# Distributional Semantics

Marcos Garcia
marcos.garcia.gonzalez@udc.gal
grupolys.org/~marcos

LyS Group, UdC




# Contents

1 Intro

2 Formal and distrib. semantics

3 Distributional meaning

4 Computing similarities

5 Contexts

6 Example uses

7 Drawbacks

# Contents

1 Intro

2 Formal and distrib. semantics

3 Distributional meaning

4 Computing similarities

5 Contexts

6 Example uses

7 Drawbacks

# Introduction

- Objectives:
  - Formal and distributional semantics.
  - Main ideas behind distributional meaning.
  - Distributional contexts.
  - Similarity and analogy.

# Contents

1 Intro

2 Formal and distrib. semantics

3 Distributional meaning

4 Computing similarities

5 Contexts

6 Example uses

7 Drawbacks

# Formal *versus* distributional semantics

- Formal semantics:
  - Precise mathematical models for understanding linguistic meaning.
  - Assign meaning to the words, and defines how to combine them to form the meaning of phrases, etc.
  - *Meaning* as a *reference*.

# Formal *versus* distributional semantics

- Formal semantics:
  - Precise mathematical models for understanding linguistic meaning.
  - Assign meaning to the words, and defines how to combine them to form the meaning of phrases, etc.
  - *Meaning* as a *reference*.
- Distributional semantics:
  - The context of a given word provides information about its meaning.
  - Quantifies semantic similarities between words (and larger expressions).
  - *Meaning* as *sense*.

# Formal *versus* distributional semantics in NLP

- Formal semantics methods (often) needs a lot of manual work:
  - Define and classify semantic features.
  - Define compositional rules.
  - ...

# Formal *versus* distributional semantics in NLP

- Formal semantics methods (often) needs a lot of manual work:
  - Define and classify semantic features.
  - Define compositional rules.
  - ...
- Distributional methods are (often) unsupervised:
  - *Do not need* annotated data (but it helps).
  - Create vector spaces of each word automatically.
  - ...

# Contents

1 Intro

2 Formal and distrib. semantics

3 Distributional meaning

4 Computing similarities

5 Contexts

6 Example uses

7 Drawbacks

# What does **sumart** means?

“They were playing **sumart** the whole day.”

# What does **sumart** means?

“They were playing **sumart** the whole day.”

“Tomorrow I’m going to a match of **sumart** with my parents.”

# What does **sumart** means?

“They were playing **sumart** the whole day.”

“Tomorrow I’m going to a match of **sumart** with my parents.”

“The local **sumart** team won the championship.”

# What does **sumart** means?

“They were playing **sumart** the whole day.”

“Tomorrow I’m going to a match of **sumart** with my parents.”

“The local **sumart** team won the championship.”

“It was the best **sumart** game I ever played.”

# What does **sumart** means?

“They were playing **sumart** the whole day.”

“Tomorrow I’m going to a match of **sumart** with my parents.”

“The local **sumart** team won the championship.”

“It was the best **sumart** game I ever played.”

“John was our **sumart** coach last year.”

# What does **sumart** means?

“They were playing **sumart** the whole day.”

“Tomorrow I’m going to a match of **sumart** with my parents.”

“The local **sumart** team won the championship.”

“It was the best **sumart** game I ever played.”

“John was our **sumart** coach last year.”

“I need a ball of **sumart** for the weekend.”

# What does **sumart** means?

“They were playing **sumart** the whole day.”

“Tomorrow I’m going to a match of **sumart** with my parents.”

“The local **sumart** team won the championship.”

“It was the best **sumart** game I ever played.”

“John was our **sumart** coach last year.”

“I need a ball of **sumart** for the weekend.”

“The **sumart** game ended in a tie.”

. . .

# What does **sumart** means?

“They were playing **sumart** the whole day.”

“Tomorrow I’m going to a match of **sumart** with my parents.”

“The local **sumart** team won the championship.”

“It was the best **sumart** game I ever played.”

“John was our **sumart** coach last year.”

“I need a ball of **sumart** for the weekend.”

“The **sumart** game ended in a tie.”

. . .

We can infer (learn?) the meaning by its distribution.

