Parsing Modern Greek verb MWEs in LFG/XLE

Niki Samaridi[1] and Stella Markantonatou[2]
[1]National and Kapodistrian University of Athens, nsamaridi@gmail.com
[2]Institute for Language and Speech Processing/ "Athena" RIC, marks@ilsp.gr

The parsing system

ILSP FBT Tagger
- Brill tagger plus rules
- 584 PAROLE tags (lemma and set of tags)

MWE recognizer (filter):
- Filter lexicon
- Filtering algorithm
- Formatter

Deep grammar (LFG/XLE)
General grammar of Modern Greek

The Filter (1): Filter lexicon: each MWE entry is specified for:

1. **Compositionality**: compositional/non-compositional interpretation
2. **'Signifier'**: lemma instructing the filter to look at the right filter lexicon entries
3. **Lemmatised form of Words_With_Spaces (WWS)**
4. **PoS and morphological constraints for the headword of a WWS**
5. **Constraints on the lemmatised forms of the remaining constituents of a WWS** that uniquely identify fixed or semi-fixed MWE substrings

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

compositionality   signifier   WWS   PoS & morphological constraints   →   constraints of the remaining constituents of a WWS

The filter (2): The filtering algorithm: (implemented in Perl)

A. No signifier: send to formatter
B. Signifiers: Any WWSs? Does string morphology match lexicon entries?
B1. No matching: send to formatter.
B2. Matching: Compositional reading? Yes: both send to the formatter and parse for WWSs.
B3. Matching: No compositional reading? Replace WWSs.
C. Replace WWS relevant substring with WWS & morphological constraints. Send to formatter.

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
Non-compositional	**περπατώ** Vb Mn Id Pr 01 Sg Xx Ip Av Xx **πάνω_σε_τεντωμένος_σκοινί** Ad Xx Ba
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

An example

8&9. Free subject (controller)-verb-object- subordinated clause with controlled subject: inflecting verb, object possibly fixed (Table 1: 9), the subordinated clause possibly semi-fixed (Table 1:8), intervening XPs, VSO/OVS word orders.

(8) Έριξαν   άδεια να πιάσουν γεμάτα
    throw-3-pl-past empty to catch-3-pl full
    'They tried to obtain information.'

(9) Έκανε   η   μάνα   του μαύρα μάτια να τον δει
    make-3-sg-past the mother-sg-nom his_j black eyes to him_j see-3-sg
    'It took his mother a long time to meet him.'

ρίχνω 'throw' is not a control verb but in (8) it behaves as such. To ensure identity of subjects, coordination could be considered alternatively. An "augmented" entry of the verb ρίχνω "throw" was eventually defined because να typically introduces (probably controlled) subordinated clauses and the constraints on verbal forms are are those of να-subordination and not of coordination.

Deep LFG analysis

Sublexical rules to parse the tagger output
- Only the lexicon has been manipulated (and not the grammar rules)
- Identification of phrasal constituents within the MWEs: permutations, ability of XPs to intervene between two words
- Grammatical functions are identified with diagnostics that apply to compositional expressions (morphological marking, WH-).

	LFG representation	Gloss	Meaning	WWS	C	FX
1	V: PRED παίρνω_πέντε	take_five	you are silly		Y	F
2	V: εγώ_**αρπάζω** <SUBJ >	(NP/pro)WEAK PRONOUN_grab	to be beaten up		Y	SF
3	V: **ρίχνω** <SUBJ,OBJ>, OBJ PRED= καρεκλοπόδαρο/ νερό_με_ο _τουλούμι	pour chair legs / water_with_the_goatskin_bag	it rains cats and dogs	νερό_με_ο_τουλούμι	N	SF
4	V: **έχω**_γούστο<COMP>, COMP COMPL=να	have_gusto to	don't tell me that	**έχω**_γούστο	N	SF
5	V: **μένω** <SUBJ,XCOMP>, XCOMP PRED=στήλη_άλας, XCOMP SUBJ=^SUBJ	(NP/pro) be left stele_of_salt	to be left speechless	στήλη_άλας	N	FL
6	V: **μένω**< SUBJ,XCOMP>, XCOMP PRED=ταπί_και_**ψύχραιμος**, XCOMP SUBJ=^SUBJ	(NP/pro) be left penniless_and_calm	to lose all one's money	ταπί_και _**ψύχραιμος**	N	FL
7	V: **τρώω/αρπάζω**<SUBJ,OBJ>, OBJ PRED=ο _ξύλο_ο _χρονιά, OBJ POSS PRED= εγώ, OBJ POSS TYPE= weak pronoun, OBJ POSS PERSON/NUMBER/GENDER =SUBJ PERSON/NUMBER/GENDER	(NP/pro)_i eat/grab the_beating_the_year _POSS_i	to be beaten up	ο_ξύλο_ο _χρονιά	N	SF
8	V: **ρίχνω** <SUBJ, OBJ, XCOMP>, XCOMP COMPL=για να/μήπως, OBJ PRED=άδειος, XCOMP PRED= **πιάνω**_γεμάτος, XCOMP SUBJ=SUBJ, XCOMP PERF=+, ¬(XCOMP TENSE)	(NP/pro)_i throw empty to pro_j catch_full	to try to obtain information	**πιάνω** _γεμάτος	N	FL
9	V: **κάνω** <SUBJ, OBJ, XCOMP>, XCOMP COMPL=να, OBJ PRED= μαύρος_μάτι, XCOMP PRED=**βλέπω** <SUBJ, OBJ>, OBJ PRED=εγώ, XCOMP SUBJ=SUBJ, XCOMP PERF=+, ¬(XCOMP TENSE)	(NP/pro)_i make black_eye to CLITIC pro_j see	it took SUBJ a very long time to meet CLITIC	μαύρος_μάτι	N	FL
10	V: **τραβάω**<SUBJ, OBJ>, OBJ PRED= ο_λινάρι_ο_πάθος	(NP/pro) suffer the_linen_ the_sufferings	to suffer a lot	ο_λινάρι _ο_πάθος	N	SF

