Lexical Resource for free subject verb MWEs

Stella Markantonatou[1], Erasmia Koletti[2], Elpiniki Margariti[2], Panagiotis Minos[1], Aimilia Stripeli[2], Georgios Zakis[2], Niki Samaridi[3]

[1]Institute for Language and Speech Processing/"Athena" RIC, marks@ilsp.gr, pminos@gmail.com
[2]National and Kapodistrian University of Athens, erkg7@yahoo.gr, elpimargariti@gmail.com, astripeli@gmail.com, georgizak@gmail.com, [3]nsamaridi@gmail.com

WP1

Introduction

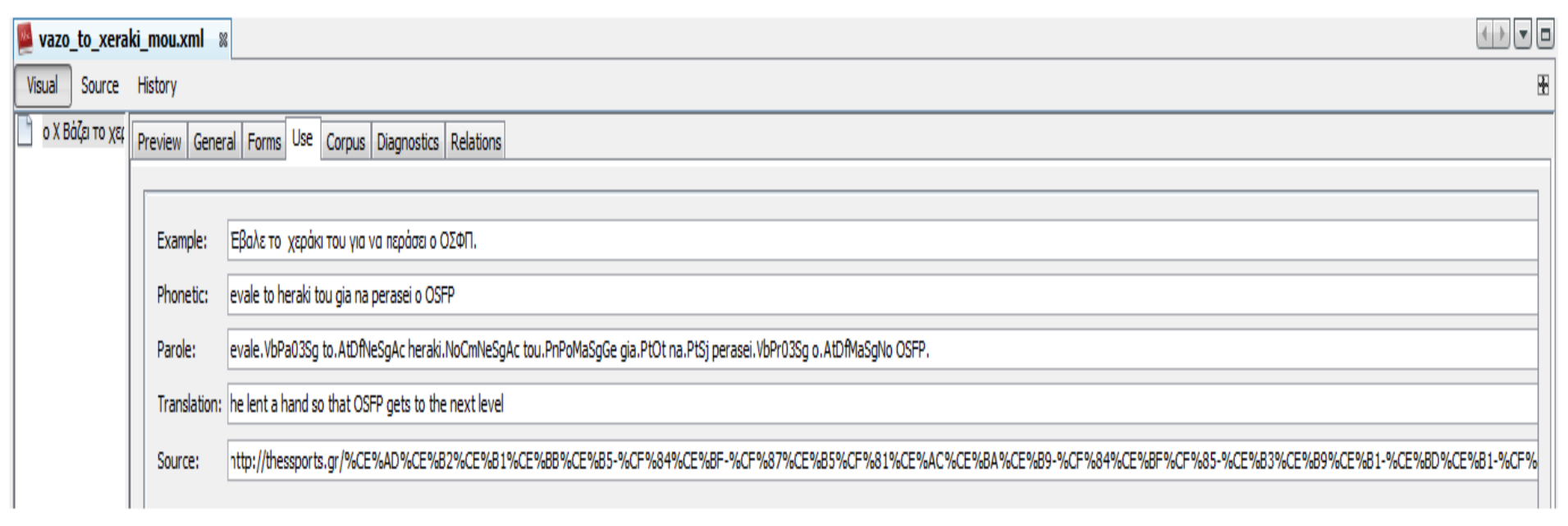
The database developed combines a wide range of linguistic information on Modern Greek free subject verb multi-word expressions (MWEs). We refer to the classification of MG MWE by Samaridi & Markantonatou (2014) and of English by Sailer (2014). The DB allows for encoding:

1. The grammatical features of a string that single it out as a MWE and possible specialized modifiers aiming at supporting parsing in different set ups
2. MWE meaning and relations among MWEs
3. Corpora of grammatical and ungrammatical strings containing the MWEs.

We have developed a custom-made Java desktop application based on the NetBeans Rich-Client Platform (RCP) framework for the editing of the XML lexical resource.

A typical DB entry

A characteristic example, along with the phonetic transcription, PAROLE transliteration and English translation is provided at the 'Use' tab for each MWE entry.

