

A MORPHOLOGICAL GRAMMAR FOR MODERN GREEK: STATE OF ART, EVALUATION AND UPGRADE

Lena Papadopoulou

Hellenic Open University

papadopoulou.lena@gmail.com

Elina Chadjipapa

Democritus University of Thrace

elinaxp@hotmail.com



CONTENT

- Background
- Corpus
- Lexicographic data & Processing
- Results
- Future work

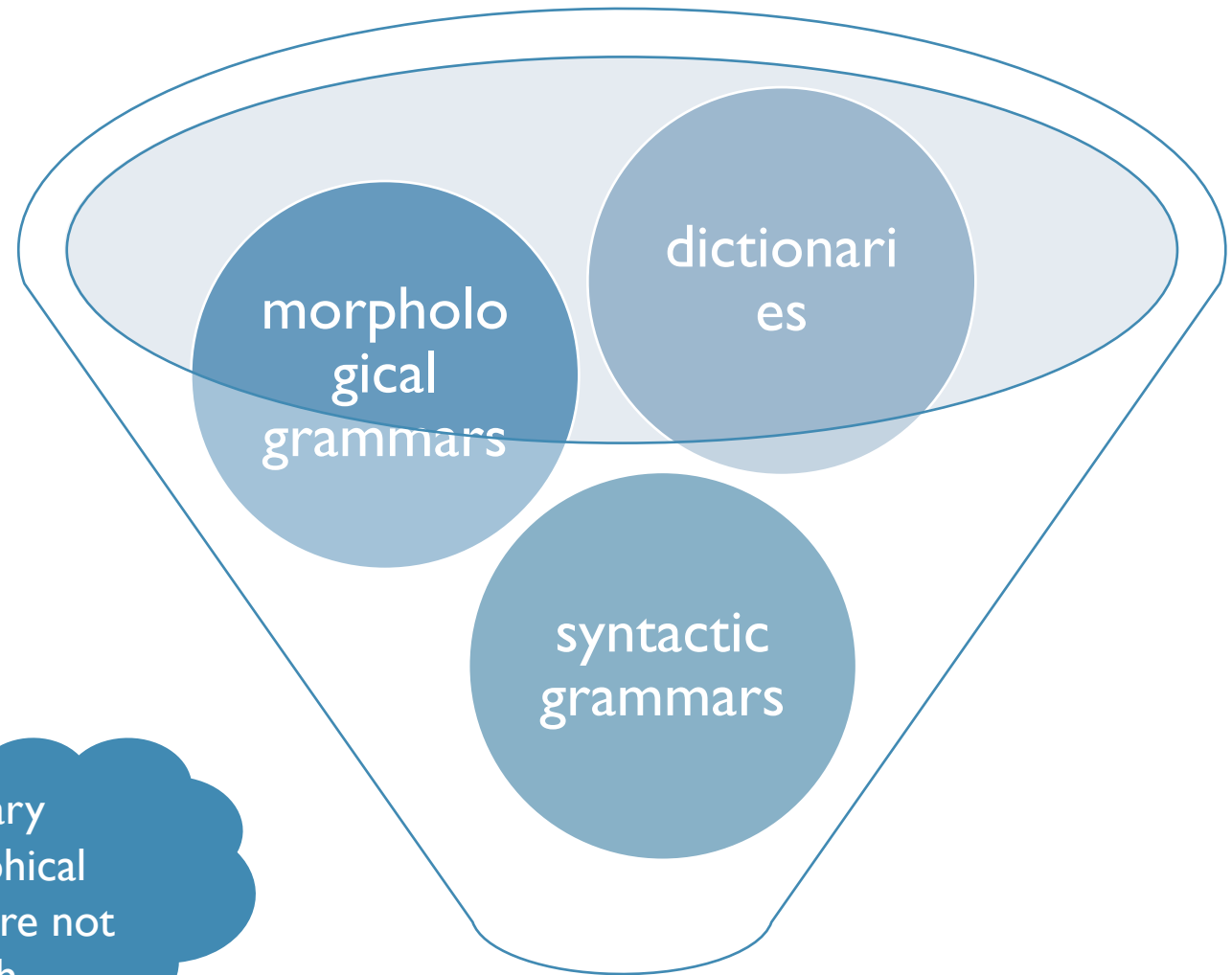
GREEK NOOJ SINCE 2007



language learning



secondary
lexicographical
resources are not
enough



