

Αξιοποίηση ηλεκτρονικών πόρων στην Ελληνική Φιλολογία-Digital Humanities (3):

ΛΕΞΙΚΟΙ ΠΟΡΟΙ (1)

Roget's, Ονομαστικόν, WordNet Ταξινομίες

Στέλλα Μαρκαντωνάτου

Υπολογιστική Γλωσσολόγος

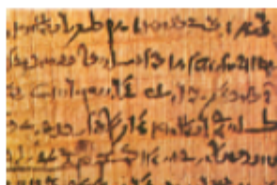
Διευθύντρια Ερευνών

Ινστιτούτο Επεξεργασίας του Λόγου/ΕΚ ΑΘΗΝΑ

stiliani.markantonatou@gmail.com

General process

Raw data (analog)



Converting analog to digital
(e.g., Scanning + OCR,
Recording + Voice-to-text)



Digital data



Analysis (e.g.,
machine learning,
NLP, Computer vision)



insights

Visualization



Graphic interface

To extend our example further, we might decide to **add a semantic linguistic analysis to each of the poems...**

```
<anthology>
  <!-- anthology markup elements here -->
  <line>
    <gram:noun>αλέκτορα</gram:noun>

    <gram:noun
wiktionary.relation=
"https://el.wiktionary.org/wiki/%CE%B1%CE%BB%CE%AD%CE%BA%CF%84%CF%89%CF%81"

enacademic.relation="http://www.enacademic.com/searchall.php?SWord=%CE%B1%CE%BB%CE%AD%CE%BA%CF%84%CF%89%CF%81&from=xx&to=en&did=&styp=0"

pyli.relation="none"/>

  </line>
</anthology>
```

....και αν, για να ενισχύσουμε την σημασιολογική αναπαράσταση, το «**αλέκτωρ**» θέλαμε να το συσχετίσουμε με άλλες λέξεις της Νέας Ελληνικής;

Θα χρειαζόμασταν ένα

ΣΗΜΑΣΙΟΛΟΓΙΚΟ ΛΕΞΙΚΟ

Ποιες σχέσεις μπορούμε να βρούμε στην Πύλη;



- Κότα;
- Κόττα;
- Όρνιθα;
- Όρνις;
- Κοτόπουλο;
- Κόκορας;
- Κόκκορας;
- Αλέκτωρ;
- Κλώσα;
- Κλωσόπουλο;
- Ωόν;
- Αυγό;
- Κοτέτσι;
- Ορνιθώνας;
- **Κλωσάω;**
- **Κακαρίζω;**
- Κακάρισμα;

