


The Geometry of Meaning: Analysing Text with Word Embeddings

Tomáš Musil, <http://tomasm.cz/slides/embs>

 June 17, 2022

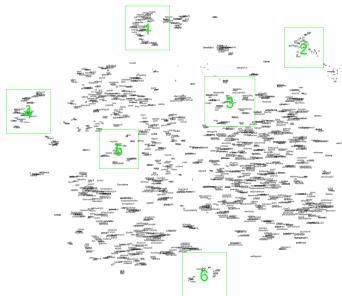


Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics



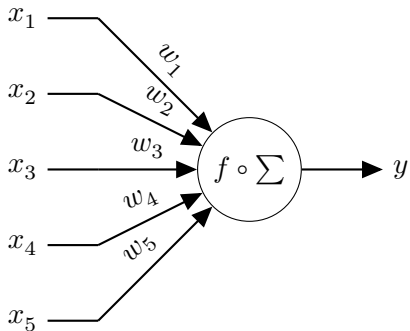
unless otherwise stated

Neural Networks and Word
Embeddings
Examples from Neural Networks
Vector Arithmetic and Word2vec
GloVe
Applications in Humanities
Structuralism



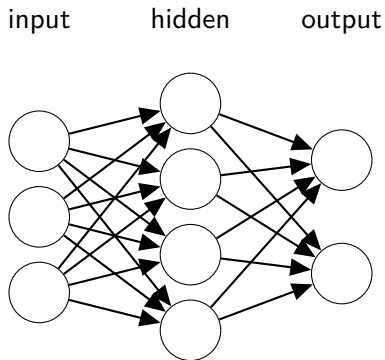
Neural Networks and Word Embeddings

Mathematical “neuron”

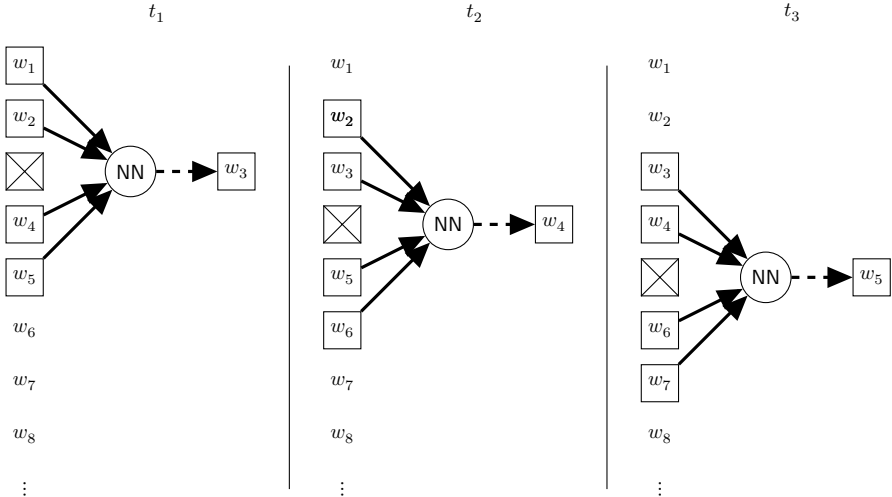


$$y = f\left(\sum_i w_i x_i\right)$$

Simple Neural Network



Language model



Ordered Vocabulary

...

456. ovšem

457. protože

458. rovněž

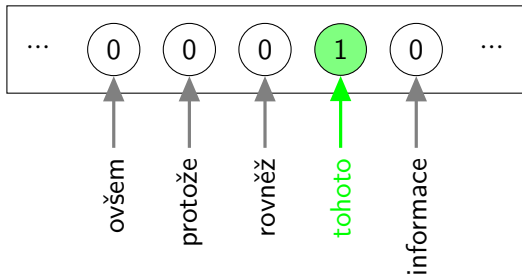
459. tohoto

460. informace

...

Numbers have order, we need *vectors*.

