

A finite-state treatment of neoclassical compounds in Modern Greek & English

Evanthia Petropoulou, Eleni Galiotou & Angela Ralli

Neoclassical compounds at a glance

- English

microcosm, atmosphere, astronaut, insecticide
pharyngolaryngitis, economical, carnivorous,
oceanographer, hamburgerology, queenomania,
jazzophile

- Modern Greek

καρδιολόγος (kardio`logos) “cardiologist”

δημοσιογράφος (dimosio`grafos) “journalist”

αιμορραγία (emora`gia) “haemorrhage”

σαρκοβόρος (sarko`voros) “carnivorous”

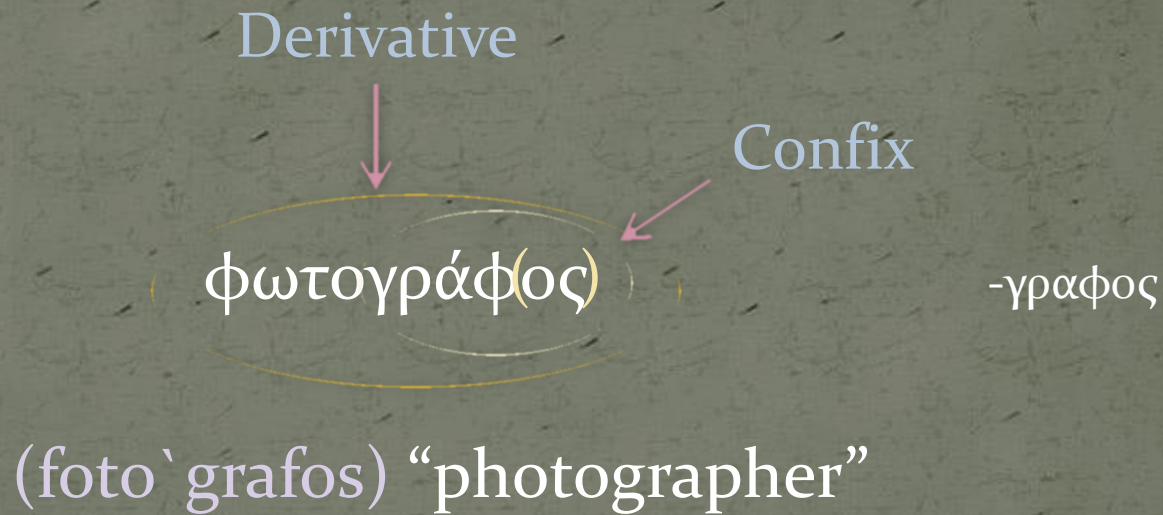
Free words vs. Bound elements

micro-, -cosm, atmo-, **sphere**, **insect**, -(i)cide,
pharyngo-, **laryngitis**, eco-, -nomical, carni-,
-vorous, **ocean**, -(o)grapher, **hamburger**, -(o)logy,
queen, **mania**, **jazz**, -(o)phile

καρδι(ά), -(ο)λόγ(ος), **δημόσι(ος)**, -(ο)γράφ(ος),
αίμ(α), -ρραγί(α), **σάρκ(α)**, -(ο)βόρ(ος)

How to analyse them? (MG)

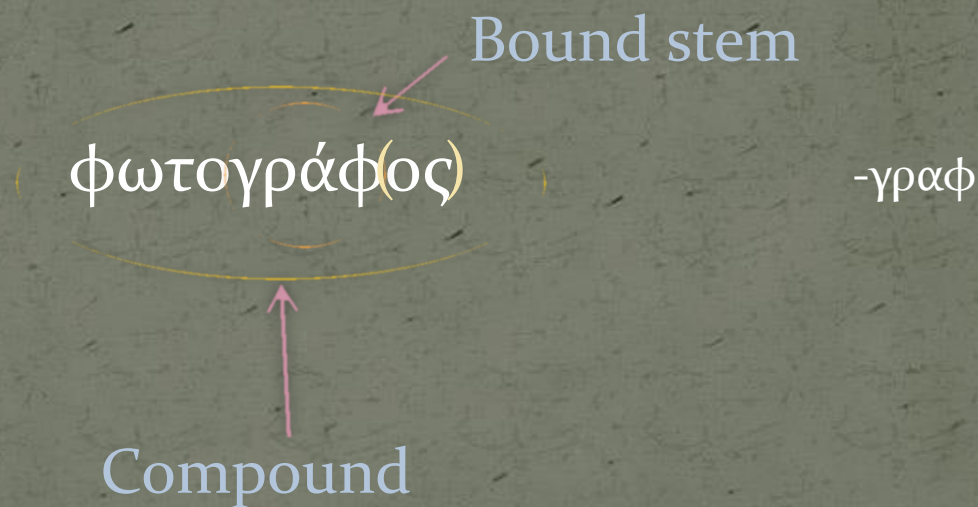
- Confixes



(Giannouloupoulou 2000)

How to analyse them?