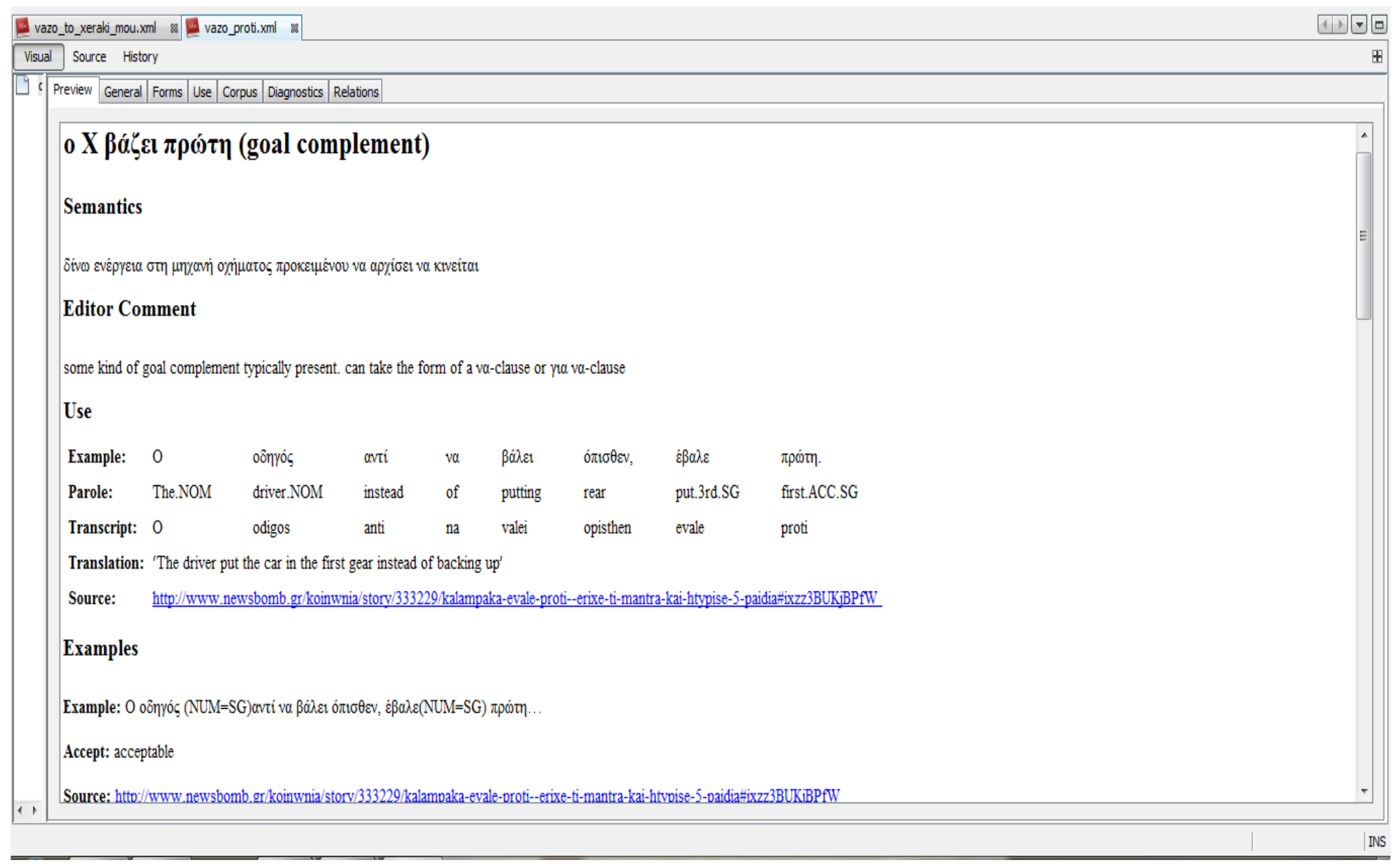


Example: Εβαλε το χεράκι του για να περάσει ο ΟΣΦΠ.
Phonetic: evale to heraki tou gia na perasei o OSFP
Parole: evale.VbPa03Sg to.AtDfNeSgAc heraki.NoCmNeSgAc tou.PnPoMaSgGe gia.PtOt na.PtSj perasei.VbPr03Sg o.AtDfMaSgNo OSFP.
Translation: he lent a hand so that OSFP gets to the next level
Source: http://thessports.gr/%CE%AD%CE%B2%CE%B1%CE%BB%CE%B5-%CF%84%CE%BF-%CF%87%CE%B5%CF%81%CE%AC%CE%BA%CE%B9-%CF%84%CE%BF%CF%85-%CE%B3%CE%B9%CE%B1-%CE%BD%CE%B1-%CF%

1. Screenshot of 'Use' tab

Grammatical structure information is encoded in the 'Forms' and 'Diagnostics' tabs. Lexicographic information is encoded in both the 'General' and the 'Relations' tabs. The corpus built in the designated tab is inter-connected to the diagnostics tab and provides examples.

The tab 'Preview' is auto-generated and presents all the information about the entry.



ο Χ βάζει πρώτη (goal complement)

Semantics

δίνω ενέργεια στη μηχανή οχήματος προκειμένου να αρχίσει να κινείται

Editor Comment

some kind of goal complement typically present. can take the form of a να-clause or για να-clause

Use

