

Parsing Modern Greek verb MWEs in LFG/XLE

Niki Samaridi¹ & Stella Markantonatou²



¹National and Kapodistrian University of Athens

²Institute for Language and Speech Processing/
“Athena” RIC

WP2

Contents

- About the parsing system
 - About the filter
 - Deep grammar approach (LFG)
- LFG analysis
 - Natural LFG analyses
 - Not-so-Natural LFG analyses
- To do...

About the parsing system



Parsing system architecture

- ❑ Text is lemmatized and tagged with the ILSP FBT Tagger
 - ❑ MWE filter that marks Words_With_Spaces in MWEs (the filter reduces parsing load on XLE in the absence of an embedded morphological component)
 - ❑ The output is formatted, and
 - ❑ Feeds an LFG/XLE grammar that has been developed independently.
-
- ❑ 2500 MWEs identified and classified
 - ❑ 40% of them processed by our system

System features

□ ILSP FBT Tagger

- Brill tagger plus rules
- 584 PAROLE tags
- assigns lemma & set of tags

□ MWE recognizer:

- Filter lexicon
- Filtering algorithm
- Formatter

□ Deep grammar (LFG/XLE)

- General grammar of Modern Greek

The filter (1): Filter lexicon

Filter lexicon: Each MWE entry is specified for:

- Compositionality.
- The ‘signifier’ (the lemma that instructs the filter to look at the appropriate filter lexicon entries).
- Lemmatised form of Words_With_Spaces (WWS) (independent fixed MWEs or substrings of a MWE).
- PoS & morphological constraints for the WWS headword.
- Constraints on the lemmatized forms of the remaining constituents of a WWS.

Some lexical entries

t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12
single	πίνω	ο αμίλητος νερό	νερό	πίνω	VbMnAv ο		SgAc	αμίλητος	AjBaNeSgAc	νερό	SgAc

t1	t2	t3	t4	t5	t6	t7	t8	t9	t10
single	κάνω	μαύρος μάτι	μάτι	κάνω	VbMnAv μαύρος		AjBaNePIAc	μάτι	NoCmNePIAc

t1	t2	t3	t4	t5	t6	t7	t8
single	~	ταπί και ψύχραιμος	ψύχραιμος	ταπί	NoCmNeSg	ψύχραιμος	AjBa

t1	t2	t3	t4	t5	t6	t7	t8	t9	t10
double	περπατώ	πάνω σε τεντωμένος σκοινί	AdXxBa	περπατώ	VbMnAv	τεντωμένος	AjBaNeSgAc	σκοινί	NeSgAc

The filter (2): The filtering algorithm

- **A.** Is there a signifier?
- **A1.** NO: Copy the string to the formatter.
- **A2.** YES: Scan the filter lexicon for some WWS entry. Do morphological constraints on the filter lexicon entries match the lemma and the tags on the input string ?
- **B1.** NO: Copy the input string to the formatter.
- **B2.** YES: Consult the filter lexicon. Can the MWE can take a compositional reading?
- **C1.** YES: Send the input string to the formatter. Go to step C2.
- **C2.** NO: Replace the recognised substring(s) with the corresponding WWS and morphological constraints. Send the new string to the formatter.

Filter's output/XIE's input

(one non-compositional output)

□ #1 non-compositional

○ At Df Ma Sg Nm **Κώστας** No Pr Ma Sg Nm **πίνω** Vb Mn Id Pa 03
Sg Xx Pe Av Xx **ο_αμίλητος_νερό** No Cm Ne Sg Ac

□ #2 non-compositional

κάνω Vb Mn Id Pa 01 Sg Xx Ip Av Xx **μαύρος_μάτι** No Cm Ne Pl
Ac **να** Pt Sj **εγώ** Pn Pe Ma 02 Sg Ac We **βλέπω** Vb Mn Id Xx 01
Sg Xx Pe Av Xx