# Foundation ideas

*“the complete meaning of a word is always contextual, and no study of meaning apart from context can be taken seriously.”*

*Firth, John R., 1935. “The technique of semantics”.*

# Foundation ideas

*“the meaning of a word is its use in the language”*

*Wittgenstein, Ludwig, 1997*

# Foundation ideas

*“It is possible to state the occurrence of any element relative to any other element, to the degree of exactness indicated above, so that distributional statements can cover all of the material of a language without requiring support from other types of information.”*

*“The fact that, for example, not every adjectives occurs with every noun can be used as a measure of meaning difference.”*

*Harris, Zellig, 1954. “Distributional Structure”.*

# Foundation ideas

*“You shall know a word by the company it keeps”*

*Firth, John R., 1957. “A synopsis of linguistic theory 1930–1955”.*

# The company it keeps. . .

- I want to drink water.
- She loves fresh water!
- . . .

- The dog is running again.
- My dog barks every morning.
- . . .

# The company it keeps. . .

- I want to drink water.
- She loves fresh water!
- . . .

- The dog is running again.
- My dog barks every morning.
- . . .

| | fresh | barks | . . . |
|---|---|---|---|
| *water* | | | |
| *dog* | | | |

# The company it keeps. . .

- I want to drink water.
- She loves fresh water!
- . . .

- The dog is running again.
- My dog barks every morning.
- . . .

| | fresh | barks | . . . |
|---|---|---|---|
| *water* | 6 | 1 | . . . |
| *dog* | | | |

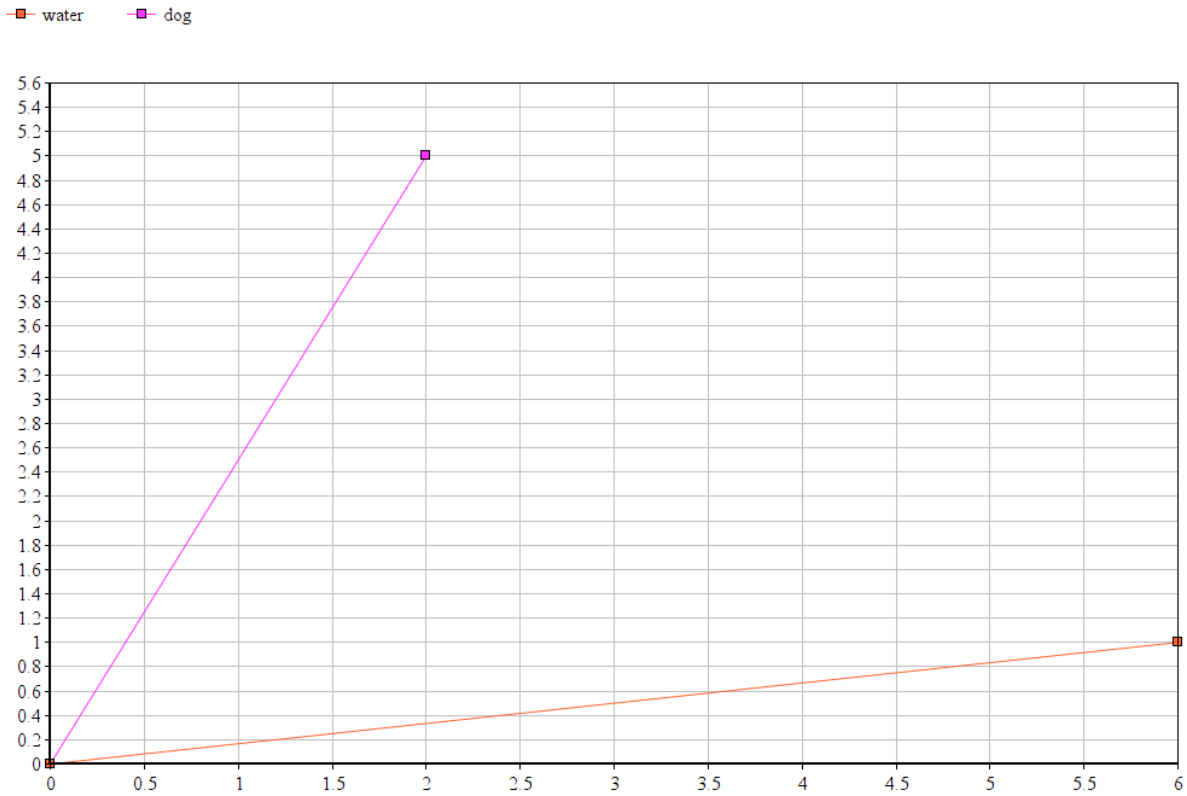
# The company it keeps. . .

- I want to drink water.
- She loves fresh water!
- . . .