- Bound stems



(Ralli 1988, 1992, 2008, 2009, 2013)

Arguments for Bound stems

- Serve as bases to prefixed words

E.g. υπέρμαχος (i`permachos) “firm supporter”
υπόλογος (i`pologos) “responsible for one’s
actions”

- Appear in compound structures, because:

i) –ο- is a Compound Marker (e.g. φωτ-ο-γραφ(ος))

ii) exhibit recursion (e.g. ζω-ο-φωτογραφία
(zoofotogra`fia)
“animal photo”)

Arguments for Bound stems

- Serve as bases for Derivatives

E.g. φωτογραφία_N, φωτογραφίζω_V, φωτογραφικός_A
(fotogra`fia) (fotogra`fizo) (fotografi`kos)
“photograph/ “take a photo” “photographic”
photography”

(Ralli 2008, 2009)

How to analyse them? (Eng)

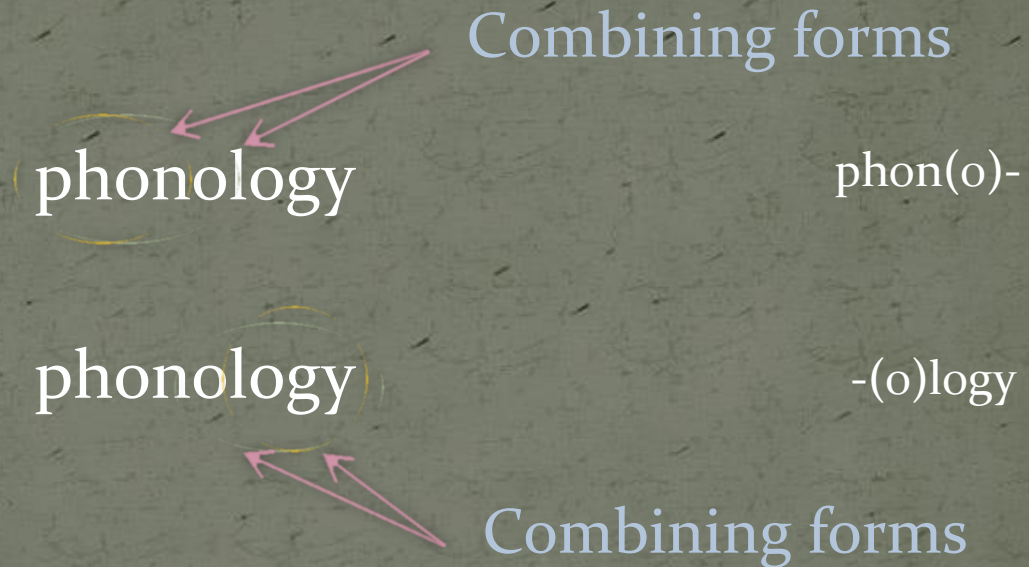
- Prefixes/Suffixes



(Scalise 1986)

How to analyse them?

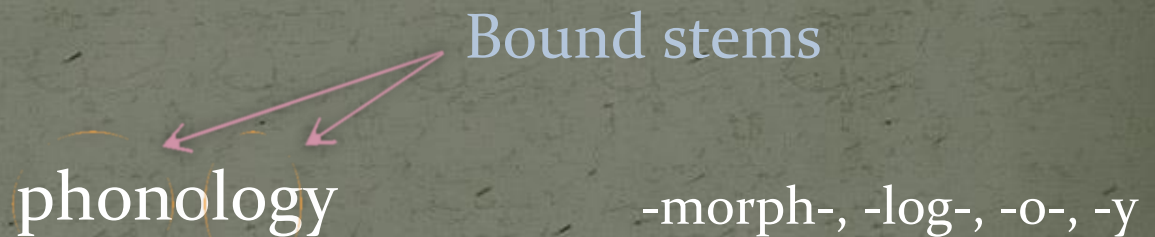
- Combining forms



Bauer (1983, 1998), Adams 1973, Warren (1990), Cannon (1992), Lehrer (1998) και Prčić (2005, 2008)

How to analyse them?

- Bound stems



e.g. logotherapy, logarithm, logic, ...
polyphonic, telephone, microphone, cacophony, ...