Natural LFG Analyses

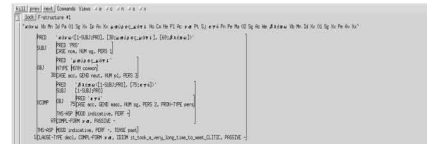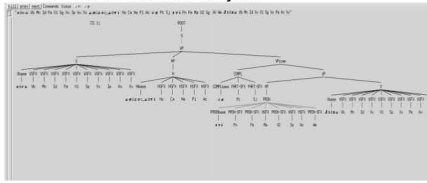Free subject (controller)-verb-object-subordinated clause with controlled subject

Fig. 1. The XLE output for the flexilbe MWE
έκανα μαύρα μάτια να σε δω, Gloss: I made black eyes to see you"
'It took me a very long time to meet you' (Table 1:9)

Verbs with "augmented" valency

"augmented" verb entries NOT in use otherwise:
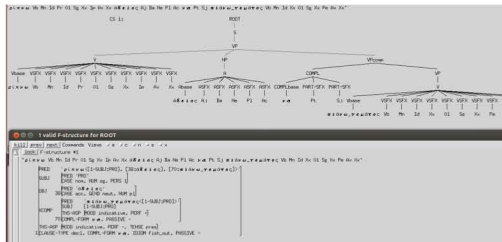**ρίχνω** (throw) with controlled sentential complement

Fig. 2. The XLE output for the flexilbe MWE
ρίχνω άδεια να πιάσω γεμάτα, Gloss: "I throw empty to catch full"
'I try to obtain information' (Table 1:8)

περπατώ (walk) with LOC argument denoting where one walks

Fig. 3. The XLE output for the flexilbe MWE
περπατώ πάνω σε τεντωμένο σκοινί, Gloss: I walk on a tightrope
'I am in a risky situation'

"Not-so-natural" LFG analysis

The fixed clitic is used non-referentially:
την έβγαλα √ καθαρή = 'get away with'
(gloss: her take_out clean-ADJECTIVE)
τα έκανα √ σαλάτα = 'make a mess'
(gloss: them made salad-NOUN)
√; where an XP may intervene

Filter output:
Fixed clitic_Verb-> one word (WWS) treated as a VERB
Alternatives:
- transitive verb + object (! the non-referential clitic is the object)
- copula + complement (! complement controlled by what?)
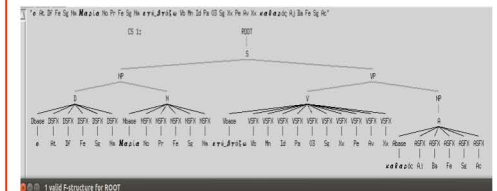- intransitive verb + OBL_manner (! the words do not denote manner normally)

Fig. 4. The XLE output for the flexilbe MWE την βγάζω καθαρή

References

Attia, Mohammed A. 2006. 'Accommodating Multiword Expressions in an Arabic LFG Grammar'. Salakoski, Tapio, Ginter, Filip, Pahikkala, Tapio, Pyysalo, Tampo. *Lecture Notes in Computer Science: Advances in Natural Language Processing, 5th International Conference, FinTAL. Turku. Finland.* Vol. 4139: 87-98. Springer-Verlag. Berlin Heidelberg

Fotopoulou, Aggeliki. 1993. *Une Classification des Phrases a Complements Figes en Grec Moderne.* Doctoral Thesis, Universite Paris VIII.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger. 2001. 'Multiword Expressions: A Pain in the Neck for NLP'. LinGO Working Paper No. 2001-03. In Alexander Gelbukh, ed., (2002) *Proceedings of CICLING-2002.* Springer.