

# Word analogies without magic: king – man + woman != queen

# A magical property of word2vec

And the only reason for being a bee that I know of is making honey.

contexts

↑  
focus  
word

contexts

Learn word vectors by predicting their contexts (or vice-versa).

Obtain vectors that solve word analogies:

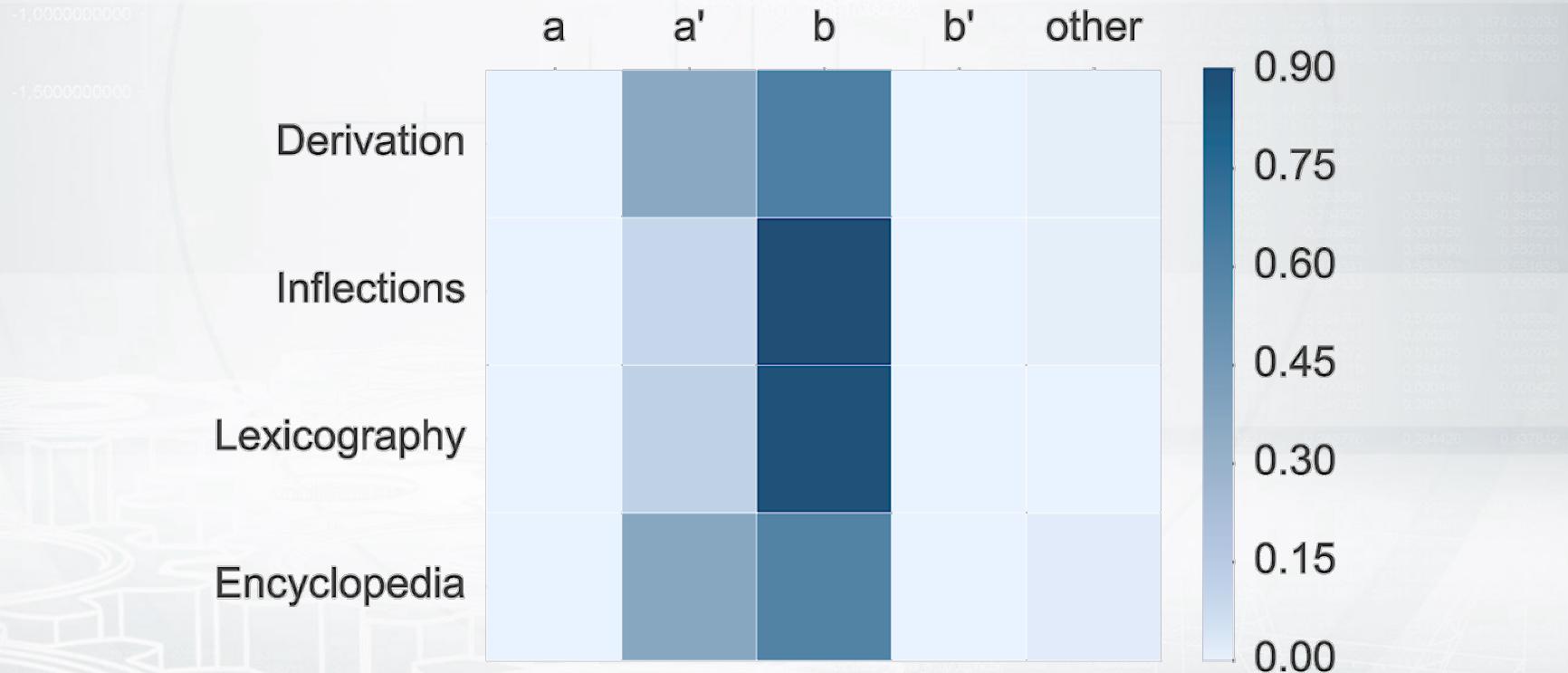
- king – man + woman = queen
- Moscow – Russia + France = Paris

Demo: [rare-technologies.com/word2vec-tutorial/](http://rare-technologies.com/word2vec-tutorial/)