- The dog is running again.
- My dog barks every morning.
- . . .

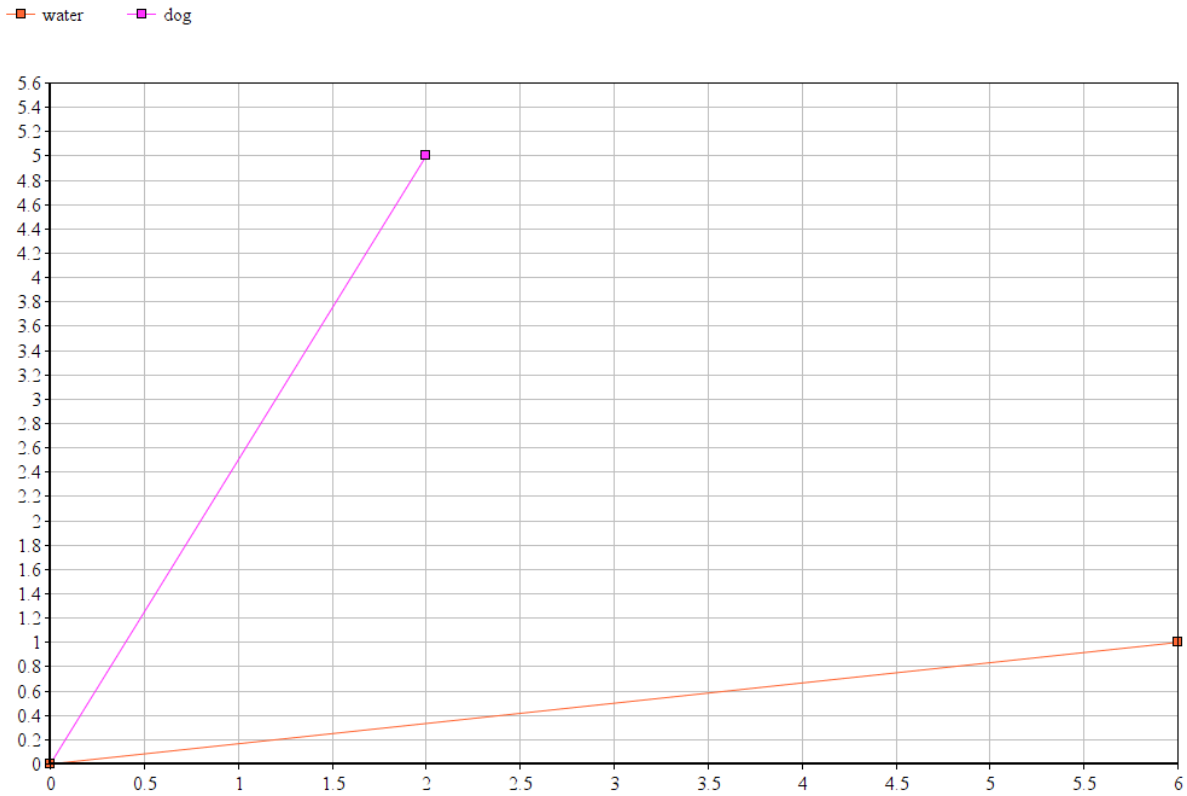
| | fresh | barks | . . . |
|---|---|---|---|
| *water* | 6 | 1 | . . . |
| *dog* | 2 | 5 | . . . |

