings are drawn from the HNC<sup>5</sup> and from the WEB while the ungrammatical ones have been evaluated by native speakers (introspection).

The purpose of the CORPUS tab is to provide a concise collection of verbatim usages and of evaluated negative data that exemplify the analysis provided in the different tabs and may be of help in future implementations.

### Caution:

<sup>5</sup> HNC: Hellenic National Corpus.

We make sure that the **corpus examples feature the same meaning of the expression**. For instance, if we have the expression ‘βάζω μέσα κτ’ that means both ‘imprison smbd’ and ‘create debts for smbd’:

-we develop two different entries, one for each meaning of the expression

-we make sure that the corpus examples in each lexical entry exemplify the right meaning

## DIAGNOSTICS TAB:

With the Diagnostics tab, we investigate:

1. whether a verb MWE has a free subject or not (dedicated diagnostic)
2. the number of constituents that a MWE contains. For instance, the MWE below contains three constituents (example 19)

(19) [κάνω] [μαύρα μάτια] [να δω NP-ACC]  
 [do] [black eyes] [to see NP-ACC]

The diagnostics ‘admission of a free XP’ and ‘word order permutations’ are used as constituency diagnostics that show whether:

3. the MWE alternates between a form with a free NP-GEN and a form with a Dative Genitive
4. the fixed parts of the MWE can be replaced with a clitic in the same predication
5. the MWE participates in the causative-inchoative alternation
6. the MWE passivizes

Each question is assigned a yes/no button and the ability to be exemplified with corpus examples illustrating the phenomenon in question.

For each question, a detailed description is given below.

- Whether a verb MWE accepts a free subject or not. This is the basic diagnostic for the existence of a free subject. If different NPs trigger agreement on the verb a free subject exists, otherwise the subject is fixed. The situation is shown with corpus examples that demonstrate subject-verb agreement with a variety of subjects (Picture 24):

**Free/fixed subject diagnostic**  
 Can different NPs trigger agreement on the verb?

Οι Γάλλοι αστυνομικοί θέλουν να πιστεύουν ότι μετά από όσα έγιναν στη Μασσαλία δεν θ' αφήσουν τίπο...  
 Τίποτα δεν αφήνει στην τύχη ο Δημήτρης Ελευθερόπουλος και ποτέ δεν θέλει να πηγαίνει η ομάδα του σ...  
 Τίποτα δεν αφήνει στην τύχη το ο Τραϊανός Δέλλας, ο οποίος έχει παρακολουθήσει από κοντά όλα τα μ...

**Picture 24. Free/fixed subject diagnostic.**

- Whether the MWE accepts a free XP constituent. The position of the free constituent delineates the constituents of the MWE.

The examples below (20a, 20b, 20c) the free XPs (here an ADJP **μόνος του** and an ADVP **στην κυριολεξία**) indicate that the fixed string is ‘το φίδι από την τρύπα’.

- (20) a. προτιμά να βγάζει **μόνος του** το φίδι από την τρύπα<sup>6</sup>  
 prefers to pull-out **alone his** the snake from the nest  
 “He prefers to do the dirty work all alone”
- b. Όλοι αυτοί άνθρωποι... βγάζουν **στην κυριολεξία** το φίδι από<sup>7</sup> την τρύπα  
 All these people... pull-out **literally** the snake from the nest  
 “All these people... are literally doing all the dirty work”
- c. \*προτιμά να βγάζει το φίδι **μόνος του/στην κυριολεξία** από την τρύπα  
 \*prefers to pull-out alone his/literally from the nest

- Word order permutations are exemplified by retrieving examples from the corpus. Ungrammatical word orders are also exemplified with corpus examples that have been supplied by the encoders.

Word order permutations involving the subject as well

Άκουσα πως η ΕΡΤ θα κόψει τα πολλά πολλά με τη Γιουροβίζιον και αναθάρρησα. \*  
 Να κόψει η κυβέρνηση τα πολλά πολλά με τους δανειστές.

**Picture 25. Example of a MWE's word order permutations.**

- If the MWE is fully fixed and doesn't allow word order permutations, it is possible to provide an ungrammatical example to show its limits in word ordering (Picture 26).

\* Στου λουτρού τα κρύα άφησαν την Μαντόνα.

**Picture 26. An unacceptable word order.**

As shown above (Picture 26), the WWS “στα κρύα του λουτρού” of the MWE “αφήνω κάποιον στα κρύα του λουτρού” is strictly fixed and, therefore, a word ordering of the form “στου λουτρού τα κρύα” is unacceptable.

- Cliticisation of the WWS

Cliticisation of the WWS is investigated in the context of the same MWE:

The value YES is assigned if cliticisation is possible in the context of the same MWE (21).