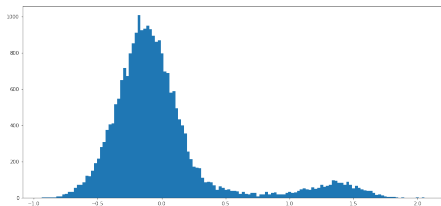
Vector Representations of Words



This is discrete, we need *continuity*.

Examples from Neural Networks

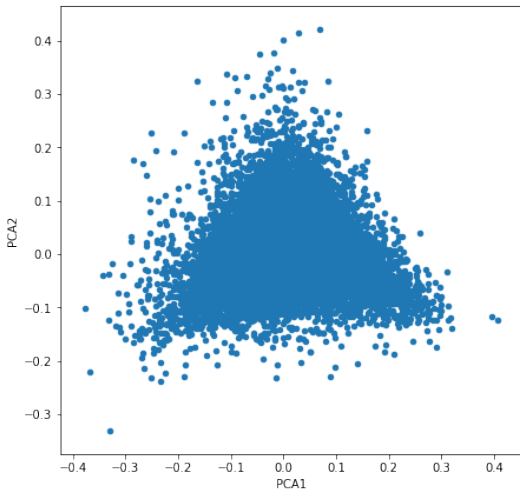
Word Embeddings in NMT: proper names



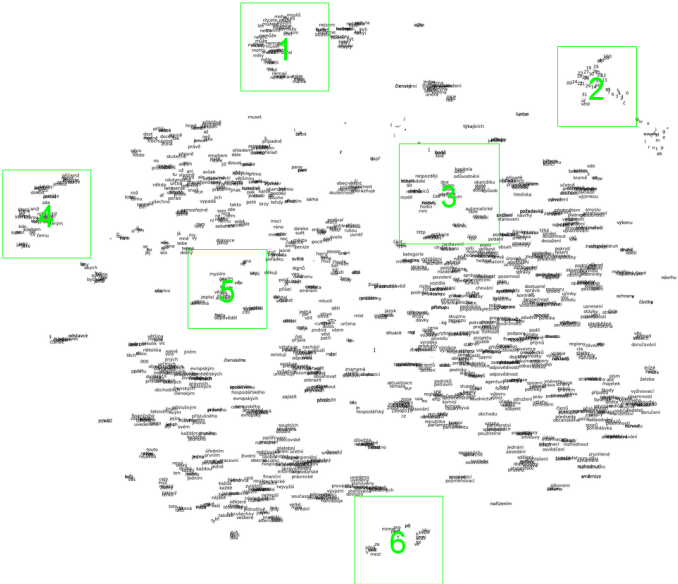
volant okně zátylek křeslem ostrůvek lucernu oknu lavičku krbem polštáři
lahvičku psovi silnicí gauči ostrovem brzdu halou prostěradlo botou
kočáru

Newt Toranaga Paithan Bremen Baniči Blake Sadi Sacharóza Vanion
Chronos Jaku Frigate Albert Culum Richards Duun Rozšafín Tynian
Anheg Hawat

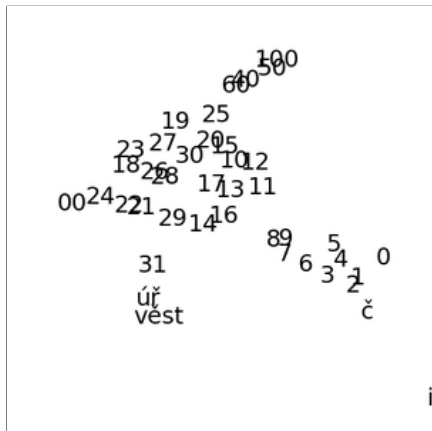
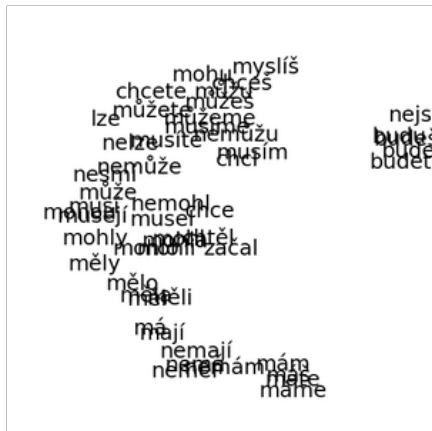
Word Embeddings in Sentiment