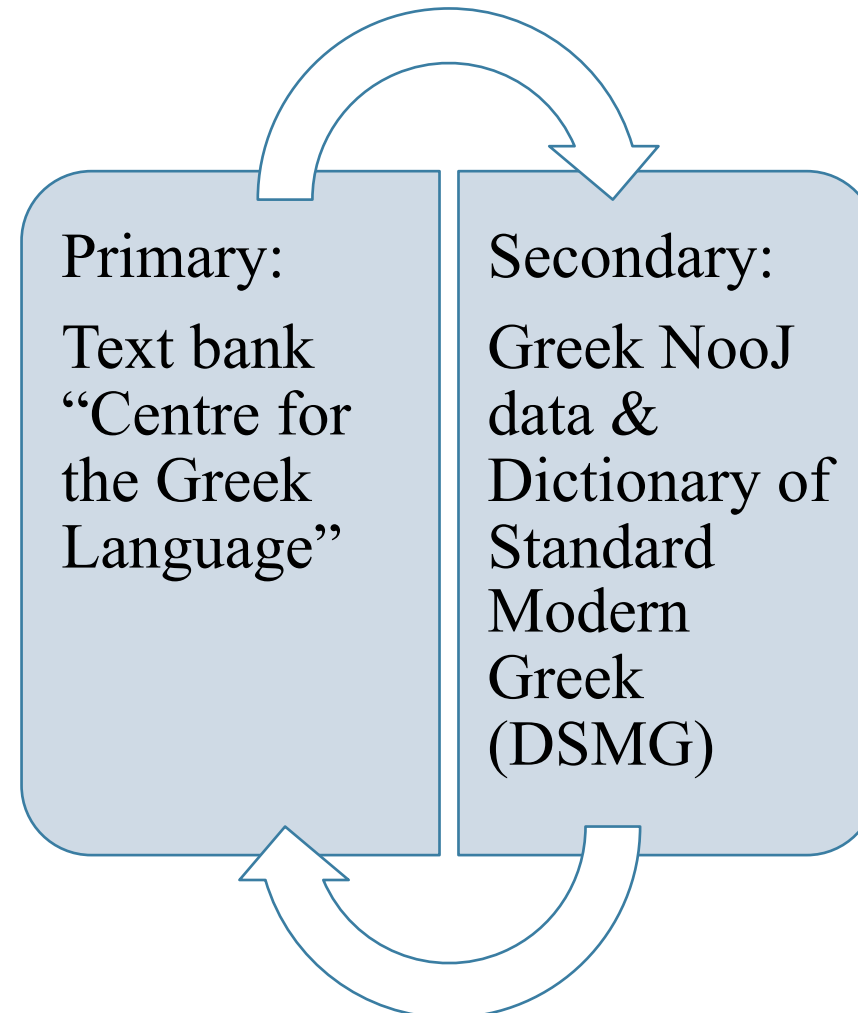
Greek language automatic
processing

**TEXT DATABANK OF THE
“CENTRE FOR THE GREEK LANGUAGE”**

- 6 text files: one for each language level (CEFR)
- 333 text units
- 128.694 tokens

<http://www.greek-language.gr/certification/dbs/teachers/index.html>

VALIDATION_NOUNS



linguistic analysis on
resources:

GLE.dic

FLEXIONES.nof

Acentuación_doble.nom

- Unknowns: 2.660
 - 175 text errors
- 1.101 noun word forms
 - 133 toponyms
 - 203 anthroponyms