# Closer look into analogy task

$$\cos(b - a + a', x) \rightarrow \max_{x \notin \{a, a', b\}}$$

king – man + woman = king:

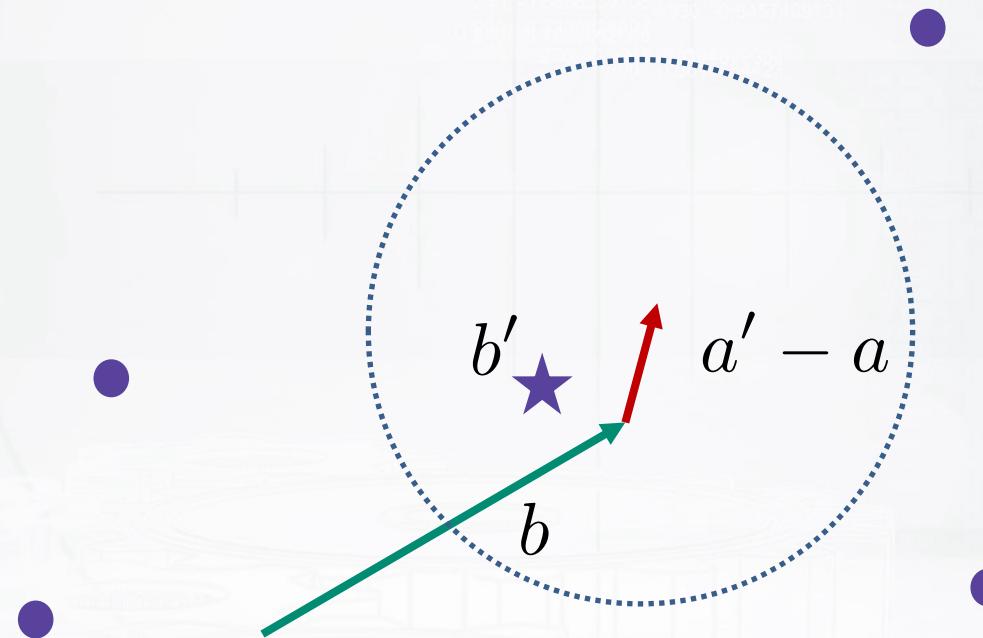


Rogers et. al. The (Too Many) Problems of Analogical Reasoning with Word Vectors, 2017.