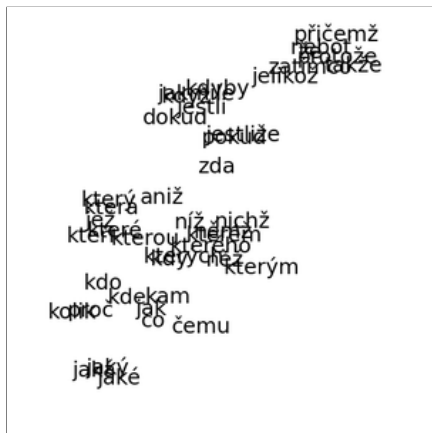
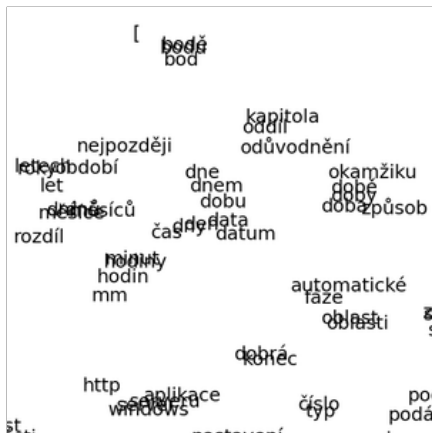
Vector Space of Meaning



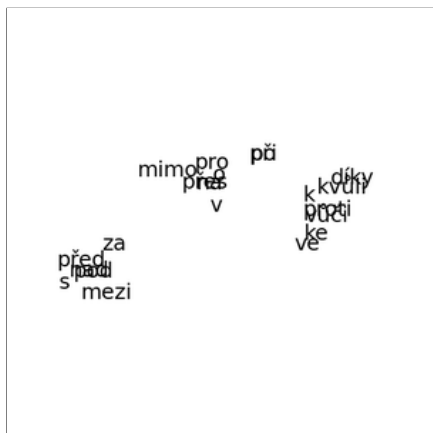
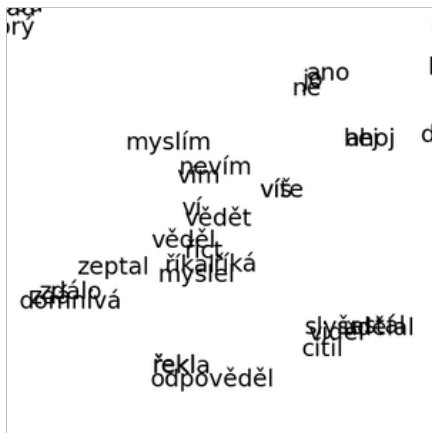
Vector Space of Meaning