Exportation of annotations
referring to nouns

- 3.464 nouns with
their inflectional
codes

1/4 PROCESS

Exclusion of forms:

- encountered only within multiword units
 - i.e. *Αγία Σοφία* EN: Hagia Sophia
- belonging to lexical units that are not used in our corpus
 - i.e. *κρατών*
 - a. nominalized participle *κρατών_{nom+s}* EN: prisoner
 - b. noun *κράτος- κρατών_{gen+p}* EN: state
- belonging to other part of speech
 - i.e. *βάρβαρος* (adjective) EN: barbarian
 - πιστεύω* (verb) EN: believe (noun: belief)
 - εγώ* (pronoun) EN: I (noun: ego)

2/4 PROCESS

- correspondence among lemmas, inflectional codification and inflectional paradigms
- alignment with inflectional categorization proposed by DSMG:

i.e.: 3:20 paradigms

O19	άγγελος	αγγέλου	άγγελο	άγγελε	άγγελοι	αγγέλων	αγγέλους	άγγελοι
O20	αντίλαλος	αντίλαλου	αντίλαλο	αντίλαλε	αντίλαλοι	αντίλαλων	αντίλαλους	αντίλαλοι
O20α	υπόκοσμος	υποκόσμου	υπόκοσμο	υπόκοσμε	υπόκοσμοι	υποκόσμων	υποκόσμους	υπόκοσμοι
		υπόκοσμου				υπόκοσμων	υπόκοσμους	

#άγγελος#

N19 = <E>/nom+m+s + <L6><A><R3><Á><RW>v/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <L6><A><R3><Á><RW><B2>ωv/gen+m+p + <L6><A><R3><Á><RW>υς/acc+m+p + ι/voc+m+p;

#Ιανουάριος#

N19_a = <E>/nom+m+s + <L4><A><R2><Á><RW>v/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <L4><A><R2><Á><RW><B2>ωv/gen+m+p + <L4><A><R2><Á><RW>υς/acc+m+p + ι/voc+m+p;

#θόρυβος#

N19_b = <E>/nom+m+s + <L5><A><R2><Á><RW>v/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <L5><A><R2><Á><RW><B2>ωv/gen+m+p + <L5><A><R2><Á><RW>υς/acc+m+p + ι/voc+m+p;

#Μάρτιος#

N19_c = <E>/nom+m+s + <L5><A><R3><Á><RW>v/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <L5><A><R3><Á><RW><B2>ωv/gen+m+p + <L5><A><R3><Á><RW>υς/acc+m+p + ι/voc+m+p;

#έλεγχος#

N19_d = <E>/nom+m+s + <L6><A><R2><Á><RW>v/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <L6><A><R2><Á><RW><B2>ωv/gen+m+p + <L6><A><R2><Á><RW>υς/acc+m+p + ι/voc+m+p;

#Δεκέμβριος#

N19_e = <E>/nom+m+s + <L6><A><R4><Á><RW>v/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <L6><A><R4><Á><RW><B2>ωv/gen+m+p + <L6><A><R4><Á><RW>υς/acc+m+p + ι/voc+m+p;

#συνάνθρωπος#

N19_f = <E>/nom+m+s + <L7><A><R4><Á><RW>v/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <L7><A><R4><Á><RW><B2>ωv/gen+m+p + <L7><A><R4><Á><RW>υ/acc+m+p + ι/voc+m+p;

#Αύγουστος#

N19_g = <E>/nom+m+s + <L7><A><R2><Á><RW>v/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <L7><A><R2><Á><RW><B2>ωv/gen+m+p + <L7><A><R2><Á><RW>υ/acc+m+p + ι/voc+m+p;