□ #3 non-compositional

ρίχνω Vb Mn Id Pr 01 Sg Xx Ip Av Xx **άδειος** Aj Ba Ne Pl Ac **να** Pt
Sj **πιάνω_γεμάτος** Vb Mn Id Xx 01 Sg Xx Pe Av Xx

Filter's output & Xle's input

(two outputs: non-compositional & compositional)

□#4 non-compositional

περπατώ Vb Mn Id Pr 01 Sg Xx Ip Av Xx

πάνω_σε_τεντωμένος_σκοινί Ad Xx Ba

□#5 compositional

περπατώ Vb Mn Id Pr 01 Sg Xx Ip Av Xx πάνω Ad Xx Ba σε As Pp

Sp τεντωμένος Aj Ba Ne Sg Ac σκοινί No Cm Ne Sg Ac

Deep analysis (LFG)

- The formatter output is parsed with an LFG/XLE grammar of MG (with sublexical rules).
 - MG MWEs are rich in syntactic structure despite any simplifications that might result from the usage of WWSs.
 - MWEs and compositional structures can be treated with more or less the same grammar.
- **We have manipulated only the lexicon but not the grammar rules.**

Classification of verb MWEs

□ Fixed MWE:

(no inflection, no intervening words, no word order variations, no alternations) → a Verb WWS

□ Semi fixed and flexible verb MWEs:

(inflected, word order permutations)

“NATURAL” LFG ANALYSES



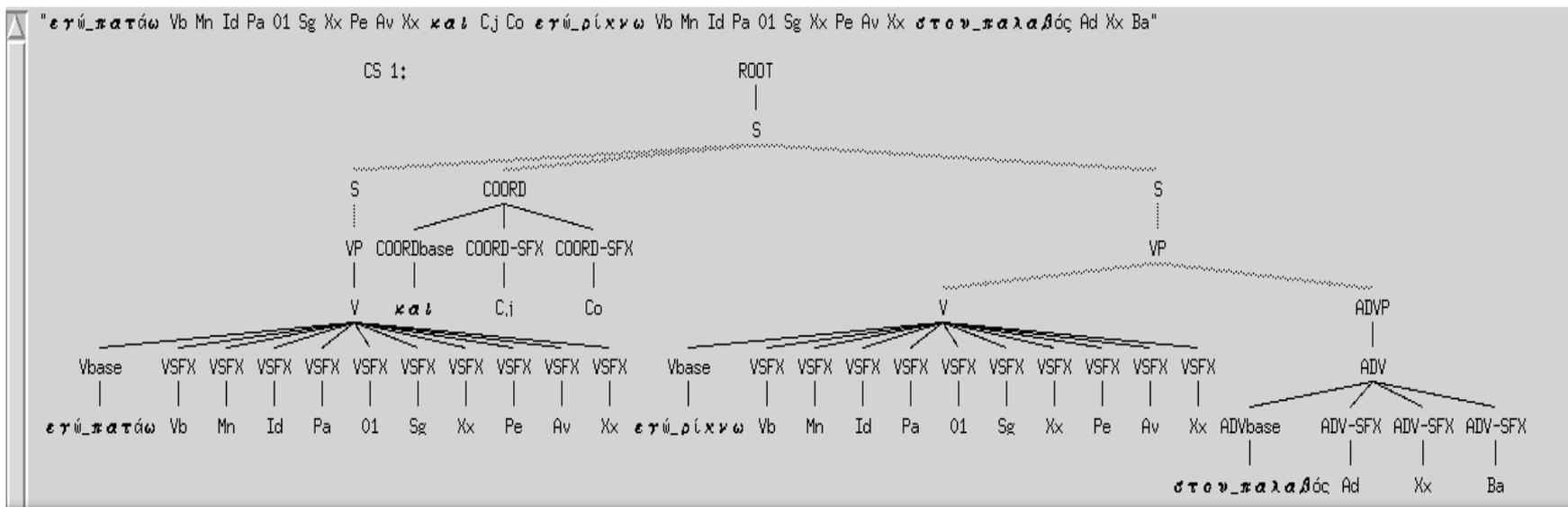
MG MWE syntactic patterns (1)

- **Fixed verb WWS:** no inflection or word permutation.
- **Free subject-verb:** inflected – SV/VS word order.
- **Impersonal verb:** inflected impersonal verb with a fixed object or a saturated sentential complement.
- **Free subject-copula-complement:** inflected copula, fixed or inflected complement, copula agrees with subject.

