

# Επεξεργασία Φυσικής Γλώσσας Διανυσματική Σημασιολογία και Ενσωματώσεις

Δημήτρης Πασχαλίδης  
Τμήμα Πληροφορικής  
Πανεπιστήμιο Κύπρου



# Ποια είναι η σημασία των λέξεων;

- ❑ Από N-Gram και Text Classification:
  - Οι λέξεις είναι ακολουθίες γραμμάτων ή δείκτες σε μια λίστα λεξιλογίου.
- ❑ Πώς μπορούμε να αναπαραστήσουμε την έννοια μιας λέξης;
- ❑ Κοιτάξτε τη Λεξική Σημασιολογία, τη γλωσσική μελέτη της σημασίας των λέξεων.  
Λέξεις, λήμματα, αισθήσεις, ορισμοί



# Lemmas και Senses

## □ Lemma “Πιπέρι”

- Sense 1: Μπαχαρικό από φυτό πιπεριού
- Sense 2: Το ίδιο το φυτό πιπεριού
- Sense 3: Ένα άλλο παρόμοιο φυτό (Τζαμάικας)
- Sense 4: Ένα άλλο φυτό με κόκκους πιπεριού (Καλιφόρνιας)
- Sense 5: Η πιπεριά



Μια sense ή "έννοια" είναι το νόημα μιας λέξης.

# Lemmas και Senses

## □ Lemma “Ποντίκι”

- Sense 1: Οποιαδήποτε πολυάριθμα μικρά τρωκτικά.
- Sense 2: Χειροκίνητη συσκευή που ελέγχει έναν κέρσορα.

- Πώς πρέπει να αντιπροσωπεύουμε την έννοια μιας λέξης;  
Χρήση λέξεων, λημμάτων, αισθήσεων και ορισμών  
Χρήση σχέσεων μεταξύ λέξεων ή αισθήσεων

# Σχέση: Συνώνυμο

❑ Τα συνώνυμα έχουν την ίδια σημασία σε ορισμένα ή σε όλα τα πλαίσια.

- Couch / Sofa
- Big / Large
- Automobile / Car
- Water / H<sub>2</sub>O

Σημειώστε ότι πιθανότατα δεν υπάρχουν παραδείγματα τέλειας συνωνυμίας.

- Ακόμη και αν οι πτυχές του νοήματος είναι ταυτόσημες.
- **Ακόμα μπορεί να μην διατηρήσει την αποδοχή με βάση τις έννοιες της ευγένειας, αργκό, μητρώο, είδος, κλπ.**

■ “My **big** sister” ≠ “My **large** sister”

# Σχέση: Αντωνυμία

- ❑ Αισθήσεις που είναι αντίθετες σε σχέση με ένα χαρακτηριστικό του νοήματος (διαφορετικά, είναι πολύ παρόμοιες) .
- Dark / Light, Short / Long, Fast / Slow, Hot / Cold, Up / Down, In / Out
- ❑ Τυπικά, τα αντώνυμα μπορούν:
  - Ορισμός δυαδικής αντίθεσης: Short / Long, Fast / Slow
  - Να είναι αντίστροφα: Up / Down, Rise / Fall



# Σχέση: Ομοιότητα

- ❑ Λέξεις με παρόμοιες σημασίες.
- ❑ **ΟΧΙ συνώνυμα**, αλλά μοιράζονται κάποιο στοιχείο του νοήματος

- Car / Bicycle, Cow / Horse

- ❑ Ρωτήστε τους ανθρώπους πόσο παρόμοιες είναι δύο λέξεις →  
SimLex-999 Dataset  
Hill et al. 2015

Word 1	Word 2	Similarity
vanish	disappear	9.80
behave	obey	7.30
belief	impression	5.95
muscle	bone	3.65
modest	flexible	0.98
hole	agreement	0.30

# Σχέση: Συγγένεια λέξεων

- ❑ Γνωστό και ως “**Word Association**”
- ❑ Οι λέξεις μπορούν να συσχετιστούν με οποιονδήποτε τρόπο, ίσως μέσω ενός σημασιολογικού πεδίου.
- *coffee and tea*:           **SIMILAR**
- *coffee and cup*:       **RELATED**, όχι **SIMILAR**
- *car, bicycle*:           **SIMILAR**
- *car, gas*:                   **RELATED**, όχι **SIMILAR**





# Σημασιολογικό πεδίο

Ένα σημασιολογικό πεδίο είναι ένα σύνολο λέξεων που καλύπτουν ένα συγκεκριμένο σημασιολογικό πεδίο και έχουν δομημένες σχέσεις μεταξύ τους.