Σήμερα μιλάμε για

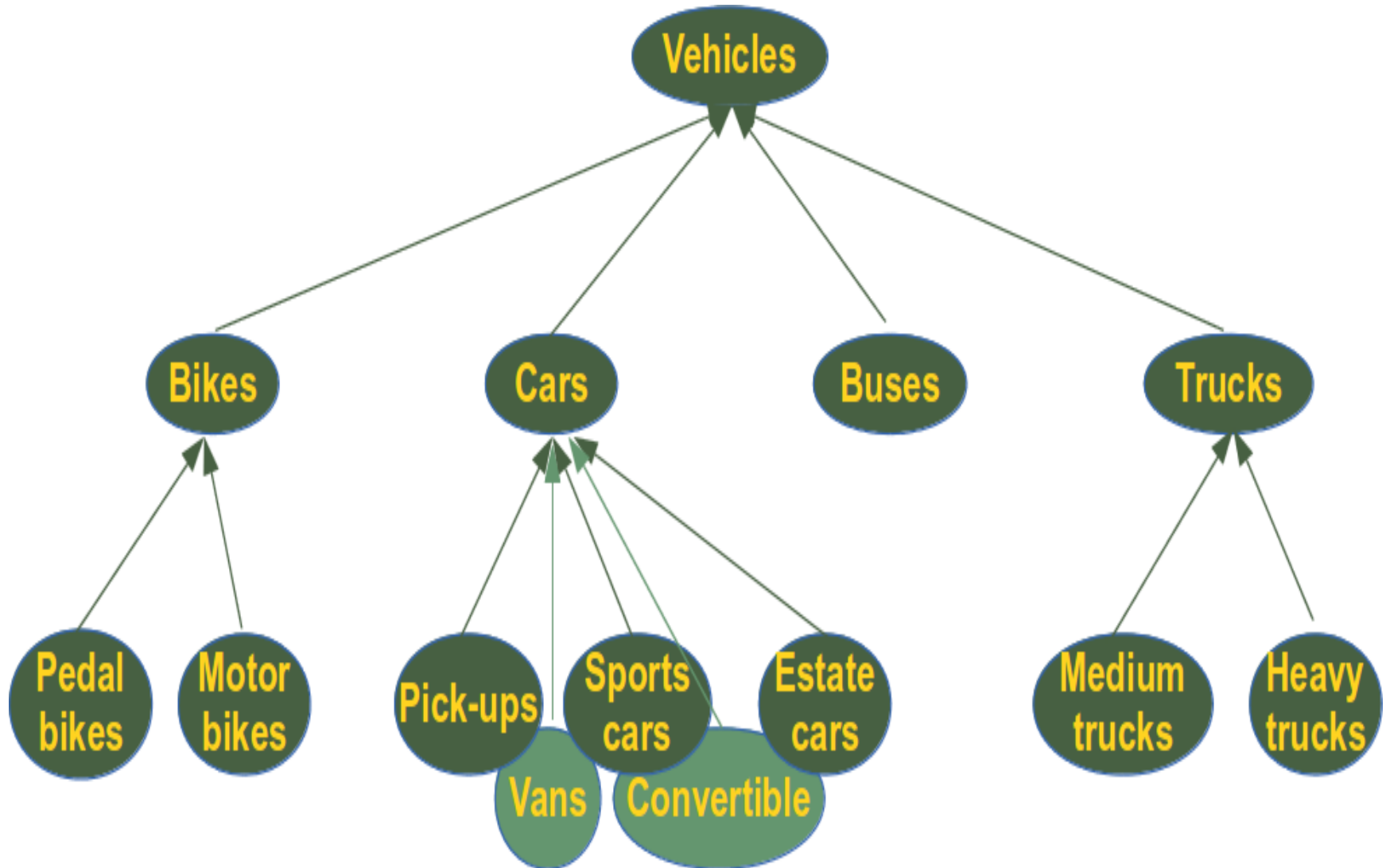
ΕΝΝΟΙΟΛΟΓΙΚΑ ΟΡΓΑΝΩΜΕΝΑ ΛΕΞΙΚΑ

που συμπεριλαμβάνουν πλήθος λεξικών
σχέσεων

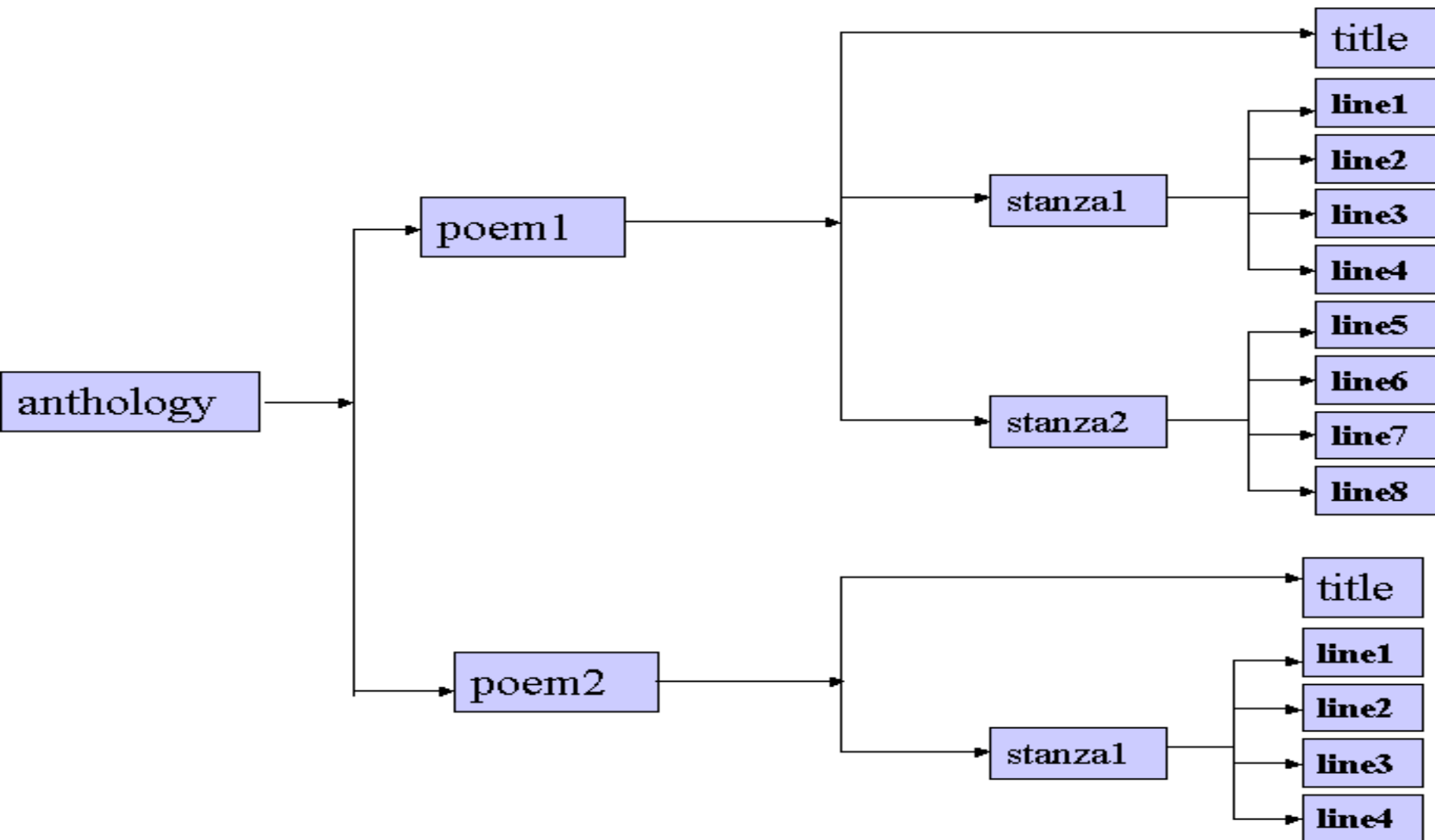
Μιλάμε για **ΛΕΞΙΚΕΣ ΟΝΤΟΛΟΓΙΕΣ**

Συχνά, η **συνωνυμία** και η **αντωνυμία** είναι
βασικές σχέσεις οργάνωσης μαζί με την σχέση
κληρονομικότητας (κληρονομικότητα κάποιου
τύπου)

Η σχέση κληρονομικότητας



XML: θα μπορούσε να θεωρηθεί ως ιεραρχική αναπαράσταση της κειμενικής δομής: σχέση μέρους-όλου



Η λογική προσέγγιση

**ΚΩΔΙΚΟΠΟΙΩΝΤΑΣ ΣΧΕΣΕΙΣ
ΓΕΝΟΥΣ-ΕΙΔΟΥΣ ΜΕΤΑΞΥ ΣΥΝΟΛΩΝ
ΟΜΟΕΙΔΩΝ ΟΝΤΟΤΗΤΩΝ**

ROGET'S THESAURUS

Publication History of Roget's Thesaurus

British

1779 Peter Mark Roget born
 1805 *Compilation of thesaurus (elements) begun*

	Editor	Publisher
1852 1st Edition published	Peter Mark Roget	Longman -- all editions
1853 2nd Edition	Peter Mark Roget	
1855 'enlarged and improved' 1856, 1857, 1858, 1864	Peter Mark Roget	
1869 <i>Death of Dr PM Roget - 28 editions to date</i>		
1879 John Lewis Roget Edition 'Greatly enlarged and improved' 1882, 1883	John Lewis Roget	
1890-1908 reprinted two-yearly	John Lewis Roget	