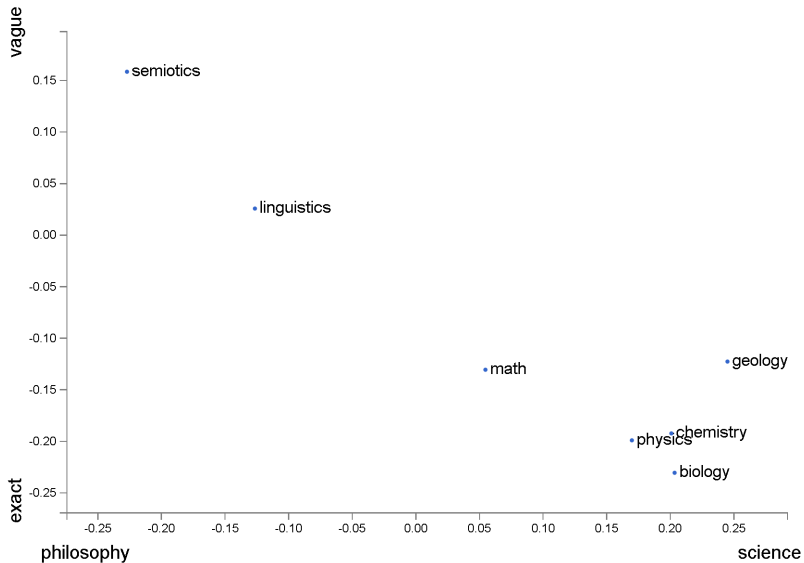
Vector Space of Meaning



Vector Space of Meaning



Vector Arithmetics of Meaning



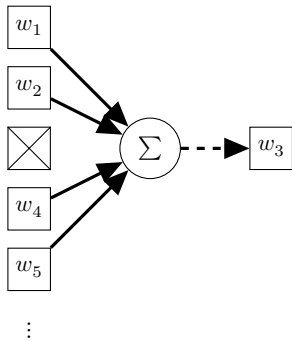
Interactive Analysis

`https://projector.tensorflow.org/`

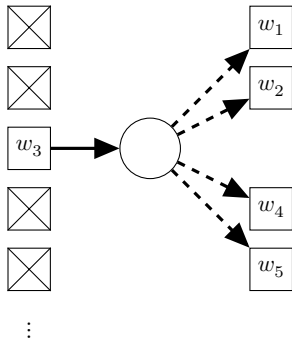
Vector Arithmetic and Word2vec

Word2vec Variants

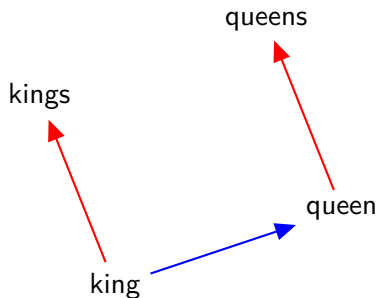
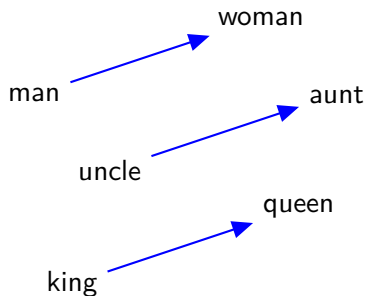
CBOW



Skip-gram



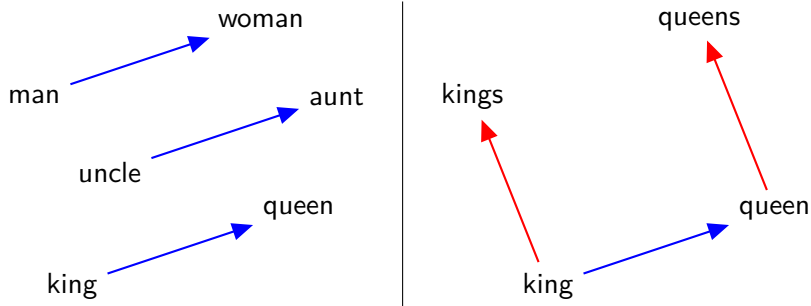
Word2vec



Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. [Linguistic regularities in continuous space word representations.](#)