Example:	Ο	οδηγός	αντί	να	βάλει	όπισθεν,	έβαλε	πρώτη.
Parole:	The.NOM	driver.NOM	instead	of	putting	rear	put.3rd.SG	first.ACC.SG
Transcript:	O	odigos	anti	na	valei	opisthen	evale	proti

Translation: 'The driver put the car in the first gear instead of backing up'
Source: http://www.newsbomb.gr/koinwnia/story/333229/kalampaka-evale-proti--erixe-ti-mantra-kai-htypise-5-paidia#ixzz3BUKjBPfW

Examples

Example: Ο οδηγός (NUM=SG)αντί να βάλει όπισθεν, έβαλε(NUM=SG) πρώτη…

Accept: acceptable

Source: http://www.newsbomb.gr/koinwnia/story/333229/kalampaka-evale-proti--erixe-ti-mantra-kai-htypise-5-paidia#ixzz3BUKjBPfW

2. Screenshot of 'Preview' tab

Grammatical features

We consider as a MWE a string with no compositional meaning. Each entry is examined to obtain a full description of its identifying features:

- form and number of the necessary words in a string (tokens)
- form and number of optional lemmata (e.g. modifiers)
- existence of Words With Spaces (WWS). We define as a WWS a set of words in fixed order with limited or no declination freedom.
- morphology
- deeper syntactic relations (control/ binding)
- internal/external modification