# Word vectors



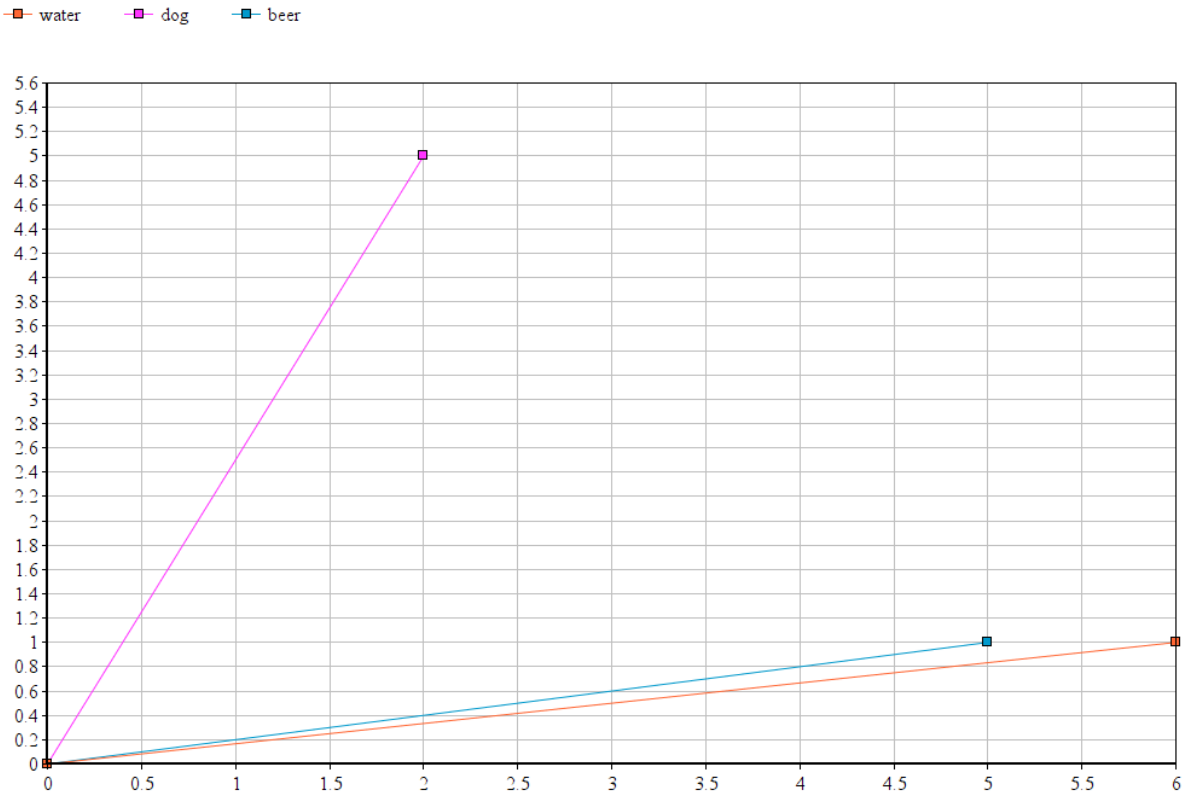

| | fresh | barks | ... |
|---|---|---|---|
| *water* | 6 | 1 | ... |
| *dog* | 2 | 5 | ... |