- (21) Έβαλε **την ουρά** του κι ο ... Τσιώλης!  
 Put **the tail**.ACC his and the... Tsiolis!  
 Ναι, **την** έβαλε.  
 Yes, **it**.PN.ACC put.3SG.PAST

The value NO is assigned if cliticisation is not possible in the context of the same MWE (22).

<sup>6</sup> (source: [www.newsbeast.gr/weekend/.../ithopoioi-pou-traumatistikan-sta-gurismat](http://www.newsbeast.gr/weekend/.../ithopoioi-pou-traumatistikan-sta-gurismat))

<sup>7</sup> (source: <http://www.zoosos.gr/article/5684/kinduneuei-e-zoe-tous-epeide-uperaspizodai-ta-kakopoiemena-zoa-sta-hania>)

- (22) Έταξε στην Ελένη **λαγούς** με πετραχήλια.  
 Promised in Helen **hares.ACC with stoles.ACC**  
 \*Ναι, **τους** έταξε.  
 \*Yes, **them.PN.ACC** promised.

The value ‘---’ is assigned when the cliticisation diagnostic is irrelevant, eg. when the WWS is a PP.

- Alternation of a free NP-GEN/ free σε-PP/ free από-PP with Dative Genitive

The value YES is assigned if the alternation with a Dative Genitive is possible (23b).

- (23) a. Ο Γιώργος έκοψε τα φτερά της Ελένης.  
 The George cut the wings.ACC Helen.GEN  
 “George discouraged Helen.”
- b. Της έκοψε τα φτερά ο Γιώργος.  
 her.DAT.GEN cut the wings.ACC the George.NOM

The value NO is assigned if the alternation with a Dative Genitive is not possible (24b).

- (24) a. Τρώω τα νύχια μου.  
 eat the nails.ACC my.POSS.GEN  
 “I’m anxious”
- b. \*Μου τρώω τα νύχια.  
 me.DAT.GEN eat the nails

It is left to future decision whether the dative genitive form of an alternating expression will be encoded explicitly or it will be derived with some rule from the without-the-dative-genitive encoded version. In fact, once we know whether it alternates or not and the structure of the MWE (both are tagged in the DB), we can apply one of the two rules of dative genitive alternation that can be defined—NP-NOM and Y are optional:

1. If the MWE alternates and is of the form [NP-NOM, X, σε κπ], it receives the form [NP-NOM, μου, X]
2. If the MWE alternates and is of the form [NP-NOM, X,NP-GEN,Y], it receives the form [NP-NOM, μου, X,Y]

**Caution: We are interested in the NP-GEN attached to a WWS only.** Dative Genitives that are not related with constituents attached to the WWS (*τους πήρανε τα σπίτια [για ένα κομμάτι ψωμί]WWS*) or cliticisation of free objects (*τα αγόρασα για ένα κομμάτι ψωμί*) are results of the application of normal grammar rules on compositional structures---and indeed free NPs within MWEs have compositional semantics. We exemplify the problem with a corpus example concerning the MWE “*παίρνω κάτι για ένα κομμάτι ψωμί*” (25a). The WWS here is “για ένα κομμάτι ψωμί” as it occurs in several contexts (25b-25g). The verb *παίρνω* is otherwise normal in that it takes a normal subject, object, SOURCE and only the VALUE “για ένα κομμάτι ψωμί” is fixed (25g).

- (25) a. Για ένα κομμάτι ψωμί θα τους τα<sub>i</sub> πάρουν τα σπίτια<sub>i</sub> οι τράπεζες.  
 For a slice bread.ACC will them.PN.ACC them<sub>i</sub>.PN.ACC

take.3PL.PRES the houses<sub>i</sub> .ACC the banks.NOM  
“Banks will buy their homes for a song.”

- b. Το αγόρασα για ένα κομμάτι ψωμί.
- c. Το πούλησα για ένα κομμάτι ψωμί.
- d. Του το ζητάνε για ένα κομμάτι ψωμί.
- e. Δουλεύω για ένα κομμάτι ψωμί.
- f. Τα ακίνητα θα βγαίνουν προς πλειστηριασμό για "ένα κομμάτι ψωμί".
- g. Κάποιος παίρνει κάτι από κάποιον για ένα κομμάτι ψωμί.

Therefore, the particular verb MWE (25a) stands for no idiomatic meaning of the verb. We still keep it in the DB as a prototypical usage of the WWS that fills in the VALUE slot.

The presence of clitics with the particular MWE are besides the point because the WWS is a για-PP that does not cliticise by definition. The example (25a) has two clitics that are not about the WWS. The one is a dative genitive that alternates with an από-PP (instantiating the SOURCE) that is not attached to the WWS and the other clitic occurs in a clitic doubling structure that involves the free object of the MWE.

- Causative-inchoative alternation

The value YES is assigned if the alternation is possible (26).

