

# Research topics



German Rigau i Claramunt

<http://adimen.si.ehu.es/~rigau>

IXA group

Departamento de Lenguajes y Sistemas Informáticos

UPV/EHU

# Research Topics

- RT1: (Yarowsky 92) using WordNet and Wikipedia
- RT2: ukb vs. w2v on similarity/relatedness
- RT3: Mapping WordNet to Wikipedia
- RT4: Mapping WordNet to EuroVoc / IATE
- RT5: New WordNet domains using BabelDomains
- RT6: New KnowNets
- RT7: Geo-country Tagging
- RT8: Domain tagging
- RT9: Domain tagging of WordNet
- ...

# RT1 ukb vs. w2v on similarity/relatedness

- UKB: <http://ixa2.si.ehu.es/ukb>
- word2vec:
  - <https://code.google.com/archive/p/word2vec>
  - <http://deeplearning4j.org/word2vec>
  - <https://radimrehurek.com/gensim/models/word2vec.html>
- Similarity/relatedness datasets:
  - <http://www.cs.cmu.edu/~mfaruqui/suite.html>

# RT2 Tools: UKB and word2vec

- UKB: <http://ixa2.si.ehu.es/ukb>
- word2vec:
  - <https://code.google.com/archive/p/word2vec>
  - <http://deeplearning4j.org/word2vec>
  - <https://radimrehurek.com/gensim/models/word2vec.html>
- Similarity/relatedness datasets:
  - <http://www.cs.cmu.edu/~mfaruqui/suite.html>

# RT4: Mapping WordNet to EuroVoc / IATE

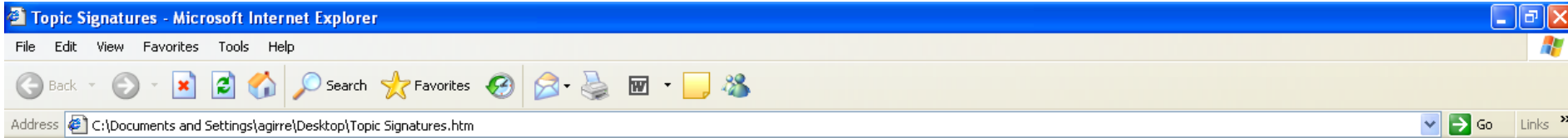
<http://eurovoc.europa.eu/5650> “greenhouse gas”

- How to perform the task?
- How to evaluate this task?

# RT6 New KnowNets

- Current KnowNets:
- <http://adimen.si.ehu.es/web/KnowNet>
  - Topic Signatures (word vectors)
  - Knowledge-based WSD (SSI-Dijkstra+)
  - => KnowNets
  - Evaluation: lexical sample WSD

# RT6 New KnowNets



## Topic Signatures Browser (all WN 1.6 polysemous nouns)

Type any noun:

### horse (definitions in WordNet 1.6)

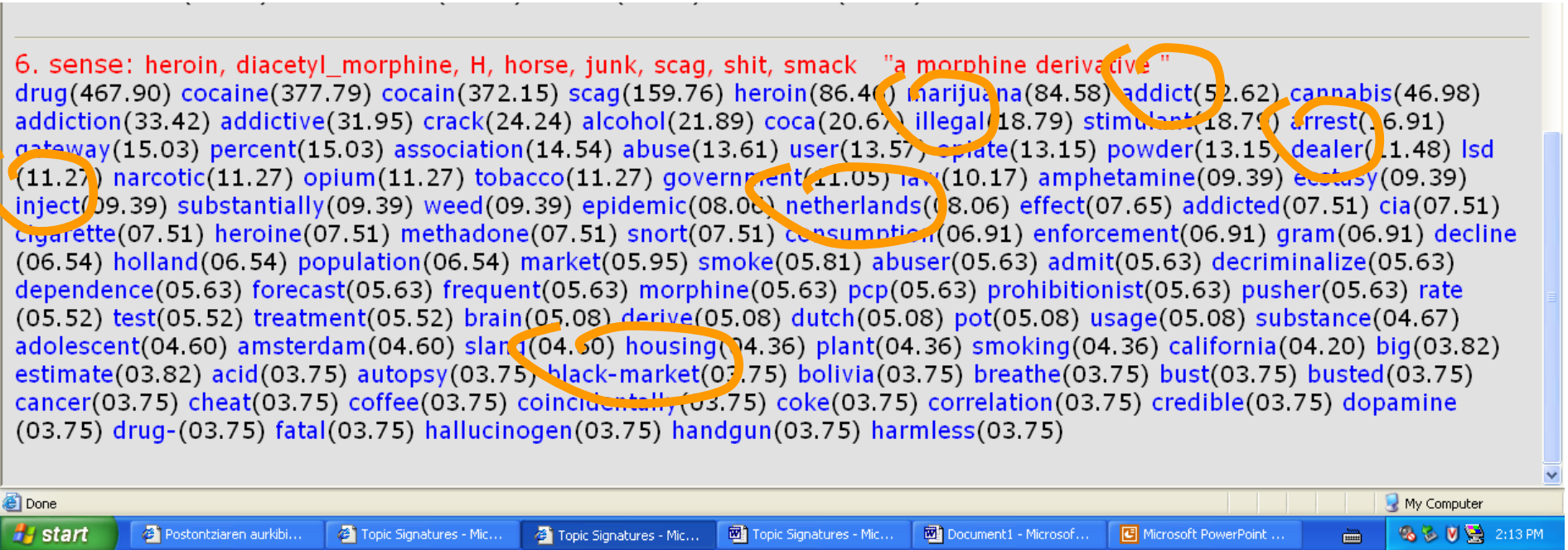
- 1. sense:** horse, Equus\_caballus "solid-hoofed herbivorous quadruped domesticated since prehistoric times "
- 2. sense:** horse "a padded gymnastic apparatus on legs "
- 3. sense:** cavalry, horse\_cavalry, horse "troops trained to fight on horseback: "
- 4. sense:** sawhorse, horse, sawbuck, buck "a framework for holding wood that is being sawed "
- 5. sense:** knight, horse "a chessman in the shape of a horse's head; can move two squares horizontally and one vertically (or vice versa) "
- 6. sense:** heroin, diacetyl\_morphine, H, horse, junk, scag, shit, smack "a morphine derivative "