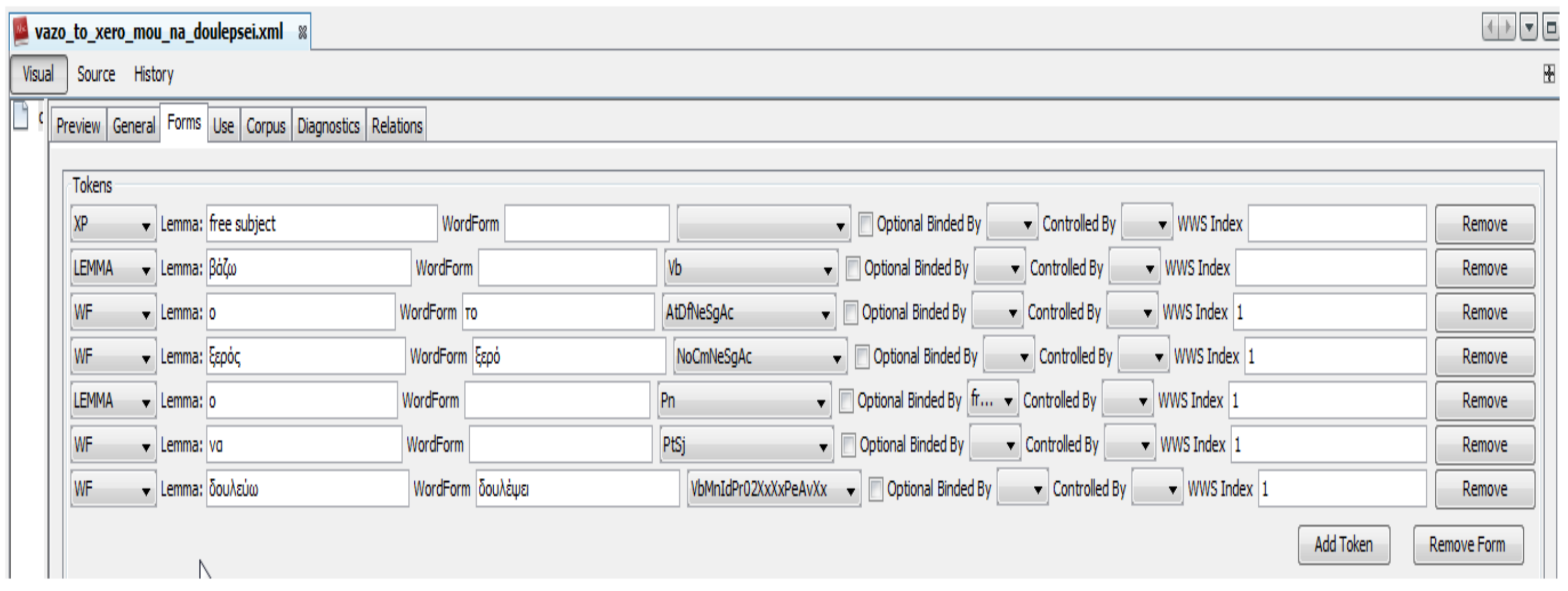
Two points are due in order to further clarify the adopted approach:

1. The syntactic properties encoded are those of the MWE in question. The 'underlying syntax' applied to create the MWE is not encoded, though the two sets of syntactic properties may diverge or intersect.
2. MWEs are not classified in terms of flexibility. We are not as yet sure whether the flexible/non-flexible/semi-flexible distinction is of parsing importance once the notion of WWS has been accounted for in the Grammar.

MWE Analysis

Information encoded in the Forms tab:

- The tokens constituting the MWE. Each token is individually encoded and tagged in PAROLE (http://nlp.ilsp.gr/nlp/tagset_examples/tagset_en/). Tokens are labelled as follows: LEMMA: declinable, WF (WORDFORM): non-declinable, CLAUSE, XP and MODIFIER. Optional tokens are marked as such in the DB
- Deep syntactic relations, i.e. binding and control
- Tokens constituting parts of a WWS. The WWS tokens are co-indexed with a common number in the option WWS Index in order to demonstrate their linking.



XP	Lemma: free subject	WordForm		☐ Optional	Binded By	Controlled By	WWS Index	Remove
LEMMA	Lemma: βάζω	WordForm	Vb	☐ Optional	Binded By	Controlled By	WWS Index	Remove
WF	Lemma: ο	WordForm το	AtDfNeSgAc	☐ Optional	Binded By	Controlled By	WWS Index 1	Remove
WF	Lemma: ξερός	WordForm ξερό	NoCmNeSgAc	☐ Optional	Binded By	Controlled By	WWS Index 1	Remove
LEMMA	Lemma: ο	WordForm	Pn	☐ Optional	Binded By fr...	Controlled By	WWS Index 1	Remove
WF	Lemma: να	WordForm	PtSj	☐ Optional	Binded By	Controlled By	WWS Index 1	Remove
WF	Lemma: δουλεύω	WordForm δουλέψει	VbMnIdPr02XxXxPeAvXx	☐ Optional	Binded By	Controlled By	WWS Index 1	Remove