# The closest neighbor of b is good enough?

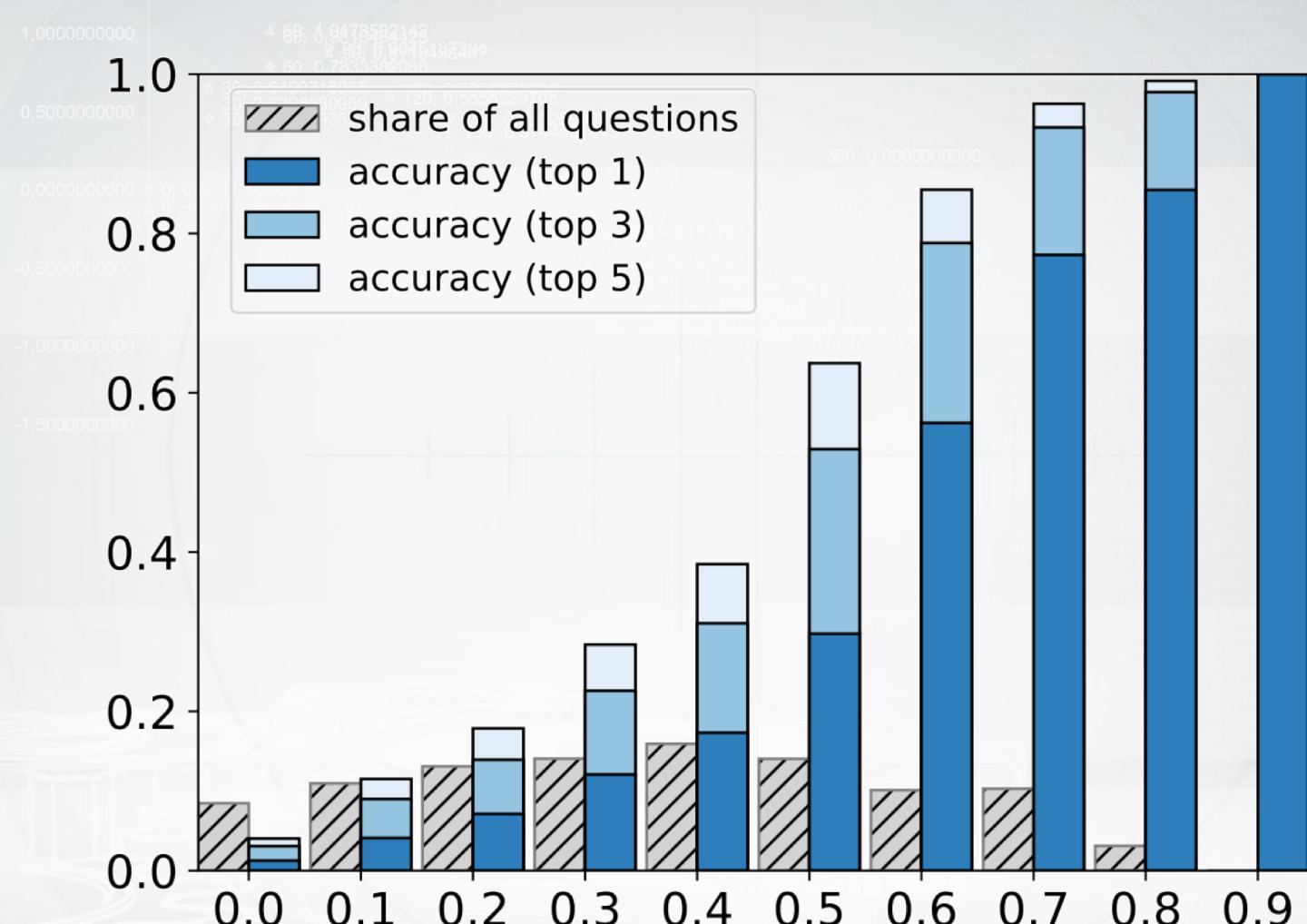
For the plural noun category in the Google test set:

70% accuracy by taking the closest neighbor of the vector b!



Linzen. Issues in evaluating semantic spaces using word analogies, 2016.