American

1854 1st American attempt
 Editor: Barnas Sears
 Publisher: Gould & Lincoln
 "the editor, among other mutilations, has altogether omitted the Phrases,
 ...and expressions borrowed from a foreign language" PMRoget

British - American 'split'

	Editor	Publisher
1911-1929 reprinted yearly	John Lewis Roget	
1925 'enlarged' - 5 reprints per year	Samuel Romilly Roget	
1933 Authorised American Edition 1936 New Edition 'revised and enlarged'	Samuel Romilly Roget	
1953 used in Masterman research <i>Sold to Penguin</i>	Samuel Romilly Roget	Longman, Greene
1962 'completely revised and modernised'	Robert Dutch	Penguin
1982 'new edition'	Susan Lloyd	
1987 'new edition'	Betty Kirkpatrick	
1998 'new edition'	Betty Kirkpatrick	
2002 '150th Anniversary Edition'	George Davidson	
		1886 1st International Edition
		1911 'practically a new book'
		1922 'revised and reset'
		1930-1939 5 editions 'enlarged and improved'
		1931 'in dictionary form'
		1946 2nd Int. Ed. "complete overhauling"
		1950-1960 3 editions
		1962 3rd International Edition
		1964-1966 2 editions <i>Sold to Harper</i>
		1977 4th International Edition
		1992 5th International Edition--new structure
		2001 6th International Edition--new structure

Roget's Thesaurus

Senses and synonyms of the word

- joy -

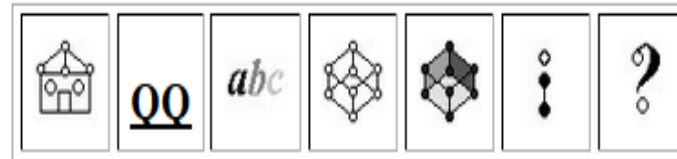
#827 **Pleasure:** Nn. [joy](#), [joyfulness](#), [joyance](#) -- [delight](#), [joy](#), [take great satisfaction](#).

#836 **Cheerfulness:** Nn. [joy](#), [joyfulness](#), [joyousness](#).

#838 **Rejoicing:** Vb. [rejoice](#), [jubilate](#), [exult](#), [glory](#), [joy](#), [delight](#), [bless one's stars](#), [thank one's stars](#), [congratulate oneself](#), [hug oneself](#), [rub one's hands](#), [clap hands](#), [throw up one's cap](#).

#884 **Boasting:** Vb. [triumph](#), [glory](#), [delight](#), [joy](#), [jubilate](#), [exult](#).

#887 **Blusterer:** Vb. [triumph](#), [glory](#), [delight](#), [joy](#), [jubilate](#), [exult](#).



Copyright 1999

Ιεραρχική αναπαράσταση της δομής

Έχει προστεθεί στην ηλεκτρονική εφαρμογή του Roget's μια νέα σχέση για τη δημιουργία των lattices: **τομή συνόλων**

Όχι ρητές σχέσεις μεταξύ ονομάτων και ρημάτων, πχ η (hen, chicken) **κότα** και το (cluck) **κακαρίζω** δεν συσχετίζονται