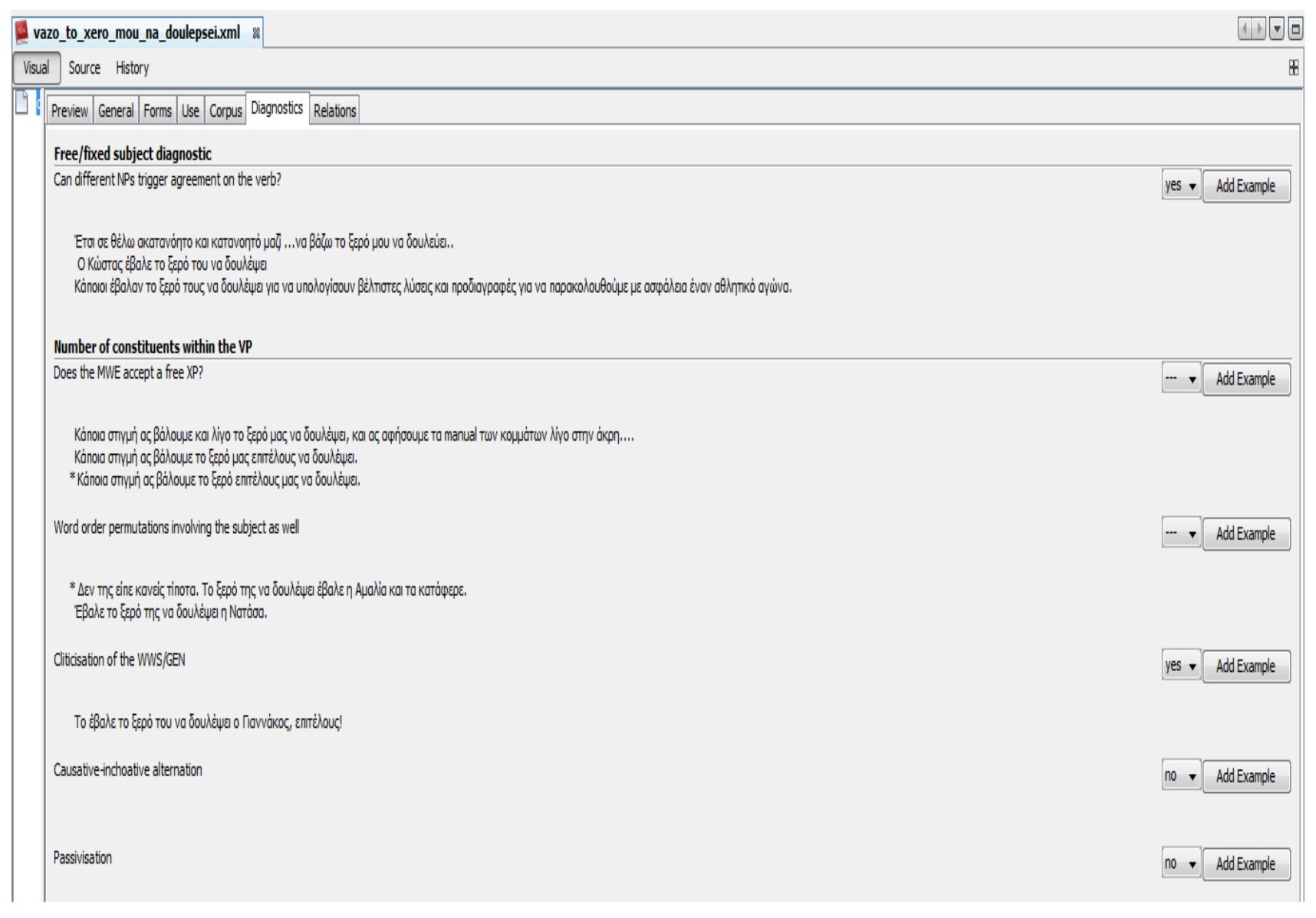
3. Screenshot of 'Forms' tab

Information encoded in the Diagnostics tab:

In the Diagnostics tab, we examine the following:

- Can different NPs trigger agreement on the verb?
- Does the MWE accept a free XP? What is its syntactic function?
- Which are the allowed word order permutations including the subject as well?
- Can the WWS be replaced by a clitic (weak personal pronoun)?
- Can the verb of the MWE show the Causative-Inchoative alternation?
- Can the verb within the MWE passivize?