**1. sense:** horse, Equus\_caballus "solid-hoofed herbivorous quadruped domesticated since prehistoric times "

polo(112.40) equus(102.66) zebra(101.61) eohippus(86.65) quagga(83.87) horse(79.18) pony(78.52) hinny(67.16) stablemate(54.63) racehorse(53.24) donkey(47.32) liver(34.45) mare(34.35) mussel(31.66) race(28.98) pinto(26.67) bangtail(26.10) workhorse(25.75) palomino(24.75) saddle(24.36) stallion(24.36) dawn(23.68) mesohippus(22.27) equid(19.18) riding(19.20) companion(18.57) harness(18.30) specie(17.71) extinct(15.66) offspring(15.66) chestnut(15.61) female(15.47) hyracotherium(15.31) foal(14.61) ass(13.92) ancestor(13.72) hybrid(13.22) stable(12.67) filly(11.30) trainer(10.66) fossil(10.09) mule(10.08) thoroughbred(09.74) dreissena(08.70) breed(08.50) burn(08.35) ride(07.50) breeding(06.96) age(06.77) wild(06.62) racing(06.61) modern(06.22) champion(06.18) ago(06.05) male(05.70) broodmare(05.56) finch(05.56) mammal(05.56) dog(05.38) printer(05.38) colt(05.33) equine(05.12) owner(05.04) derby(04.87) midget(04.87) oligocene(04.87) sterile(04.87) arabian(04.69) ownership(04.69) genus(04.48) rescue(04.48) domestic(04.44) trail(04.30) eocene(04.17) mustang(04.17) subspecies(04.17) animal(03.85) bean(03.84) stud(03.84) gelding(03.82) sheep(03.82) evolution(03.63) tail(03.50) breeder(03.48) protohippus(03.48) dressage(03.41) prehistoric(03.41) rider(03.36) toe(03.23) creature(03.20) equidae(03.13) feral(03.13) sorrel(03.13) sire(03.09) mane(02.98) native(02.98) retire(02.98) evolve(02.96) tooth(02.96) cave(02.78)

# RT6 New KnowNets

- Topic Signatures:
  - [http://ixa3.si.ehu.es/cgi-bin/signatureak/signature\\_lem.cgi](http://ixa3.si.ehu.es/cgi-bin/signatureak/signature_lem.cgi)
  - Word vectors (acquired from the web) associated to synsets





# RT6 New KnowNets

- Similar process in:
  - <http://adimen.si.ehu.es/cgi-bin/WSDbyEvocation.v1/index.php>
  - <http://adimen.si.ehu.es/cgi-bin/WSDbyEvocation.v3/index.php>
- Using LDA on POS tagged BNC to obtain the word vectors
- Using SSSI-Dijkstra+ for WSD
- Try:
  - bank.n + river.n
  - bank.n + money.n

# RT6 New KnowNets

- New KnowNets:
  - word2vec (word embeddings)
  - Knowledge-based WSD (SSI-Dijkstra+, UKB)
  - => New KnowNets
  - Evaluation: lexical sample WSD

# RT6 New KnowNets

- New KnowNets:

- word2vec (word embeddings)

- <https://code.google.com/p/word2vec>

- <http://deeplearning4j.org/word2vec.html>

- <https://radimrehurek.com/gensim/models/word2vec.html>

- ...

- Knowledge-based WSD (SSI-Dijkstra+, UKB)

- <http://ixa2.si.ehu.es/ukb>

- => New KnowNets (a very dense graph!)

- Evaluation: lexical sample WSD

- <http://adimen.si.ehu.es/web/WSD-WN-Glosses>

# RT7 Geo-country Tagging

Mariano Rajoy announced yesterday in Madrid that the budget cuts will continue next year.

Madrid => Spain?

Mariano\_Rajoy => Spain?

- Where is that knowledge?
- How to use that knowledge to perform the task?
- How to evaluate this task?

# RT8: Domain Tagging

Mariano Rajoy announced yesterday in Madrid that the budget cuts will continue next year.

budget\_cuts => ECONOMY? POLITICS?

Mariano\_Rajoy => ECONOMY? POLITICS?

Madrid => ECONOMY? POLITICS? ...

- Where is that knowledge?
- How to use that knowledge to perform the task?
- How to evaluate this task?

# RT8: Domain Tagging of WordNet

hospital#n#1 : health facility where patients receive treatment

hospital => MEDICINE?

health\_facility => MEDICINE?

patients => MEDICINE?

treatment => MEDICINE?

- Where is that knowledge?
- How to use that knowledge to perform the task?
- How to evaluate this task?

# Research topics



German Rigau i Claramunt

<http://adimen.si.ehu.es/~rigau>

IXA group

Departamento de Lenguajes y Sistemas Informáticos

UPV/EHU