# Word vectors



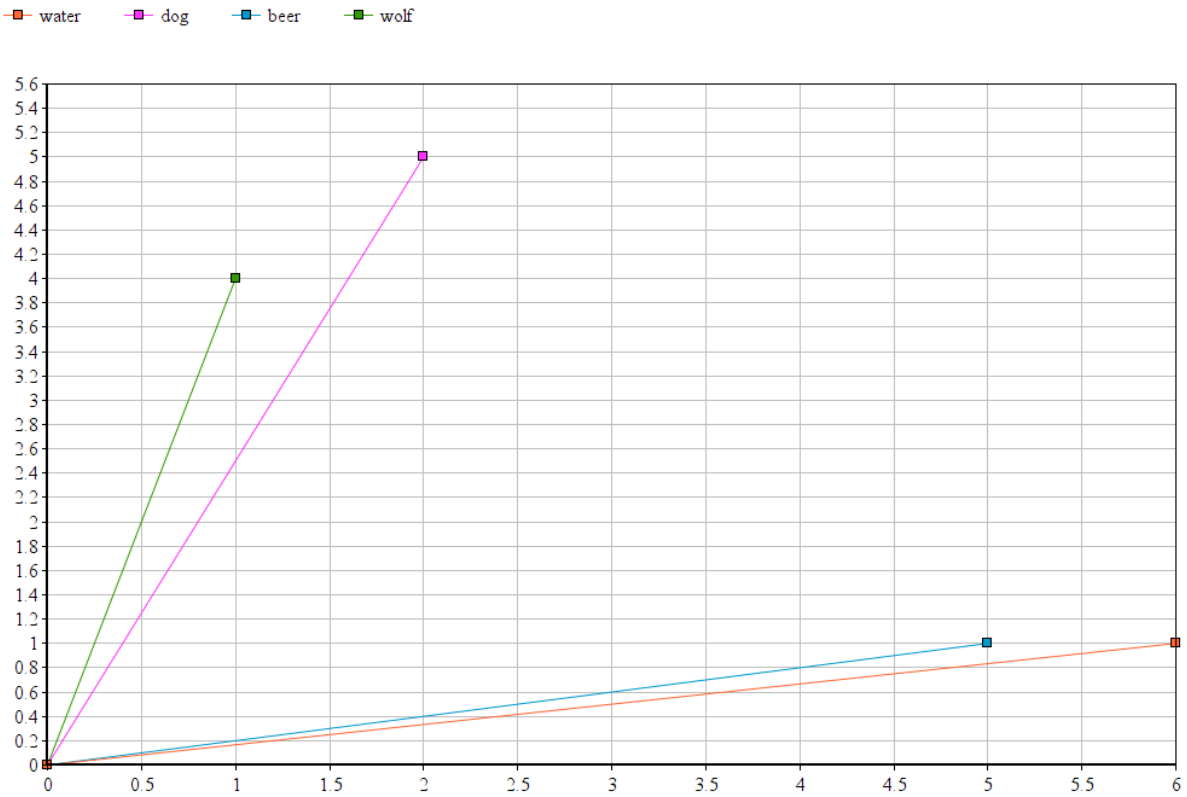

| | fresh | barks | ... |
|---|---|---|---|
| *water* | 6 | 1 | ... |
| *dog* | 2 | 5 | ... |
| *beer* | 5 | 1 | ... |

# Word vectors




| | fresh | barks | ... |
|---|---|---|---|
| *water* | 6 | 1 | ... |
| *dog* | 2 | 5 | ... |
| *beer* | 5 | 1 | ... |