Each question is assigneda yes/no button and the ability to draw corpus examples illustrating the phenomenon in question.



Free/fixed subject diagnostic

Can different NPs trigger agreement on the verb? yes

Έτσι σε θέλω ακατανόητο και κατανοητό μαζί ...να βάζω το ξερό μου να δουλεύει..
Ο Κώστας έβαλε το ξερό του να δουλέψει
Κάποιοι έβαλαν το ξερό τους να δουλέψει για να υπολογίσουν βέλτιστες λύσεις και προδιαγραφές για να παρακολουθούμε με ασφάλεια έναν αθλητικό αγώνα.

Number of constituents within the VP

Does the MWE accept a free XP? ---

Κάποια στιγμή ας βάλουμε και λίγο το ξερό μας να δουλέψει, και ας αφήσουμε τα manual των κομμάτων λίγο στην άκρη....
Κάποια στιγμή ας βάλουμε το ξερό μας επιτέλους να δουλέψει.
*Κάποια στιγμή ας βάλουμε το ξερό επιτέλους μας να δουλέψει.

Word order permutations involving the subject as well ---

*Δεν της είπε κανείς τίποτα. Το ξερό της να δουλέψει έβαλε η Αμαλία και τα κατάφερε.
Έβαλε το ξερό της να δουλέψει η Νατάσα.

Cliticisation of the WWS/GEN yes

Το έβαλε το ξερό του να δουλέψει ο Γιαννάκος, επιτέλους!

Causative-inchoative alternation no

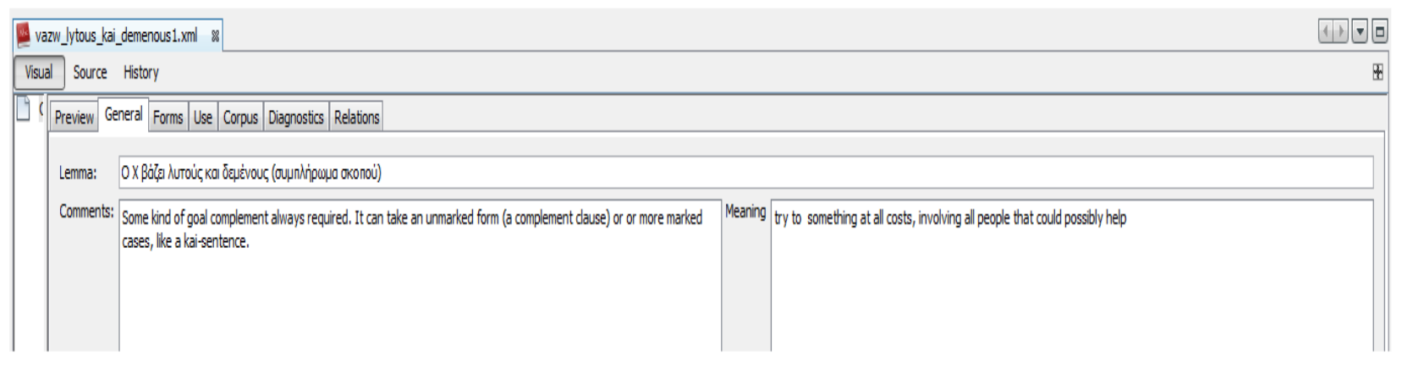
Passivisation no

4. Screenshot of "Diagnostics" tab

Lexicographic Information

Information encoded in the "General" tab.

In the "General" tab the meaning(s) of the MWE is encoded. There is also space to add some general comments/notes about the MWE examined.