In *HLT-NAACL*, pages 746–751, 2013

Word2vec

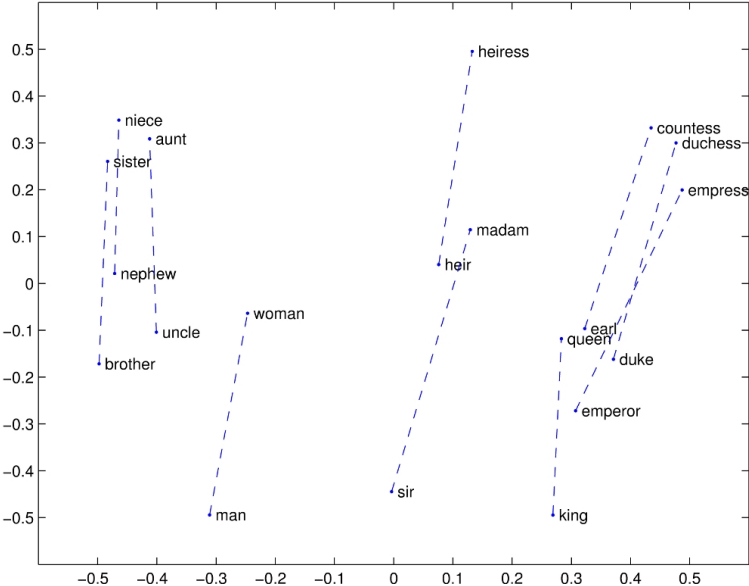


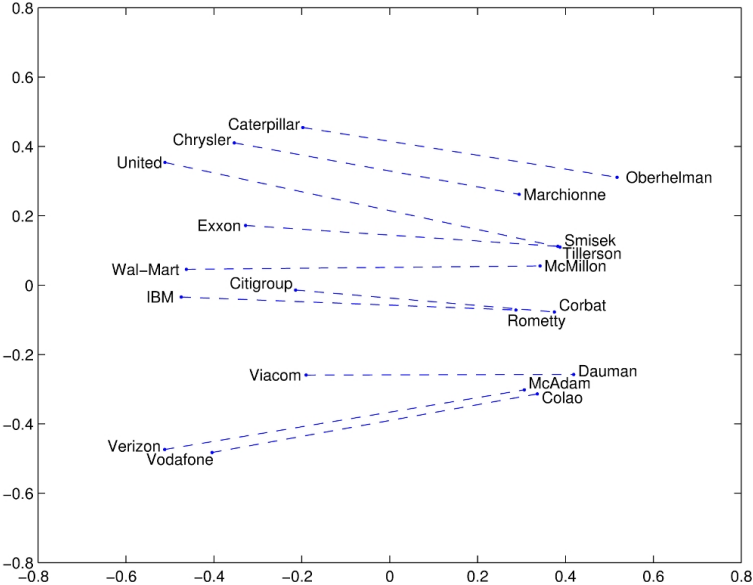
$$v_{king} - v_{man} + v_{woman} \approx v_{queen}$$

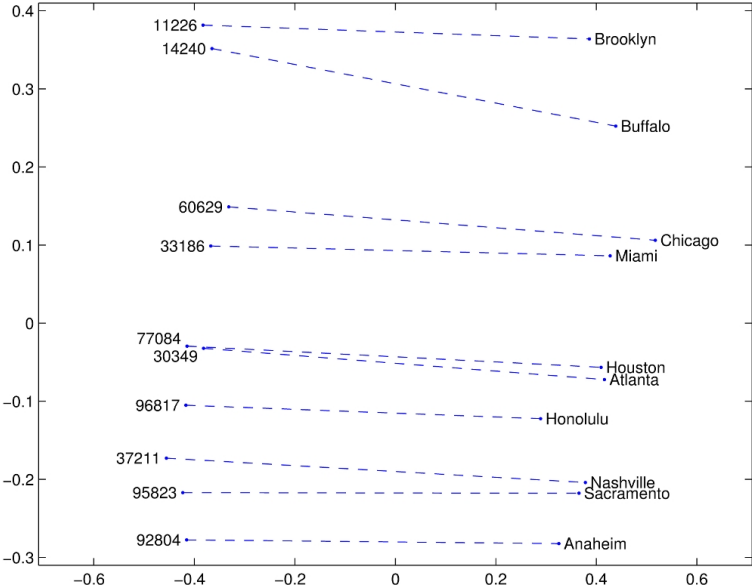
Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. [Linguistic regularities in continuous space word representations.](#)