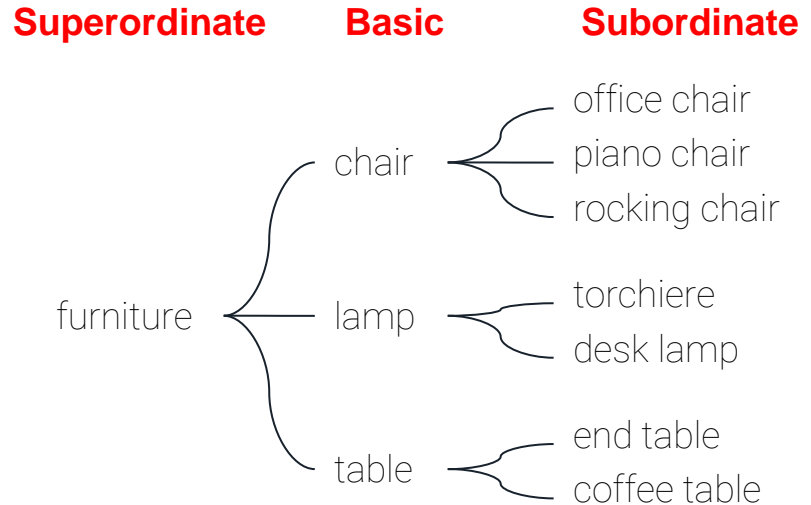
- Νοσοκομεια:
  - Surgeon, scalpel, nurse, anesthetic, hospital
- Εστιατορια:
  - Waiter, menu, plate, food, chef
- Σπίτια:
  - Door, roof, kitchen, family, bed



# Relation: Superordinate / Subordinate

☐ Μια αίσθηση είναι **subordinate** μιας άλλης αν η πρώτη αίσθηση είναι πιο συγκεκριμένη, δηλώνοντας μια υποκατηγορία της άλλης

- *Car* is a subordinate of *vehicle*.
- *Mango* is a subordinate of *fruit*.
- ☐ Conversely superordinate:
  - *Vehicle* is a superordinate of *car*.
  - *Fruit* is a superordinate of *Mango*.



# Connotation / Sentiment

Οι λέξεις έχουν επίσης συναισθηματικές έννοιες

- Θετικοί συνειρμοί Happy
- Αρνητικοί συνειρμοί Sad

Connotations can be subtle

- Θετικοί συνειρμοί Happy
- Αρνητικοί συνειρμοί Sad

Evaluation - Sentiment

- Θετικοί συνειρμοί Great, Love
- Αρνητικοί συνειρμοί Terrible, Hate



# Connotation / Sentiment

☐ Οι λέξεις ποικίλλουν σε 3 συναισθηματικές διαστάσεις:

- Valence: η ευχαρίστηση του ερεθίσματος
- Arousal: η ένταση του συναισθήματος που προκαλείται από το ερέθισμα
- Dominance: ο βαθμός ελέγχου που ασκείται από το ερέθισμα

	Word	Score	Word	Score
<b>Valence</b>	love	1.000	toxic	0.008
	happy	1.000	nightmare	0.005
<b>Arousal</b>	elated	0.960	mellow	0.069
	frenzy	0.965	napping	0.046
<b>Dominance</b>	powerful	0.991	weak	0.045
	leadership	0.983	empty	0.081

Values from NRC VAD Lexicon (Mohammad 2018)

# Σημασιολογικό Πλαίσιο

- Ένα σύνολο λέξεων που υποδηλώνουν προοπτικές ή συμμετέχοντες σε έναν συγκεκριμένο τύπο εκδήλωσης:
  - “buy”: το γεγονός από την οπτική γωνία του αγοραστή
  - “sell” : από την οπτική γωνία του πωλητή
  - “pay”: εστίαση στη νομισματική πτυχή
  - *John hit Bill*
  - *Bill was hit by John*
  
- Τα πλαίσια έχουν σημασιολογικούς ρόλους (όπως buyer, seller, goods, money) και οι λέξεις σε μια πρόταση δεν μπορούν να πάρουν αυτούς τους ρόλους.