Concept Neighbourhoods in Roget's Thesaurus

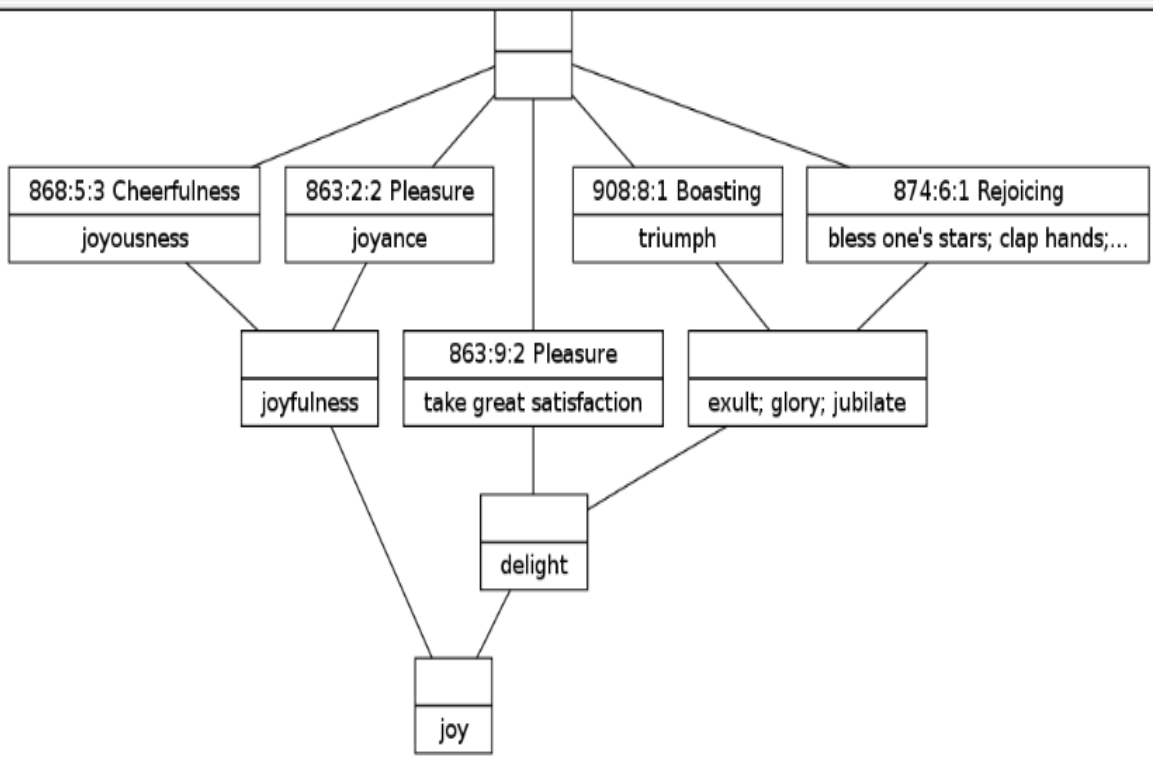
joy

- png
- svg
- cxt
- cex
- csc

submit

Type in any word to see the concept neighbourhood of the word in Roget's Thesaurus.

[Help and more information](#)



Ιεραρχική αναπαράσταση: Σύνδεση των ανεξάρτητων εννοιών και συνολική ανάσυρσή τους

Inbox (89) - stiliani.mark... x Stella Markantonatou sh... x Enriching Digitized Med... x Princeton WordNet 3.1 x ψη Λέξη: "μνησιπήμων" x Roget's Hy
www.roget.org/scripts/qq.php 90% Search
Most Visited Getting Started Admin Docs | Εσωτερ... IDION EDITOR PMWE 2 - Overleaf Multiword expressions... http://cultdict.ioperm...

Roget's Thesaurus

Senses and synonyms of the word

- joy -

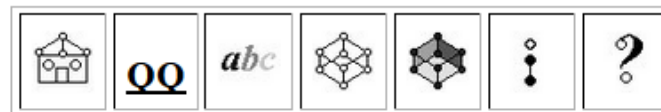
#827 Pleasure: Nn. joy, joyfulness, joyance -- delight, joy, take great satisfaction.

#836 Cheerfulness: Nn. joy, joyfulness, joyousness.

#838 Rejoicing: Vb. rejoice, jubilate, exult, glory, joy, delight, bless one's stars, thank one's stars, congratulate oneself, hug oneself, rub one's hands, clap hands, throw up one's cap.

#884 Boasting: Vb. triumph, glory, delight, joy, jubilate, exult.