MG MWE syntactic patterns (2)

- ❑ **Free subject-verb-object:** inflected verb, fixed or non-fixed object.
- ❑ **Free subject-verb-fixed object with subject-bound possessive:** inflected verb, object modified by a possessive weak pronoun bound by the subject.
- ❑ **Free subject (controller)-verb-object-subordinated sentence with controlled subject:** inflected verb, the object may be fixed and/or the subordinated sentence may be semi-fixed.

Free subject with conjunction



1 valid F-structure for ROOT

kill prev next Commands Views a c n s x

lock F-structure #1

```
"ερώ_πατάω Vb Mn Id Pa 01 Sg Xx Pe Av Xx και Cj Co ερώ_ρίχνω Vb Mn Id Pa 01 Sg Xx Pe Av Xx στον_παλαβός Ad Xx Ba"
[
  [
    [
      PRED 'ερώ_πατάω<[1-SUBJ;PRO]>'
      SUBJ [
        PRED 'PRO'
        CASE nom, NUM sg, PERS 1
      ]
      TNS-ASP [MOOD indicative, PERF +, TENSE past]
      1[CLAUSE-TYPE decl, IDIOM deceived, PASSIVE -]
    ]
    [
      PRED 'ερώ_ρίχνω<[45-SUBJ;PRO], [82:στον_παλαβός]>'
      SUBJ [
        PRED 'PRO'
        CASE nom, NUM sg, PERS 1
      ]
      OBL-LOC 82[PRED 'στον_παλαβός']
      TNS-ASP [MOOD indicative, PERF +, TENSE past]
      45[CLAUSE-TYPE decl, IDIOM pretend_the_indifferent, PASSIVE -]
    ]
  ]
  38[COORD-FORM και]
]
```

Verbs with “augmented” valency

- To obtain “natural” LFG analyses, sometimes we were forced to assume “augmented” verb entries that are NOT in use otherwise:
- *ρίχνω* (throw) normally does not take a controlled sentential complement (XCOMP)
- *περπατώ* (walk) normally does not take a LOC argument denoting where one walks; arguably, this is an adjunct’s job

ρίχνω (throw) as a control verb

- V: **ρίχνω** <SUBJ, OBJ, XCOMP>, XCOMP COMPL=για να/μήπως, **v** OBJ PRED=άδειος, **v** XCOMP PRED= **πιάνω**_γεμάτος, XCOMP SUBJ=SUBJ, XCOMP PERF=+, ¬(XCOMP TENSE)