# Word vectors



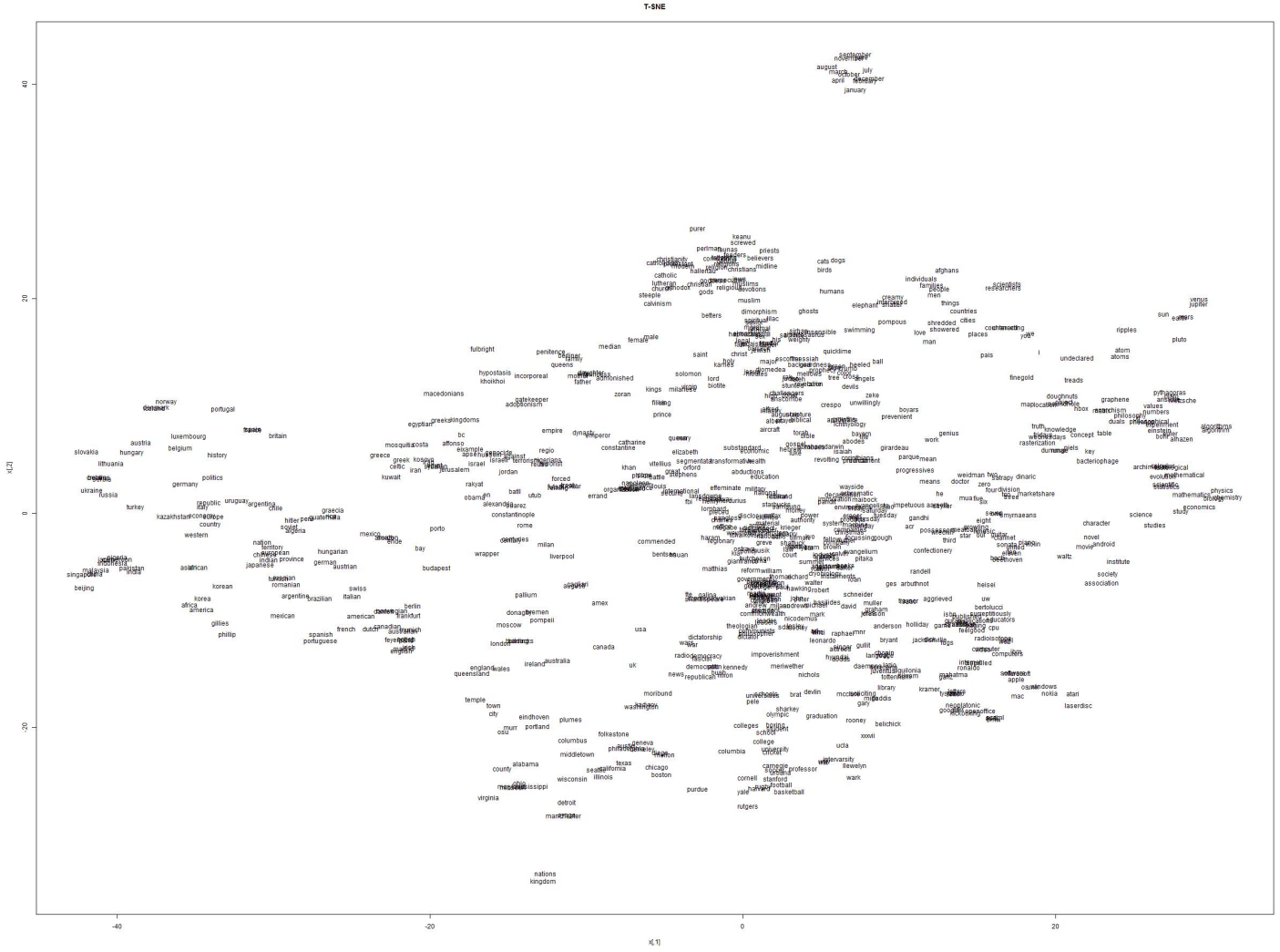

| | fresh | barks | ... |
|---|---|---|---|
| *water* | 6 | 1 | ... |
| *dog* | 2 | 5 | ... |
| *beer* | 5 | 1 | ... |
| *wolf* | 1 | 4 | ... |

# Dimensionality and weighting

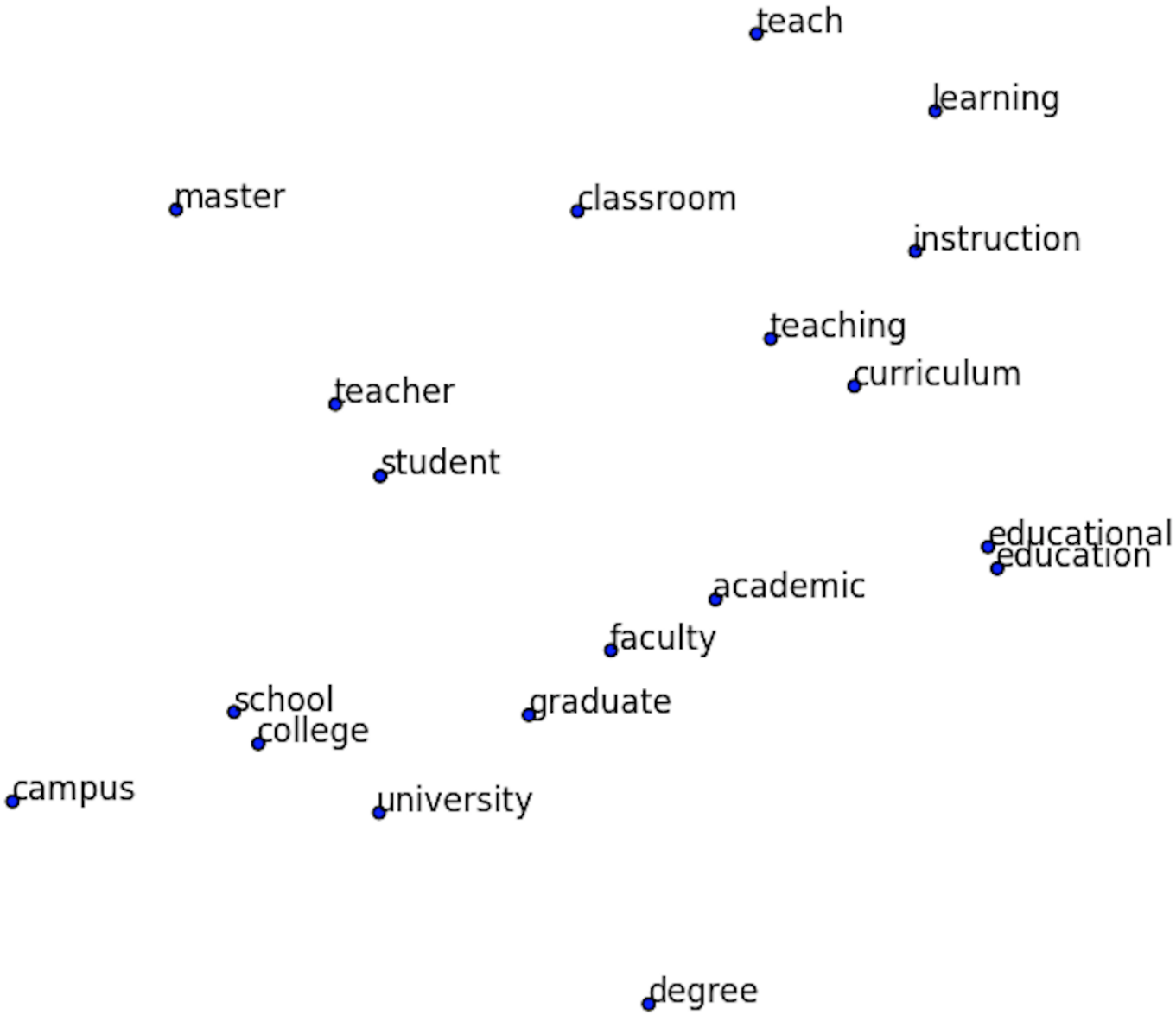
- Each context in the corpus == vector dimension.
- Huge vectors $\longrightarrow$ dimensionality reduction (LSA, SVD, etc.).
- Frequency is not a good association measure $\longrightarrow$ weighting (PMI, tf-idf, etc.).