# Good accuracy when $b$ and $b'$ are close



Buckets are based on similarity between  $b$  and  $b'$

# BATS dataset

## Inflectional morphology:

- student:students, strong:stronger, follow:following, ...*

## Derivational morphology:

- bake:baker, edit:editable, home:homeless, ...*

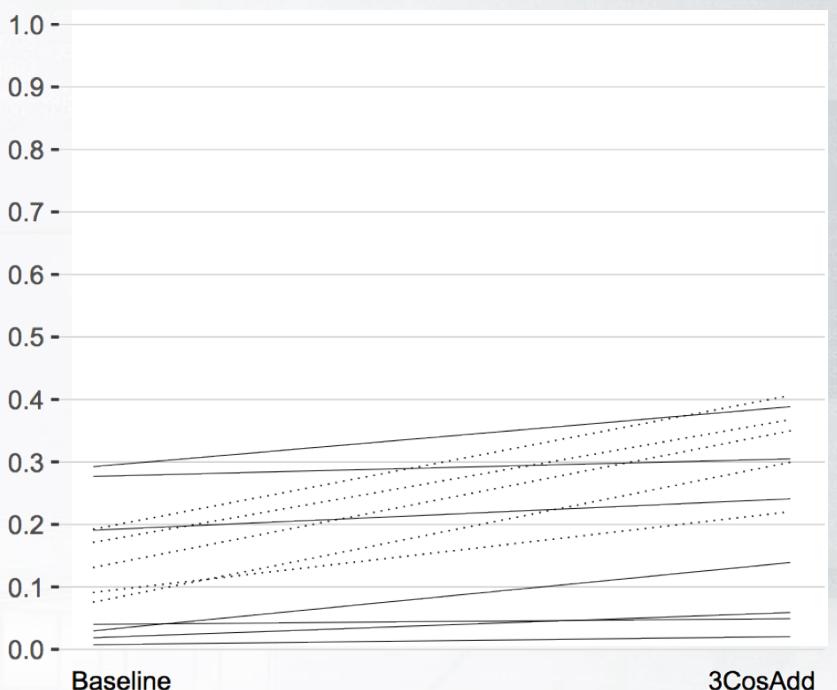
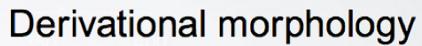
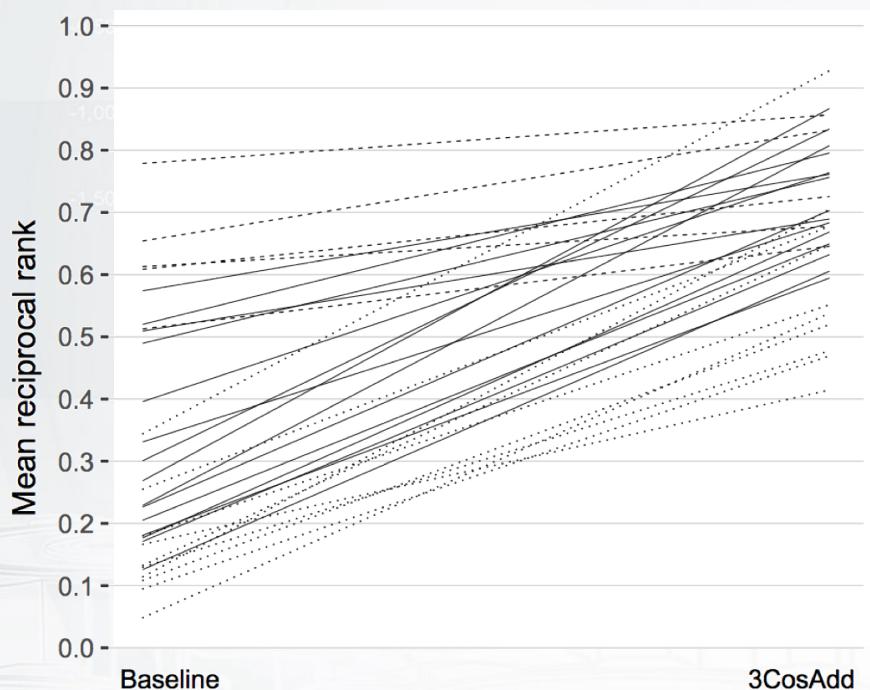
## Lexicographic semantics:

- Hypernyms, meronyms : peach:fruit, sea:water, player:team,  
...
- Antonyms, synonyms: up:down, clean:dirty, cry:scream, ...

## Encyclopedic semantics:

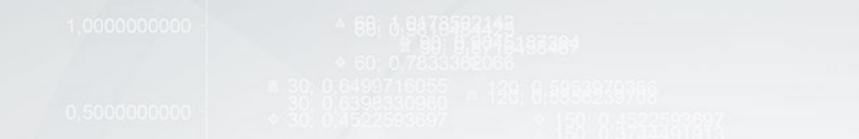
- Animals: cat:kitten, dog:bark, ...
- Geography: Athens:Greece, Peru:Spanish, ...
- People: Lincoln:president, Lincoln:American, ...
- Other: blood:red, actor:actress, ...