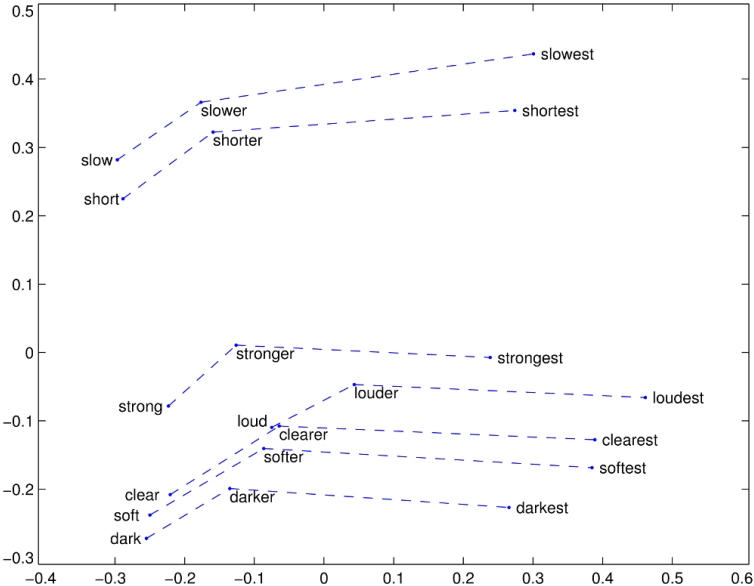
In *HLT-NAACL*, pages 746–751, 2013

GloVe









Games with Word Embeddings

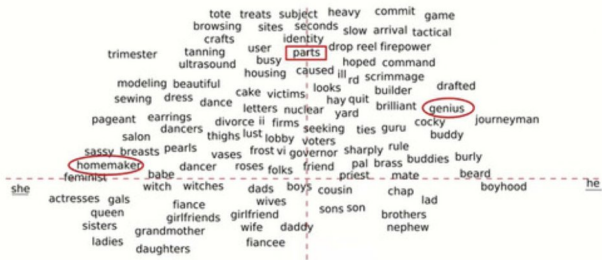
`https://semantle.com/`

`https://research.google.com/semantris/`

Applications in Humanities

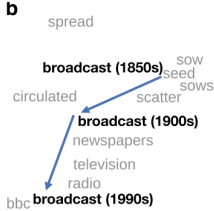
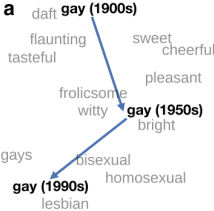


Gender Bias in Word Embeddings.



Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings
Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai, in NeurIPS, 2016.

Word Paths



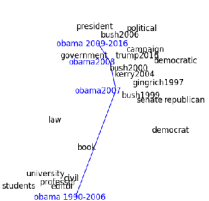
Word Paths



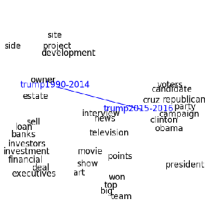
(a) apple



(b) amazon



(c) obama



(d) trump

Polarization

We Don't Speak the Same Language:

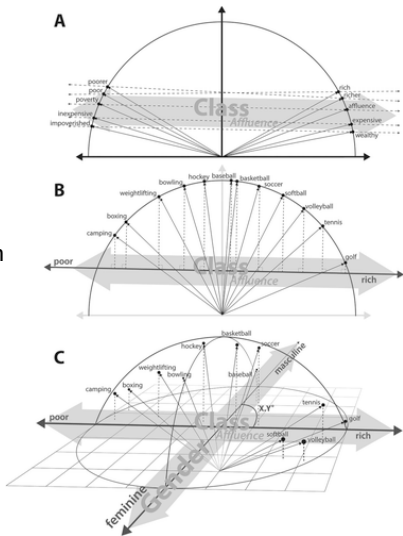
Interpreting Polarization through Machine Translation