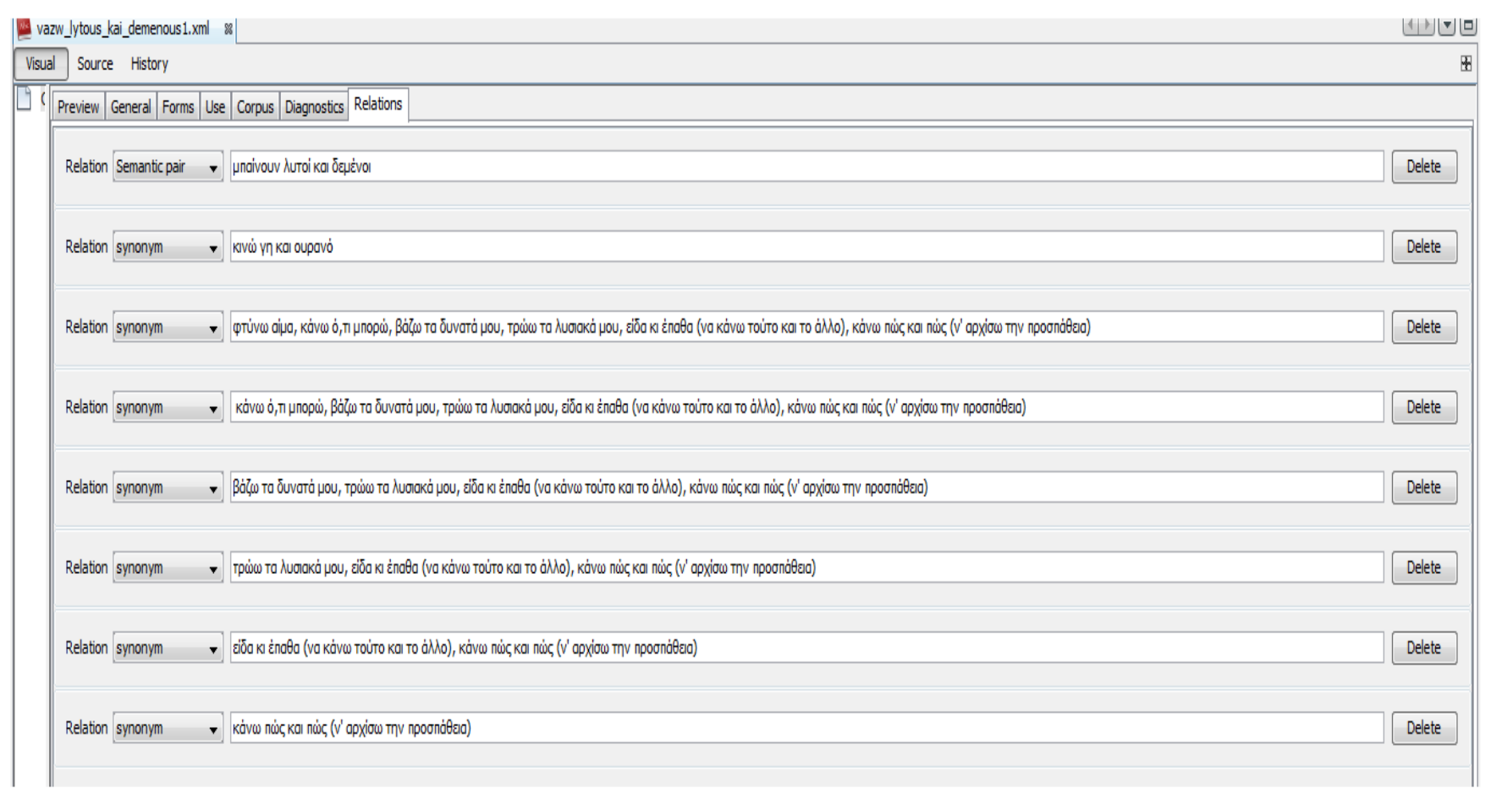
Lemma: Ο Χ βάζει λυτούς και δεμένους (συμπλήρωμα σκοπού)
Comments: Some kind of goal complement always required. It can take an unmarked form (a complement clause) or or more marked cases, like a kai-sentence.
Meaning: try to something at all costs, involving all people that could possibly help

5. Screenshot of "General" tab

Information encoded in the "Relations" tab.