- **Gloss**

(NP/pro)_j throw empty to pro_j catch_full

- **Meaning**

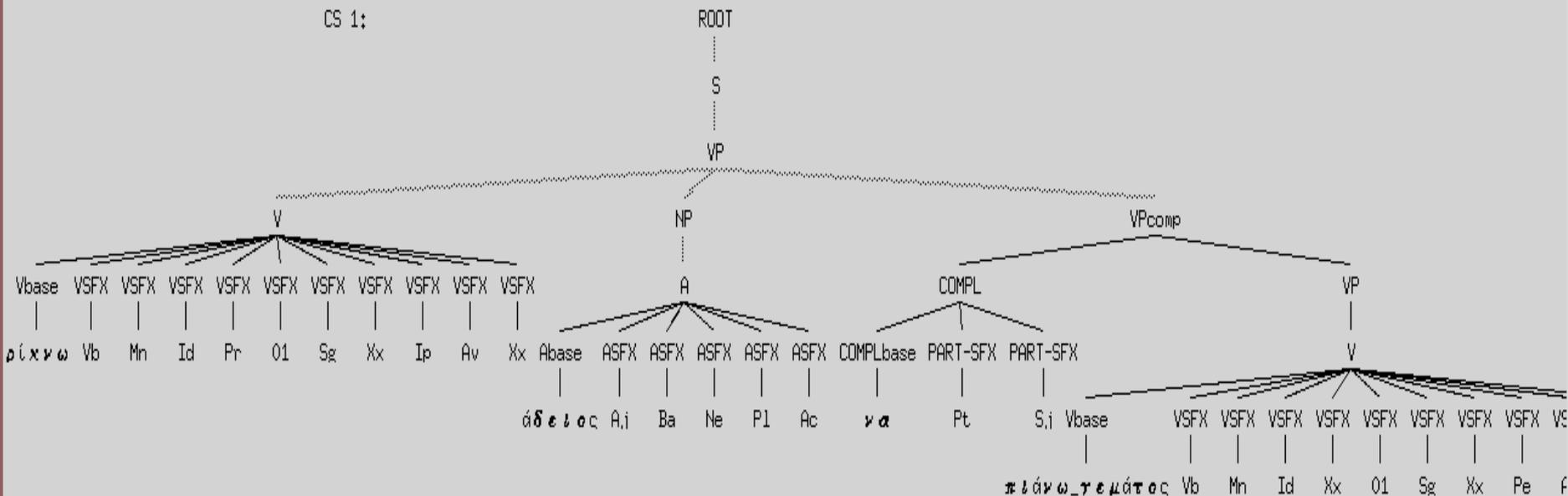
fish out (of/from)

- **Sub-WWS**

πιάνω_γεμάτος

v; where an XP may intervene

ρίχνω Vb Mn Id Pr 01 Sg Xx Ip Av Xx άδειος Aj Ba Ne Pl Ac να Pt Sj πιάνω_τεμάτος Vb Mn Id Xx 01 Sg Xx Pe Av Xx"



1 valid F-structure for ROOT

kill prev next Commands Views a c n s x

lock F-structure #1

"ρίχνω Vb Mn Id Pr 01 Sg Xx Ip Av Xx άδειος Aj Ba Ne Pl Ac να Pt Sj πιάνω_τεμάτος Vb Mn Id Xx 01 Sg Xx Pe Av Xx"

```

[PRED 'ρίχνω<[1-SUBJ;PRO], [38:άδειος], [70:πιάνω_τεμάτος]>'
SUBJ [PRED 'PRO'
CASE nom, NUM sg, PERS 1]
OBJ [PRED 'άδειος'
38CASE acc, GEND neut, NUM pl]
XCOMP [PRED 'πιάνω_τεμάτος<[1-SUBJ;PRO]>'
SUBJ [1-SUBJ;PRO]
TNS-ASP [MOOD indicative, PERF +]
70COMPL-FORM να, PASSIVE -]
TNS-ASP [MOOD indicative, PERF -, TENSE pres]
1CLAUSE-TYPE decl, COMPL-FORM να, IDIOM fish_out, PASSIVE -]

```

περπατώ (walk) with a LOC complement

1 valid F-structure for ROOT

kill prev next Commands Views a c n s x

lock F-structure #1

"περπατώ Vb Mn Id Pr 01 Sg Xx Ip Av Xx πάνω Ad Xx Ba σε As Pp Sp τεττωμένος Aj Ba Ne Sg Ac σκοινί No Cm Ne Sg Ac"

```
PRED 'περπατώ<[1-SUBJ;PRO]>'
SUBJ [ PRED 'PRO'
      CASE nom, NUM sg, PERS 1 ]
ADJUNCT { PCOMP [ PRED 'πάνω<[50;σε]>'
                  PRED 'σε<[91;σκοινί]>'
                  PRED 'σκοινί'
                  ADJUNCT { PRED 'τεττωμένος'
                            [59 CASE acc, GEND neut, NUM sg]
                          ]
                  NTYPE [NSYN common]
                        91 CASE acc, GEND neut, NUM sg, PERS 3
                  ]
          ]
        ]
TNS-ASP [MOOD indicative, PERF -, TENSE pres]
1 [CLAUSE-TYPE decl, PASSIVE -
```