- Low-dimensional vectors.

# Vector spaces



# Vector spaces



# Contents

1 Intro

2 Formal and distrib. semantics

3 Distributional meaning

4 Computing similarities

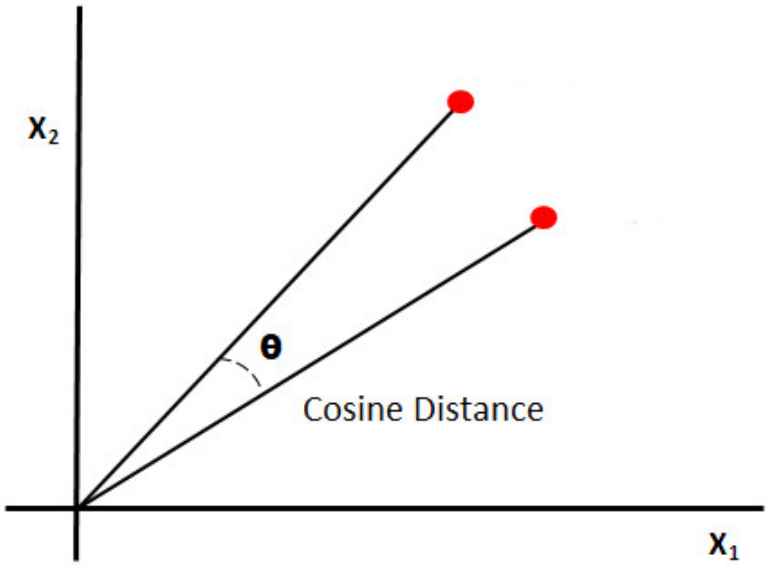
5 Contexts

6 Example uses

7 Drawbacks

# Cosine distance

- Words are represented by vectors.
- How to measure the similarity between two words?
- Cosine similarity (others: Jaccard, Dice, etc.).



# Contents

1 Intro

2 Formal and distrib. semantics

3 Distributional meaning

4 Computing similarities

5 Contexts

6 Example uses

7 Drawbacks

# Contexts for distributional semantics

- *Bag-of-words* models.

| | fresh | barks | ... |
|---|---|---|---|
| *water* | 6 | 1 | ... |
| *dog* | 2 | 5 | ... |

# Contexts for distributional semantics

- *Bag-of-words* models.
- PoS-tags.

| | fresh_ADJ | barks_VERB | ... |
|---|---|---|---|
| *water* | 6 | 1 | ... |
| *dog* | 2 | 5 | ... |

# Contexts for distributional semantics

- *Bag-of-words* models.
- PoS-tags.
- Lemmatized contexts (with/without PoS).

| | fresh (ADJ) | bark (VERB) | ... |
|---|---|---|---|
| *water* | 6 | 1 | ... |
| *dog* | 2 | 5 | ... |

# Contexts for distributional semantics

- *Bag-of-words* models.
- PoS-tags.
- Lemmatized contexts (with/without PoS).
- Dependency (syntactic) contexts.

| | *amod*_fresh | *subj*_bark | ... |
|---|---|---|---|
| *water* | 6 | 1 | ... |
| *dog* | 2 | 5 | ... |

# Contexts for distributional semantics

- *Bag-of-words* models.
- PoS-tags.
- Lemmatized contexts (with/without PoS).
- Dependency (syntactic) contexts.
- Context selection / filter.

| | *amod*_fresh | *subj*_bark | ... |
|---|---|---|---|
| *water* | 6 | 1 | ... |
| *dog* | 2 | 5 | ... |