# Performance by categories

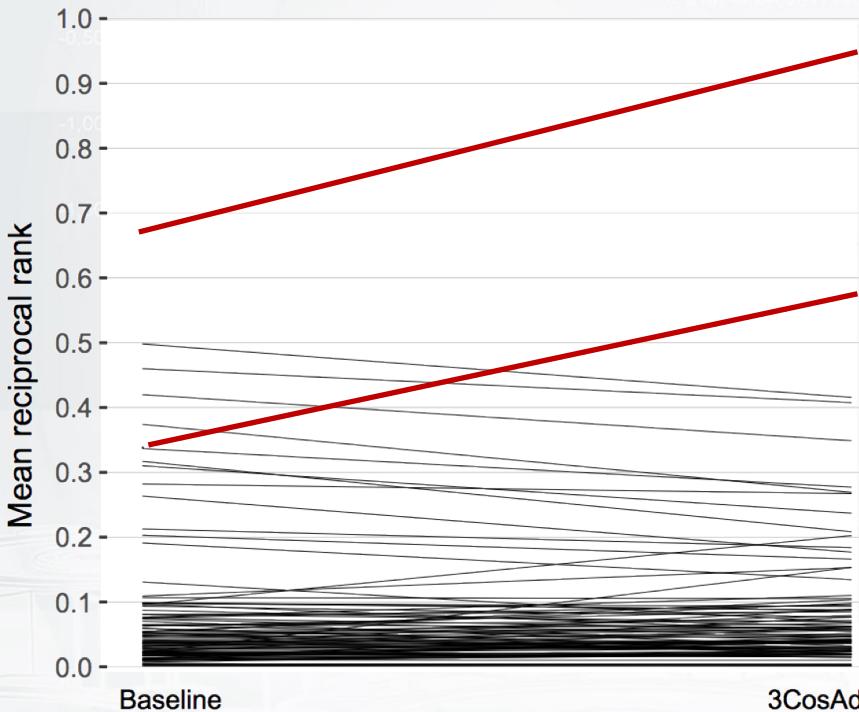


Finley et. al. What Analogies Reveal about Word Vectors and their Compositionality, 2017.

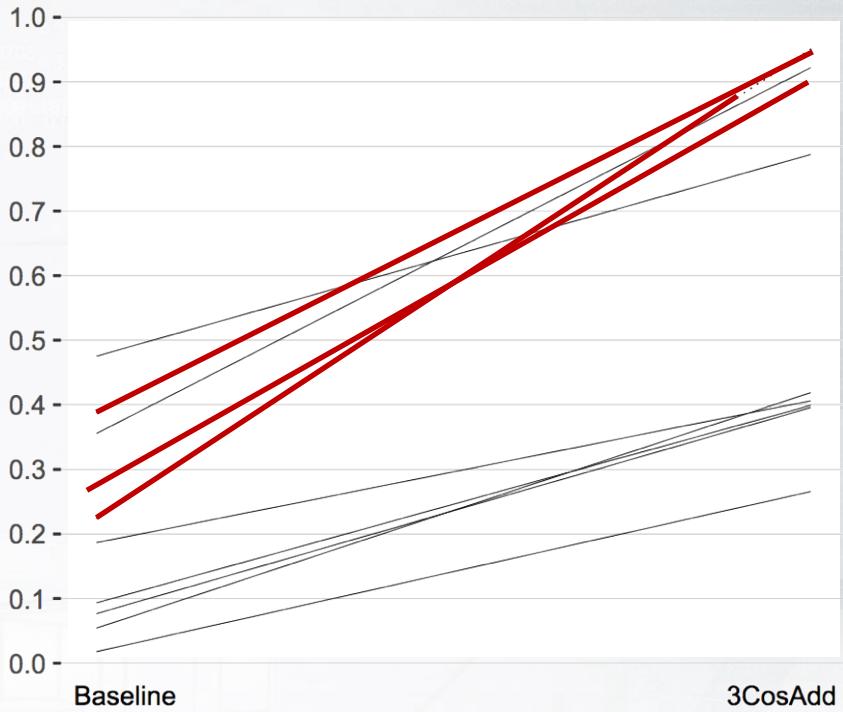
# Genders and counties are cherry-picks



Lexical semantics



Named entities



Finley et. al. What Analogies Reveal about Word Vectors and their Compositionality, 2017.

# Resume

- *word2vec* works fine for word similarities
- But there are many questions with word analogies
- Be careful about hype!