1 valid F-structure for ROOT

kill prev next Commands Views a c n s x

lock F-structure #1

"περπατώ Vb Mn Id Pr 01 Sg Xx Ip Av Xx πάνω_σε_τεττωμένος_σκοινί Ad Xx Ba"

```
PRED 'περπατώ<[1-SUBJ;PRO], [38;πάνω_σε_τεττωμένος_σκοινί]>'
SUBJ [ PRED 'PRO'
      CASE nom, NUM sg, PERS 1 ]
OBL-LOC 38 [ PRED 'πάνω_σε_τεττωμένος_σκοινί' ]
TNS-ASP [MOOD indicative, PERF -, TENSE pres]
1 [CLAUSE-TYPE decl, IDIOM risk, PASSIVE -
```

περπατώ (walk) with a LOC complement

```
1 valid F-structure for ROOT
kill prev next Commands Views a c n s x
lock F-structure #1
"περπατώ Vb Mn Id Pr 01 Sg Xx Ip Av Xx πάνω Ad Xx Ba σε As Pp Sp τεττωμένος Aj Ba Ne Sg Ac σκοινί No Cm Ne Sg Ac"

[PRED 'περπατώ<[1-SUBJ;PRO]>'
 [SUBJ [PRED 'PRO'
 [CASE nom, NUM sg, PERS 1]
 [ADJUNCT {
 [PCOMP [OBJ [ADJUNCT {
 [PRED 'τεττωμένος'
 [59[CASE acc, GEND neut, NUM sg]
 [NTYPE [NSYN common]
 [91[CASE acc, GEND neut, NUM sg, PERS 3]
 [50[PTYPE sem
 [38[PTYPE sem
 [TNS-ASP [MOOD indicative, PERF -, TENSE pres]
 [1[CLAUSE-TYPE decl, PASSIVE -
```

```
1 valid F-structure for ROOT
kill prev next Commands Views a c n s x
lock F-structure #1
"περπατώ Vb Mn Id Pr 01 Sg Xx Ip Av Xx πάνω_σε_τεττωμένος_σκοινί Ad Xx Ba"

[PRED 'περπατώ<[1-SUBJ;PRO], [38:πάνω_σε_τεττωμένος_σκοινί]>'
 [SUBJ [PRED 'PRO'
 [CASE nom, NUM sg, PERS 1]
 [OBL-LOC 38[PRED 'πάνω_σε_τεττωμένος_σκοινί']
 [TNS-ASP [MOOD indicative, PERF -, TENSE pres]
 [1[CLAUSE-TYPE decl, IDIOM risk, PASSIVE -
```

“NOT-SO-NATURAL” LFG ANALYSES



The fixed clitic is used non-referentially:

‘**την** έβγαλα **v** καθαρή’ = get away with

(gloss: her get away clean-ADJECTIVE)

‘**τα** έκανα **v** σαλάτα’ = make a mess

(gloss: them made salad-NOUN)

‘**την** περνάω **v** ζάχαρη’ = have a nice time with little effort

(gloss: her pass sugar-NOUN)