#887 Blusterer: Vb. triumph, glory, delight, joy, jubilate, exult.



Copyright 1999

Θεολόγος Βοσταντζόγλου, 1962

«ΟΝΟΜΑΣΤΙΚΟΝ»

- Το Ελληνικό αντίστοιχο του Roget's
- Πιο πυκνό σημασιολογικό δίκτυο (από το αρχικό Roget's) με μέρη του λόγου και πολυλεκτικά και με δική του λεξική οντολογία
- Όχι ρητές σχέσεις μεταξύ ονομάτων και ρημάτων σε γεγονοτικό πλαίσιο, πχ η **κότα** και το **κακαρίζω** δεν συσχετίζονται

Τα έντυπα λεξικά θέτουν περιορισμούς χώρου

Ποιες σχέσεις μπορούμε να βρούμε στο Ονομαστικόν;



- Κότα;
- Κόττα;
- Όρνιθα;
- Όρνις;
- Κοτόπουλο;
- Κόκορας;
- Κόκκορας;
- Αλέκτωρ;
- Κλώσα;
- Κλωσόπουλο;
- Ωόν;
- Αυγό;
- Κοτέτσι;
- Ορνιθώνας;
- **Κλωσάω;**
- **Κακαρίζω;**
- Κακάρισμα;

WordNet

<https://wordnet.princeton.edu/>

WordNet

- **WordNet[®]**: μεγάλη βάση λεξικών δεδομένων της Αγγλικής
 - Ονόματα, ρήματα, επίθετα και επιρρήματα ομαδοποιούνται σε σύνολα **γνωσιακών συνωνύμων (synsets)**. Κάθε synset δηλώνει μία ξεχωριστή έννοια
 - Τα synsets διασυνδέονται με εννοιολογικές-σημασιολογικές και λεξικές σχέσεις
 - Συνολικά, δημιουργείται ένα δίκτυο λέξεων και εννοιών συνδεδεμένων με νοηματοδοτημένες σχέσεις
- Το WordNet έχει τη μορφή θησαυρού
 - Διασυνδέει συγκεκριμένες σημασίες των λέξεων → η στενή γειτνίαση των λέξεων μέσα στο δίκτυο οδηγεί στην άρση των σημασιολογικών αμφισημιών
 - Διαθέτει ένα πολύ μεγάλο σύνολο ετικετών για τις σημασιολογικές σχέσεις μεταξύ των λέξεων

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) [hen](#), [bidly](#) (adult female chicken)
 - [direct hyponym](#) / [full hyponym](#)
 - [S:](#) (n) [cackler](#) (a hen that has just laid an egg and emits a shrill squawk)
 - [S:](#) (n) [brood hen](#), [broody](#), [broody hen](#), [setting hen](#), [sitter](#) (a domestic hen ready to brood)
 - [S:](#) (n) [mother hen](#) (a hen with chicks)
 - [S:](#) (n) [layer](#) (a hen that lays eggs)
 - [S:](#) (n) [pullet](#) (young hen usually less than a year old)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - [S:](#) (n) [chicken](#), [Gallus gallus](#) (a domestic fowl bred for flesh or eggs; believed to have been developed from the red jungle fowl)
- [S:](#) (n) [hen](#) (adult female bird)
- [S:](#) (n) [hen](#) (flesh of an older chicken suitable for stewing)
- [S:](#) (n) [hen](#) (female of certain aquatic animals e.g. octopus or lobster)

Η σχέση γένους-είδους

- Η πιο συχνή σχέση ανάμεσα στα synsets είναι η σχέση **γένους-είδους, κληρονομικότητας (υπερωνυμίας-υπωλυμίας / super-subordinate relation, hyperonymy, hyponymy, ISA relation)**.
- Από τα γενικά synsets {furniture, piece_of_furniture} πηγαίνουμε στα ειδικά {bed}, {bunkbed}.
- Δηλαδή, το WordNet δηλώνει ότι η έννοια **furniture** περιλαμβάνει την έννοια **bed** που με την σειρά της περιλαμβάνει την έννοια **bunkbed**.
- Η σχέση γένους-είδους είναι μεταβατική: εάν το μπαουλοντίβανο είναι ένα είδος κρεβατιού και το κρεβάτι ένα είδος επίπλωσης τότε το μπαουλοντίβανο είναι ένα είδος επίπλωσης.
- Τελικά όλες οι ονοματικές ιεραρχίες συναντώνται στον αρχικό κόμβο {entity}.