(26) a. Ανάβω τα λαμπάκια σε κάποιον  
Turn-on.1SG.PRES the lights.ACC to someone  
“To make someone furious”

b. Μου ανάβουν τα λαμπάκια.  
me.DAT.GEN turn-on.3PL.PRES the lights.NOM  
“Something makes me furious”

The value NO is assigned if the alternation is NOT possible (27).

- (27) a. Κόβω τις φλέβες μου.  
cut.1SG.PRES the veins,ACC my.POSS.GEN  
“I’m bored to death”
- b. \*Οι φλέβες μου κόβουν.  
the veinsSUB.NOM my-POSS.GEN cut..3PL.PRES

- Passivisation. Whether the verb of the MWE passivises or not (Picture 27):

Passivisation

Αντε, μπράβο! τώρα που κόπηκαν τα πολλά πολλά με τις επιχορηγήσεις, ευκαιρία να βάλουν οι συνήθεις...

**Picture 27. Passivisation of a MWE.**

The value YES is assigned if the alternation is possible (28).

(28) Τοποθετήθηκε ο πήχης ψηλά.  
Placed-3sg.PAST.PASS the cubit-NOM up-ADV.  
“A high target was set”

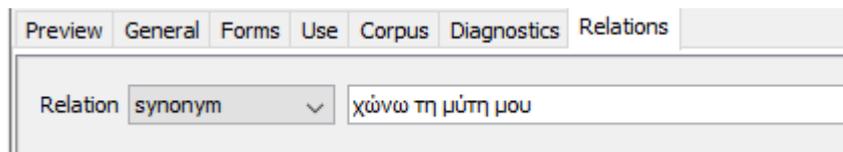
The value NO is assigned if the alternation is NOT possible (29).

- (29) \* Αφέθηκαν παιδιά πίσω του.  
\* Let-3pl.PAST.PASS children-NOM behind him-PN.GEN.  
\* “He had a family when he died”

## RELATIONS TAB:

In the RELATIONS tab the semantic relations among MWEs are documented. The following relations can be encoded:

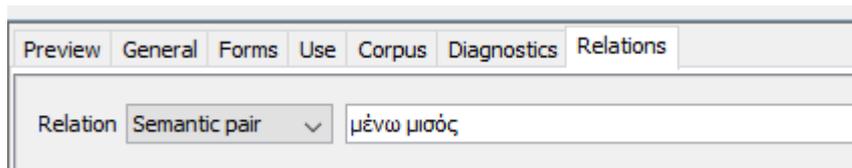
- Synonymous MWEs (Picture 28)



The screenshot shows a software interface with a 'Relations' tab selected. A dropdown menu is set to 'synonym' and the text 'χώνω τη μύτη μου' is entered in the adjacent field.

**Picture 28. Synonymous MWE with the MWE “βάζω την ουρά μου”.**

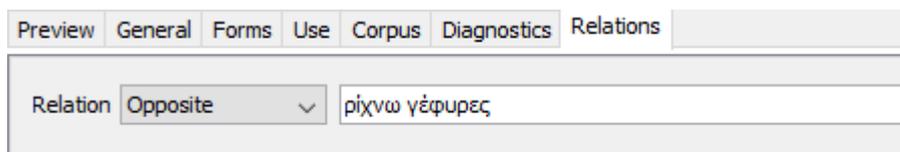
- Semantic pair (MWEs with common nominal phrase but with different verb normally standing in a causative-non causative relation) (Picture 29)



The screenshot shows a software interface with a 'Relations' tab selected. A dropdown menu is set to 'Semantic pair' and the text 'μένω μισός' is entered in the adjacent field.

**Picture 29. Semantic pair with the MWE “αφήνω μισό κάτι”.**

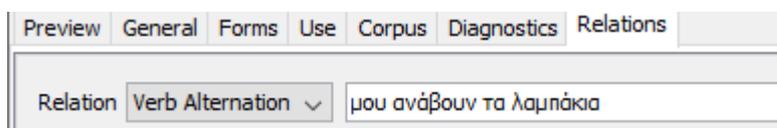
- Opposite MWEs (Picture 30)



The screenshot shows a software interface with a 'Relations' tab selected. A dropdown menu is set to 'Opposite' and the text 'ρίχνω γέφυρες' is entered in the adjacent field.

**Picture 30. Opposite MWE with the MWE “χωρίζω τα τσανάκια μου”.**

- Verb alternations (MWEs where the verb alternates between different syntactic patterns) (Picture 31)



The screenshot shows a software interface with a 'Relations' tab selected. A dropdown menu is set to 'Verb Alternation' and the text 'μου ανάβουν τα λαμπάκια' is entered in the adjacent field.

**Picture 31. The MWE “μου ανάβουν τα λαμπάκια” participates in the causative-inchoative alternation.**