In the "Relations" tab we store the semantic relations among MWEs. The following relations can be encoded:

- Synonymous MWEs
- Semantic pair (MWEs with common nominal phrase but with different verb)
- Antonymous MWEs
- Verb alternations (causative-inchoative, active-passive)



Relation Semantic pair: μπαίνουν λυτοί και δεμένοι
Relation synonym: κινώ γη και ουρανό
Relation synonym: φτύνω αίμα, κάνω ό,τι μπορώ, βάζω τα δυνατά μου, τρώω τα λυσιακά μου, είδα κι έπαθα (να κάνω τούτο και το άλλο), κάνω πώς και πώς (ν' αρχίσω την προσπάθεια)
Relation synonym: κάνω ό,τι μπορώ, βάζω τα δυνατά μου, τρώω τα λυσιακά μου, είδα κι έπαθα (να κάνω τούτο και το άλλο), κάνω πώς και πώς (ν' αρχίσω την προσπάθεια)
Relation synonym: βάζω τα δυνατά μου, τρώω τα λυσιακά μου, είδα κι έπαθα (να κάνω τούτο και το άλλο), κάνω πώς και πώς (ν' αρχίσω την προσπάθεια)
Relation synonym: τρώω τα λυσιακά μου, είδα κι έπαθα (να κάνω τούτο και το άλλο), κάνω πώς και πώς (ν' αρχίσω την προσπάθεια)
Relation synonym: είδα κι έπαθα (να κάνω τούτο και το άλλο), κάνω πώς και πώς (ν' αρχίσω την προσπάθεια)
Relation synonym: κάνω πώς και πώς (ν' αρχίσω την προσπάθεια)

6. Screenshot of "Relations" tab

The Corpus

The DB currently contains ~300 Modern Greek MWEs drawn from a list of ~1120 MWEs (Samaridi 2014, http://users.sch.gr/samaridi/attachments/article/3/Lexical%20Resources.pdf).

The text corpora accommodated in the DB contain both grammatical and ungrammatical strings featuring MWEs and are directly linked to the Diagnostics tab. The grammatical strings are drawn from the HNC and from the WEB while the ungrammatical ones have been evaluated by native speakers (introspection).

Future Plans

Further analysis of WWS creating a series of WWS diagnostics that yields extra information on:

- whether it can be replaced by a definite/indefinite pronoun
- whether it can form the target of an anaphoric clause.
- whether it can be assigned any reference at all

A MWE may be assigned a meaning that is related through some modifier to another MWE that has been assigned a "basic meaning". We therefore plan to capture such relations by adding the relation: is modified form of… in the relations tab.

For instance, the MWE "μου.GEN.DAT ανάβουν.V.INCH τα.DET λαμπάκια.PL.NOM" (my lights are switched on) means "to become angry". A modified form of this MWE exists that has an extended meaning. The modified MWE contains the adjective ("όλα" (all)) that modifies "λαμπάκια" and increases semantic intensity: "μου.GEN.DAT ανάβουν.V.INCH όλα.ADJ τα.THE λαμπάκια.PL.NOM (all my lights are switched on) means "to become extremely angry".

Another goal is to make the lexical resource available on line.

References

- Manfred Sailer. 2014. MWE Template: English http://wiki.studiumdigitale.unifrankfurt.de/FB10_Parseme/index.php/MWE_Template:_English
- Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. In Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002), Mexico City, Mexico, pp. 1-15
- Samaridi, Niki and Stella Markantonatou. 2014. Classification of Modern Greek Verb MWEs. Frankfurt Workshop on Multi-word Expressions, co-hosting the PARSEME 3rd general meeting
- Villavicencio, A., T. Baldwin, B. Waldron. 2004. A Multilingual Database of Idioms. In Proceedings of the 4th International Conference On Language Resources and Evaluation, LREC-2004, Lisboa.