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) [hen](#), [bidly](#) (adult female chicken)
 - [direct hyponym](#) / [full hyponym](#)
 - [S:](#) (n) [cackler](#) (a hen that has just laid an egg and emits a shrill squawk)
 - [S:](#) (n) [brood hen](#), [broody](#), [broody hen](#), [setting hen](#), [sitter](#) (a domestic hen ready to brood)
 - [S:](#) (n) [mother hen](#) (a hen with chicks)
 - [S:](#) (n) [layer](#) (a hen that lays eggs)
 - [S:](#) (n) [pullet](#) (young hen usually less than a year old)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - [S:](#) (n) [chicken](#), [Gallus gallus](#) (a domestic fowl bred for flesh or eggs; believed to have been developed from the red jungle fowl)
- [S:](#) (n) [hen](#) (adult female bird)
- [S:](#) (n) [hen](#) (flesh of an older chicken suitable for stewing)
- [S:](#) (n) [hen](#) (female of certain aquatic animals e.g. octopus or lobster)

Η σχέση μέρους-όλου (μερωνυμία)

Η μερωνυμία ισχύει μεταξύ synsets, όπως για παράδειγμα {seat} and {leg}.

Τα μέρη κληρονομούνται από τις ευρύτερες έννοιες: εάν το κάθισμα έχει πόδια τότε και η πολυθρόνα έχει πόδια!



κάθισμα



SEMANTIC RELATIONS

- **S: (n) body**, [organic structure](#) (the entire physical structure of an organism (an animal, plant, or human being)) *"he felt as if his whole body were on fire"*
 - [direct hyponym](#) / [full hyponym](#)
 - [part meronym](#)
 - **S: (n) body substance** (the substance of the body)
 - **S: (n) articulatory system** (the system of joints in the body)
 - **S: (n) cavity**, [bodily cavity](#), [cavum](#) ((anatomy) a natural hollow or sinus within the body)
 - **S: (n) digestive system**, [gastrointestinal system](#), [systema alimentarium](#), [systema digestorium](#) (the system that makes food absorbable into the body)
 - **S: (n) endocrine system** (the system of glands that produce endocrine secretions that help to control bodily metabolic activity)
 - **S: (n) lymphatic system**, [systema lymphaticum](#) (the interconnected system of spaces and vessels between body tissues and organs by which lymph circulates throughout the body)
 - **S: (n) musculoskeletal system** (the system of muscles and tendons and ligaments and bones and joints and associated tissues that move the body and maintain its form)
 - **S: (n) nervous system**, [systema nervosum](#) (the sensory and control apparatus consisting of a network of nerve cells)
 - **S: (n) respiratory system**, [systema respiratorium](#) (the system for taking in oxygen and giving off carbon dioxide; in terrestrial animals this is accomplished by breathing)
 - **S: (n) sensory system** (the body's system of sense organs)
 - **S: (n) vascular system** (the vessels and tissue that carry or circulate fluids such as blood or lymph or sap through the body of an animal or plant)
 - **S: (n) circulatory system**, [cardiovascular system](#) (the organs and tissues involved in circulating blood and lymph through the body)
 - **S: (n) head**, [caput](#) (the upper part of the human body or the front part of



Σχέσεις μεταξύ μερών του λόγου (λεξικές σχέσεις)

Cross-POS relations

Οι περισσότερες σημασιολογικές σχέσεις ορίζονται μεταξύ των ιδίων μερών του λόγου. Άρα το WordNet στην παραγματικότητα περιλαμβάνει 4 υπο-δίκτυα.