Scalise (1986), ten Hacken 1994, Adams 2001,
Baeskow 2004, Kastovsky 2009, Petropoulou 2009)

The Prototype

- Neoclassical Compounds

E.g. [micro][scope], [astr]o[naut],

- Neoclassical Compound Derivatives

E.g. [[bi]o[log]]+y, [[ge]o[graph]]#er,

[[anthrop]o[morph]]+ic, [[astr]o[log]]#er

(Baeskow 2004)

Neoclassical Compounds (MG, Eng)

- Compound structure
- Bound stem status of constituents
- Recognition of -o- as a linking element
- Nominal characteristics of bound stems

Combining forms *vs.* Bound stems

-(o)logy

-(o)graphy

-(o)logical

-(o)graphical

astro-

electro-

ous

hydro-

-cratic

-cracy ...

-scop-

-graph-

-log-

-phon-

-hydr-

-electr-

-phil-

-crat/crac-

-morph- ...

+-y

#-er

+-ic

+-

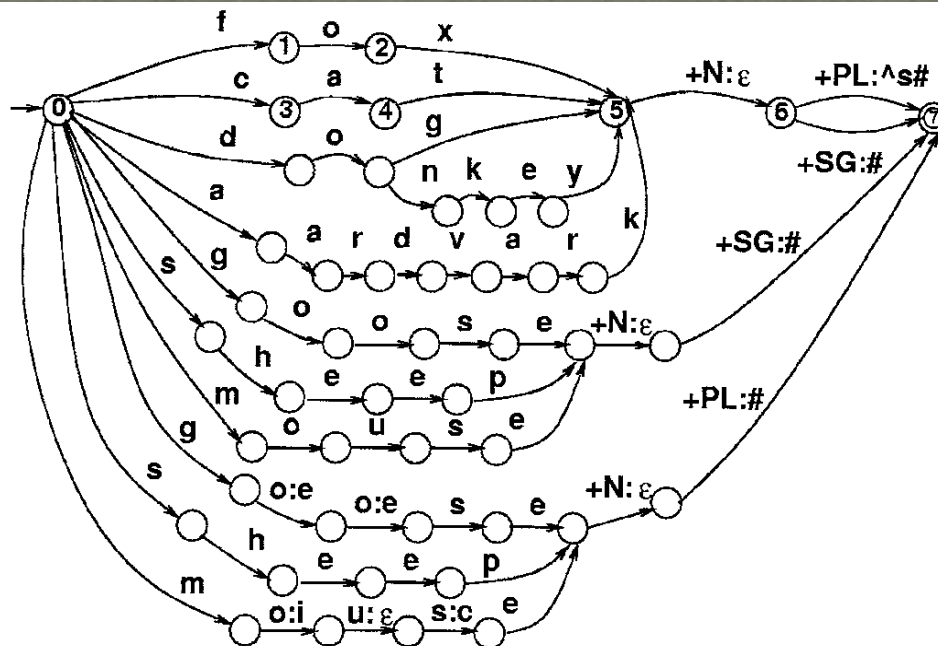
Finite-state methods

Lexical

c	a	t	+N	+PL		
---	---	---	----	-----	--	--

Surface

c	a	t	s			
---	---	---	---	--	--	--



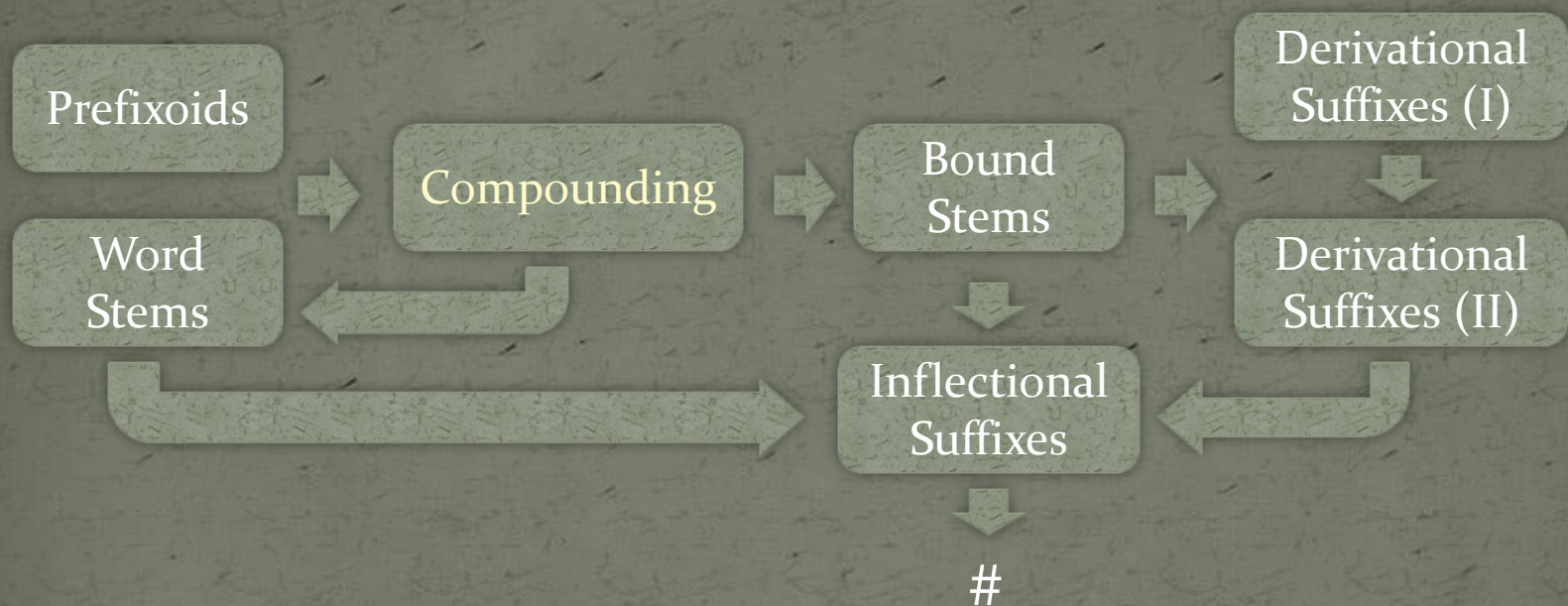
(Jurafsky & Martin 2000)

XEROX Lexical Compiler (LEXC)

- A high-level declarative language and associated compiler for defining finite-state automata and transducers; particularly well-suited for defining natural-language lexicons.
- A kind of right-recursive phrase-structure grammar.
- Continuation classes are the basic mechanism for describing morphotactics in LEXC and translate into concatenation.

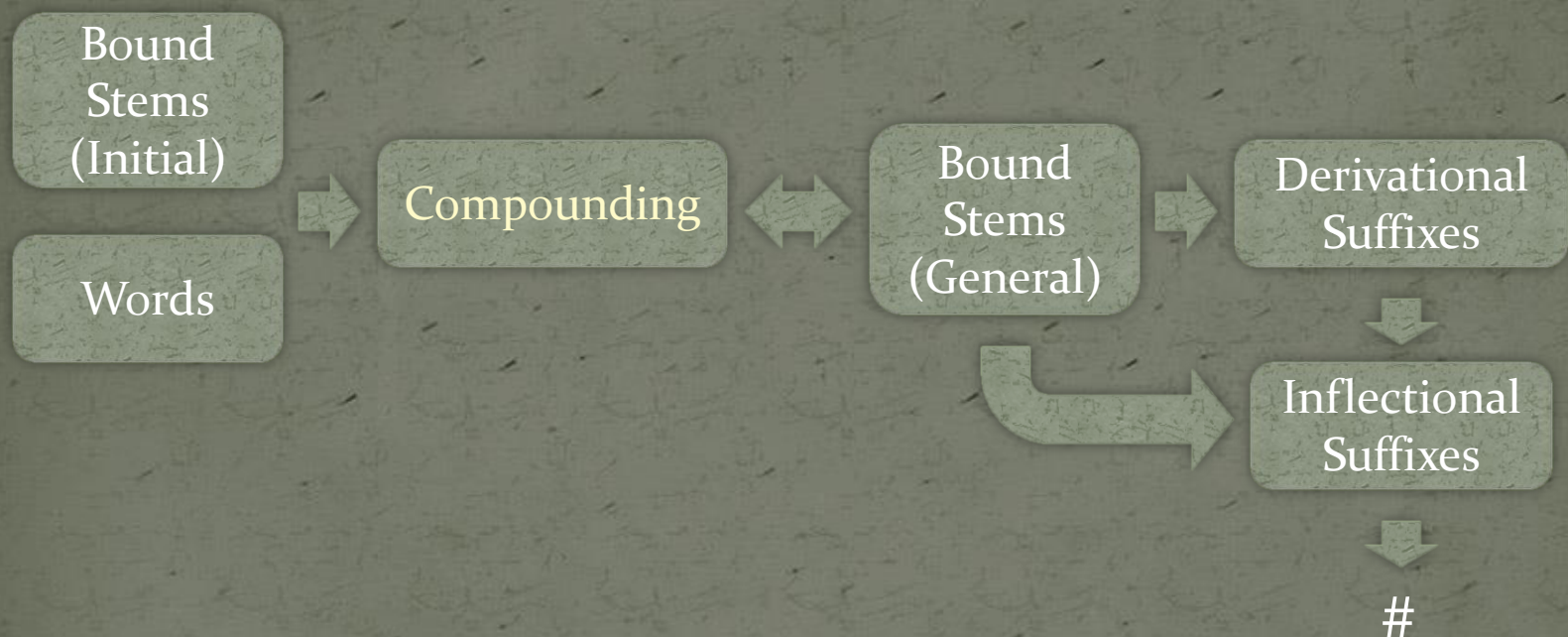
(Beesley & Karttunen 2003)