#πρόεδρος#

N19_h = <E>/nom+m+s + <L5><A><R><Á><R4>v/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <L5><A><R><Á><R4><B2>ωv/gen+m+p + <L5><A><R><Á><R3>υ/acc+m+p + ι/voc+m+p;

#λαχανόκηπος#ΛΚΝαντίλαλος#

N20 = <E>/nom+m+s + υ/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <B2>ων/gen+m+p + <B2>ουζ/acc+m+p + ι/voc+m+p;

#ιππόκαμπος#ΛΚΝυπόκοσμος#

N20a = <E>/nom+m+s + υ/gen+m+s + <L6><A><R2><Á><RW>υ/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <B2>ων/gen+m+p + <L6><A><R2><Á><RW><B2>ων/gen+m+p + <L6><A><R2><Á><RW>υζ/acc+m+p +
υζ/acc+m+p + ι/voc+m+p;

#παράδεισος#

N20a_a = <E>/nom+m+s + υ/gen+m+s + <L6><A><R3><Á><RW>υ/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <B2>ων/gen+m+p + <L6><A><R3><Á><RW><B2>ων/gen+m+p + <L6><A><R3><Á><RW>υζ/acc+m+p +
υζ/acc+m+p + ι/voc+m+p;

#μονόλογος#

N20a_b = <E>/nom+m+s + υ/gen+m+s + <L5><A><R2><Á><RW>υ/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <B2>ων/gen+m+p + <L5><A><R2><Á><RW><B2>ων/gen+m+p + <L5><A><R2><Á><RW>υζ/acc+m+p +
υζ/acc+m+p + ι/voc+m+p;

#οδοντίατρος#

N20a_c = <E>/nom+m+s + υ/gen+m+s + <L5><A><R><Á><RW>υ/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <B2>ων/gen+m+p + <L5><A><R><Á><RW><B2>ων/gen+m+p + <L5><A><R><Á><RW>υζ/acc+m+p +
υζ/acc+m+p + ι/voc+m+p;

#απόηχος#

N20a_d = <E>/nom+m+s + υ/gen+m+s + <L4><A><R><Á><RW>υ/gen+m+s + /acc+m+s + <B2>ε/voc+m+s +
ι/nom+m+p + <B2>ων/gen+m+p + <L4><A><R><Á><RW><B2>ων/gen+m+p + <L4><A><R><Á><RW>υζ/acc+m+p +
υζ/acc+m+p + ι/voc+m+p;

MANUAL

- Information for encoding correspondence for:
 - Inflectional irregularities
 - i.e. `_gp`
 - Proper nouns
 - Nominalized
 - verbs
 - adjectives
 - pronouns and
 - Particles
- i.e. nominalized perfect passive participle masculine: N19

3/4 PROCESS

Simplification inflectional paradigms' internal structure

NEW

#επισκέπτης#

N10_a = <E>/nom+m+s + /gen+m+s + /acc+m+s + /voc+m+s +
<L̄>ε/nom+m+p + <L4><A><RW><B2>ών/gen+m+p + <B2>ες/acc+m+p + <B2>ες/voc+m+p;

PREVIOUS

#ΑΦΕΝΤΗΣ

N10_a = <E>/nom+m+s + /gen+m+s + /acc+m+s + /voc+m+s +
<L>ε/nom+m+p + <L4><A><R2><RW><B2>ών/gen+m+p + <L>ε/acc+m+p +
<L>ε/voc+m+p;

4/4 PROCESS

~~non active paradigms~~

158 → 165

RESULTS

- Dynamic morphological grammar for simple nouns based both on primary and secondary lexicographic resources
- Easier attribution of the inflectional codification of new lemmas
 - 947 new lemmas – 0 new inflectional codes

FUTURE WORK

- Other parts of speech
- Lemmas as lexical units
- Improvement of linguistic analysis
 - Homonymy
 - Polysemy
 - Phrasemes
- Study of lexical richness & core vocabulary

**THANK YOU FOR
YOUR ATTENTION!**