Οι λεξικές σχέσεις είναι κατά βάση μορφολογικές όπου η μορφολογία δεν έχει αλλάξει την σημασία.

Display Options: (Select option to change)

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- **S: (n) cluck**, [clucking](#) (the sound made by a hen (as in calling her chicks))
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - [derivationally related form](#)
 - **W: (v) cluck** [Related to: [cluck](#)] (make a clucking sounds, characteristic of hens)

Verb

- **S: (v) cluck**, [click](#), [clack](#) (make a clucking sounds, characteristic of hens)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (v) utter**, [emit](#), [let out](#), [let loose](#) (express audibly; utter sounds (not necessarily words)) "*She let out a big heavy sigh*"; "*He uttered strange sounds that nobody could understand*"
 - [derivationally related form](#)
 - **W: (n) cluck** [Related to: [cluck](#)] (the sound made by a hen (as in calling her chicks))
 - **W: (n) clack** [Related to: [clack](#)] (a sharp abrupt noise as if two objects hit together; may be repeated)
 - [sentence frame](#)
 - Something ----s
 - Somebody ----s

**LEXICAL
RELATIONS**

Ποιες σχέσεις μπορούμε να βρούμε στο WordNet;



- Κότα;
- Κόττα;
- Όρνιθα;
- Όρνις;
- Κοτόπουλο;
- Κόκορας;
- Κόκκορας;
- Αλέκτωρ;
- Κλώσα;
- Κλωσόπουλο;
- Ωόν;
- Αυγό;
- Κοτέτσι;
- Ορνιθώνας;
- **Κλωσάω;**
- **Κακαρίζω;**
- Κακάρισμα;

BalkaNet-Greek WordNet

- The Balkan WordNet aims at the development of a multilingual lexical database comprising individual WordNets for the Balkan languages. The most ambitious feature of the BalkaNet is its attempt to represent semantic relations between words in each Balkan language and link them in order to develop an on line multilingual semantic network. The main objective is the development of WordNets for each individual language from available resources covering the general vocabulary. Semantic relations will be classified in the independent WordNets according to a shared ontology. Then, all individual WordNets will be organized into a common database providing linking across them. Each of the WordNets will be structured along the same lines as the EuroWordNet through a WordNet Management System. This project is an excellent opportunity to explore the less studied Balkan languages and combine and compare them cross-linguistically.
- **Objectives**
- Investigate the less studied Balkan languages.
- Create the possibility to expand the EuroWordNet by adding to it Wordnets built for the Balkan languages.
- Give access to the European community to new sources of information stemming from the Balkan languages that are not so widely studied.
- Facilitate access to information written in less studied languages.
- Create and strengthen ties with the academic and information technology community in Balkan countries.
- Create a common large-scale linguistic resource.

Ελληνικό WordNet

<http://wordnet.okfn.gr/index.php?id=search&super=1&query=%CE%BF%CE%BD%CF%84%CF%8C%CF%84%CE%B7%CF%84%CE%B1>

- Αναπτύχθηκε στο Πανεπιστήμιο Πατρών με σύμπραξη του ΕΚΠΑ.
- Περιέχει 18.677 synsets και 24,811 λήμματα

Panagiotis Blitsas and Maria Gregoriadou. 2011. Greek WordNet and Its Extension with Terms of the Computer Science Domain. [Universal Access in Human-Computer Interaction. Applications and Services: 6th International Conference, UAHCI 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part IV](#) (pp.22-31)

Για το επόμενο μάθημα:

Συνδέοντας το FrameNet με το WordNet

<https://dkm-static.fbk.eu/resources/images/5/52/Sergieny-a-bryl-tonelli.pdf>

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Adjective

- [S:](#) (adj) [disgusted](#), **fed up**, [sick](#), [sick of](#), [tired of](#) (having a strong distaste from surfeit)
"grew more and more disgusted"; "fed up with their complaints"; "sick of it all"; "sick to death of flattery"; "gossip that makes one sick"; "tired of the noise and smoke"