Filter output: **Fixed clitic_Verb-> one word (WWS) treated as a VERB**

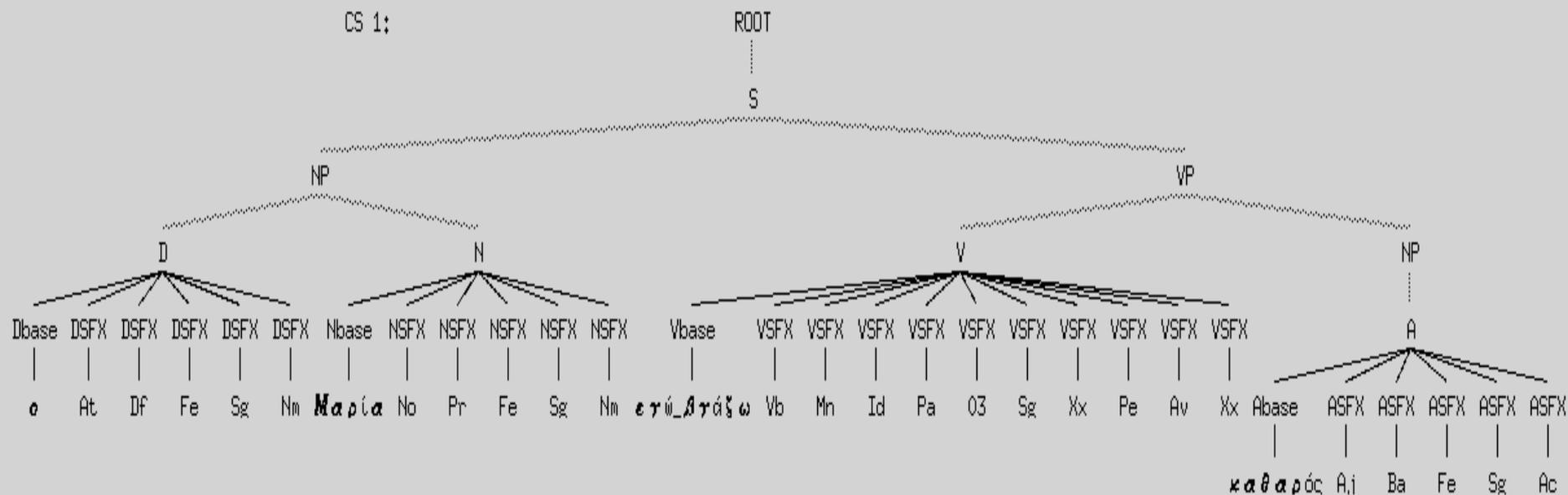
Alternatives:

transitive verb + object (! the non-referential clitic is the object normally)

copula + complement (! complement controlled by what?)

intransitive verb + OBL_manner (! the words do not denote manner normally)

"ο At Df Fe Sg Nm **Μαρία** No Pr Fe Sg Nm **ερώ_βράζω** Vb Mn Id Pa 03 Sg Xx Pe Av Xx **καθαρός** Aj Ba Fe Sg Ac"



1 valid F-structure for ROOT

kill prev next Commands Views a c n s x

lock F-structure #1

"ο At Df Fe Sg Nm **Μαρία** No Pr Fe Sg Nm **ερώ_βράζω** Vb Mn Id Pa 03 Sg Xx Pe Av Xx **καθαρός** Aj Ba Fe Sg Ac"

```

[
  PRED 'ερώ_βράζω<[29:Μαρία],[96:καθαρός]>'
  ]
[
  PRED 'Μαρία'
  NTYPE [NSYN proper]
  SUBJ
  SPEC [DET 1[CASE nom, DET-TYPE def, GEND fem, NUM sg, ο +]]
  29[CASE nom, GEND fem, NUM sg, PERS 3]
  ]
[
  PRED 'καθαρός'
  96[CASE acc, GEND fem, NUM sg]
  ]
  TNS-ASP [MOOD indicative, PERF +, TENSE past]
  59[CLAUSE-TYPE decl, IDIOM overcome_a_difficult_situation, PASSIVE - ]

```

To do...

We obtained (several...) natural analyses of the MWEs with the standing machinery of our MG LFG grammar

We will consider:

- a more sophisticated design of the filter
- semantic treatment of MWEs

Thank you!

*2nd PARSEME General Meeting
Athens, 2014, March, 10-11.*