Modern Greek LEXC Description



E.g. [vivli]o[de]+ti(s) “book binder”
[vivli]o[de]+s+i(a) “book binding”
[vivli]o[pol](is) “bookseller”
[vivli]o[pol]+i(o) “book store”

English LEXC Description



E.g. [tele][graph]
[[film]o[graph]]+y
[[phot]o[graph]]
[[phot]o[graph]]#er

LEXICON Compound

<"^CM^":o "@U.VOWEL.OFF@">

General;

<"^CM^":0 "@U.VOWEL.ON@">

General;

<"^CM^":o "@U.VOWEL.OFF@">

Initial; (Eng.)

"^CM^":0 "@U.VOWEL.OFF@">

Initial;

<"^CM^":o "@U.VOWEL.OFF@">

BoundStems;

<"^CM^":0 "@U.VOWEL.ON@">

BoundStems; (MG)

<"^CM^":o "@U.VOWEL.OFF@">

WordStems;

<"^CM^":0 "@U.VOWEL.OFF@">

WordStems;

Flag Diacritics

- βιολόγ(ος) (vio`logos) “biologist”
<"@U.IC1A.ON@" @U.IC7.OFF@" @U.IC8.OFF@"
"+Masc":o"+Sg":o"+Nom":ος> #;
- δάσ(ος) (`dassos) “forest”
<"@U.IC7.ON@" @U.IC1A.OFF@" @U.IC8.OFF@"
"+Neut":o"+Sg":o"+Nom":ος> #;
- σώματ(ος) (`somatos) “of the body”
<"@U.IC8.ON@" @U.IC7.OFF@" @U.IC1A.OFF@"
"+Neut":o"+Sg":o"+Gen":ος> #;

LEXC Testing

- Correct distribution of the linking element

apply up> psychopaths

psych^CM^path+Noun+Pl

apply up> psychpaths

???

- Selection of the appropriate suffix

apply up> monogamy apply up>

homophobia

mon^CM^gam+Noun+Sg

hom^CM^phob+Noun+Sg

apply up> monogamia apply up> homophoby

???

???

LEXC Testing

apply up> polygonic

poly^CM^gon+Adj+Sg

poly^CM^gon+Adj+Pl

apply up> polygonous

poly^CM^gon+Adj+Sg

poly^CM^gon+Adj+Pl

apply up> polygonal

poly^CM^gon+Adj+Sg

poly^CM^gon+Adj+Pl

apply up> autophagous

aut^CM^phag+Adj+Pl

aut^CM^phag+Adj+Sg

apply up> autophagic

aut^CM^phag+Adj+Pl

aut^CM^phag+Adj+Sg

apply up> autophagal

???

LEXC Testing

- Selection of appropriate allomorphs

apply up> homicide apply up> homicidal

homi^CM^cid+Noun+Sg ???

apply up> homicidious apply up> homicidous

homi^CM^cid+Adj+Sg ???

homi^CM^cid+Adj+Pl

apply up> homicidal

homi^CM^cid+Adj+Sg

homi^CM^cid+Adj+Pl

LEXC Testing

- Recognition of recursive structures

apply up> biotechnology

bi^CM^techn^CM^log+Noun+Sg

apply up> sociomorphologist

soci^CM^morph^CM^log+Noun+Sg

apply up> electromyographic

electr^CM^myel^CM^graph+Adj+Sg

electr^CM^myel^CM^graph+Adj+Pl

apply up> osteomyopathy

oste^CM^my^CM^path+Noun+Sg

Future plans

- Automatic testing through application to large corpora, with the Lookup function of the LEXC
- Description of the same phenomenon in other languages