# Contents

1 Intro

2 Formal and distrib. semantics

3 Distributional meaning

4 Computing similarities

5 Contexts

6 Example uses

7 Drawbacks

# Algebraic operations

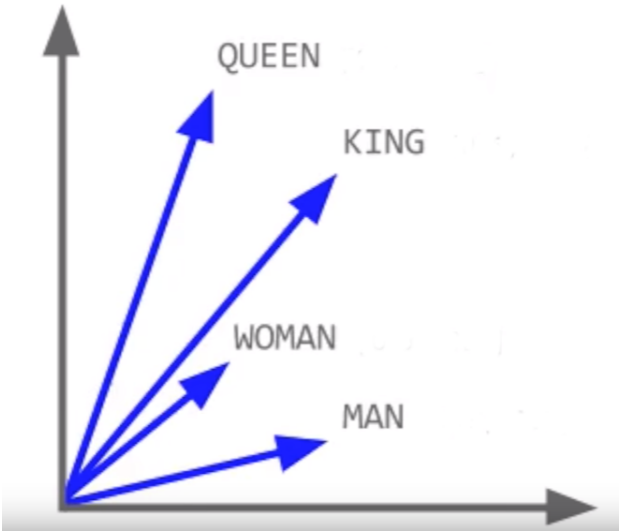
- Algebraic operations on vectors.
- Prime example:
  - king - man + woman =

# Algebraic operations

- Algebraic operations on vectors.
- Prime example:
  - king - man + woman = queen

# Algebraic operations

- Algebraic operations on vectors.
- Prime example:
  - king - man + woman = queen




Example: *SketchEngine embeddings*

# Analogy

- Capable of finding analogies such as:
  - conscious, unconscious; responsible: irresponsible
  - Madrid, Spain; Paris: France
  - dog, dogs; corpus: corpora
  - play, playing; sleep: sleeping
  - man, woman; boy: girl
  - Russia, rouble; Europe: euro
  - Italia, italian; France: french
  - ...

# Clustering

- mammals: dog, elephant, cat, cow, lion, . . .
- birds: robin, sparrow, pigeon, chicken, eagle, . . .
- fish: cod, goldfish, minnow, salmon, trout, tuna.
- vegetables: broccoli, spinach, lettuce, potato, onion, . . .
- fruit: apple, orange, grape, peach, strawberry, . . .
- trees: oak, pine, birch, cedar.
- vehicles: boat, car, ship, truck, motorcycle, helicopter, . . .
- clothes: shirt, pants, socks, jacket, sweater, . . .
- tools: screwdriver, hammer, chisel, wrench, sandpaper, pliers.
- kitchenware: cup, bowl, spoon, pan, pot, blender, . . .

# Software

- Software:
  - *word2vec*: `https://code.google.com/archive/p/word2vec/`
  - GloVe: `https://nlp.stanford.edu/projects/glove/`
  - fastText: `https://fasttext.cc/`
- Demos:
  - LX-SemanticSimilarity:

    `http://lxcenter.di.fc.ul.pt/services/en/LXServicesSemanticSimilarity.html`
  - TensorBoard: `http://projector.tensorflow.org/`
  - Diachronic explorer:

    `https://tec.citius.usc.es/explorador-diacronico/`
  - Turku NLP: `http://bionlp-www.utu.fi/wv_demo/`

# Contents

1 Intro

2 Formal and distrib. semantics

3 Distributional meaning

4 Computing similarities

5 Contexts

6 Example uses

7 Drawbacks

# Drawbacks and on-going work

- Polysemy.
- Antonymy.
- Compositionality.
- Idioms.
- ...

# References


- References:
  - Firth, John R., 1935. “The technique of semantics”. *Transactions of the Philological Society* 34(1). 36–73.
  - Firth, John R., 1954. “A synopsis of linguistic theory 1930–1955”. In *Studies in linguistic analysis*, 1–32. Oxford: Blackwell.
  - Harris, Zellig, 1954. “Distributional Structure”. *Word* 10(23). 146–162.
  - Wittgenstein, Ludwig. 1997. Philosophical investigations. Oxford, England: Blackwell Publishers. Original work published 1953.
- Further reading:
  - Erk, Katrin, 2012. “Vector Space Models of Word Meaning and Phrase Meaning: A Survey”. *Language and Linguistics Compass* 6/10: 635–653.
  - Jurafsky, Dan & James H. Martin, 2017. “Vector Semantics”. In *Speech and Language Processing (3rd ed. draft)*. Chapter 15.

# Thanks! :-)

Questions?