# Λεξική Σημασιολογία

- Πώς πρέπει να αντιπροσωπεύουμε την έννοια μιας λέξης;
  1. Λέξεις, λήμματα, αισθήσεις και ορισμοί
  - Σχέσεις μεταξύ λέξεων ή αισθήσεων
  - Ταξινόμηση των σχέσεων
  - Ομοιότητα και συγγένεια λέξεων
  - Συνειρμοί και συναίσθημα
  - Σημασιολογικά πλαίσια και ρόλοι



# Πηγές και λεξικά

## □ WordNet

```
from nltk.corpus import wordnet as wn
panda = wn.synset('panda.n.01')
hyper = lambda s: s.hypernyms()

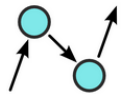
list(panda.closure(hyper))
```

```
[Synset('procyonid.n.01'),
Synset('carnivore.n.01'),
Synset('placental.n.01'),
Synset('mammal.n.01'),
Synset('vertebrate.n.01'),
Synset('chordate.n.01'),
Synset('animal.n.01'),
Synset('organism.n.01'),
Synset('living_thing.n.01'),
Synset('whole.n.02'),
Synset('object.n.01'),
Synset('physical_entity.n.01'),
Synset('entity.n.01')]
```

# Πηγές και λεξικά

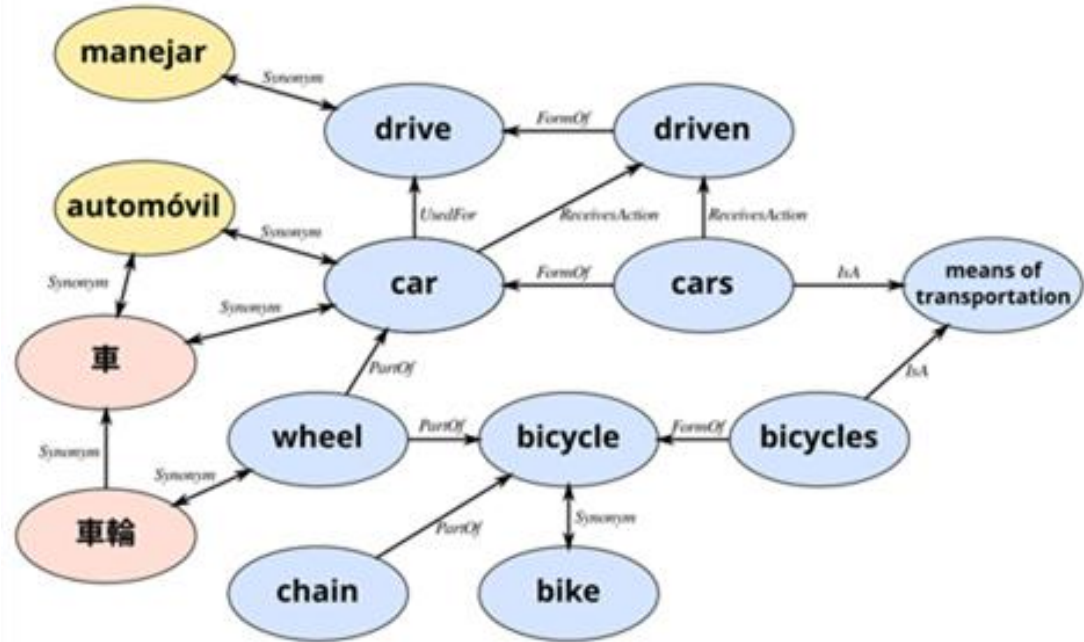
## □ ConceptNet

- ConceptNet is a freely-available semantic network, designed to help computers understand the meanings of words that people use.



**ConceptNet**

An open, multilingual knowledge graph





# Προβλήματα με τη διακριτή αναπαράσταση

- Προσεγγίσεις όπως BoW και 1-Hot-Encoding είναι πολύ χονδροειδείς, αραιές και ακριβές.
- Expert → Skillful
- Σχέσεις λέξεων είναι δύσκολο να υπολογιστούν!

<b>expert</b>	0	<b>1</b>	0	0	0	0	0	0
<b>skillful</b>	0	0	0	0	0	<b>1</b>	0	0

Κοίταξε στο Vector Semantics



# Περίληψη

- Τύποι σημασιολογικών σχέσεων λέξεων:
  - Συνωνυμία
  - Αντωνυμία
  - Ομοιότητα
  - Συγγένειας
  - Υπερκείμενος / Υποδεέστερος
  - Αίσθημα
- Πόροι και Σημασιολογικά Λεξικά π.χ. **ConceptNet**
- Ζητήματα με διακριτές αναπαραστάσεις



# Πόροι

- Jurafsky, D. and H. Martin Justin, Chapter 6. "Vector Semantics and Embeddings" Speech and Language Processing
- ConceptNet: <https://conceptnet.io/>
- WordNet: <https://wordnet.princeton.edu/>