KhudaBukhsh, A.R., Sarkar, R., Kamlet, M.S., and Mitchell, T.M. (2021).

| Category | Misaligned pairs |
|--------------------|--|
| Political entities | \langle democrats, republicans \rangle , \langle nunes, schiff \rangle , \langle dem, republican \rangle , \langle dnc, gop \rangle , \langle kushner, burisma \rangle , \langle gop, democrats \rangle , \langle flynn, hillary \rangle |
| News entities | \langle fox, cnn \rangle , \langle hannity, cuomo \rangle , \langle tapper, hannity \rangle , \langle tucker, cuomo \rangle |
| Derogatory | \langle trumptards, snowflakes \rangle , \langle chump, trump \rangle , \langle liberals, libtards \rangle , \langle pelosi, pelousy \rangle , \langle obamas, obummer \rangle , \langle cooper, giraffe \rangle , \langle biden, creep \rangle , \langle schiff, schitt \rangle , \langle barr, weasel \rangle |
| (Near) synonyms | \langle lmao, lol \rangle , \langle allegations, accusations \rangle , \langle puppet, stooge \rangle , \langle bs, bullshit \rangle , \langle potus, president \rangle , \langle hahaha, lol \rangle |
| Spelling errors | \langle mueller, muller \rangle , \langle kavanaugh, cavanaugh \rangle , \langle hillary, hilary \rangle , \langle isreal, israel \rangle |
| Ideological | \langle kkk, blm \rangle , \langle christianity, multiculturalism \rangle , \langle sham, impeachment \rangle , \langle antifa, nazi \rangle , \langle liberals, conservatives \rangle , \langle communism, nazism \rangle , \langle leftists, fascists \rangle , \langle liberalism, conservatism \rangle , \langle communists, nazis \rangle , \langle immigrants, illegals \rangle |

The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings

Kozlowski, Austin C., et al. (2019)



Programming Exercise

`http://tomasm.cz/slides/embs`

`https://colab.research.google.com/drive/
1AGa2Q-JSAcF4TLBc1lqUlAKaN2xV4QpB?usp=sharing`

Structuralism

Syntagms

(horizontal)

Paradigms

(vertical)

| | | | | | |
|-------|-------|---------|-------|-----------|---------|
| Their | tiger | stood | by | my | friend. |
| That | wolf | sat | over | your | tree. |
| The | dog | walked | on | the | floor. |
| My | cat | bounced | under | Nicole's | mat. |
| Your | cow | jumped | above | Anthony's | fridge. |

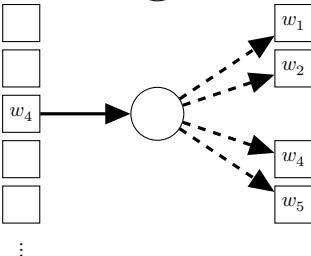
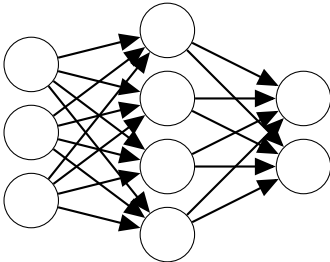
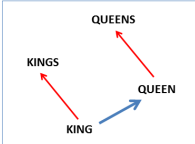
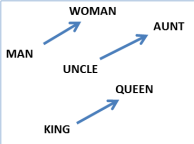
The elements of a structure have neither extrinsic designation, nor intrinsic signification. Then what is left? As Levi-Strauss recalls rigorously, they have nothing other than a sense: a sense which is necessarily and uniquely “positional”.

– G. Deleuze, How do we recognize structuralism?

Vector Space of Meanings



Thank you for your attention



References I

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *HLT-NAACL*, pages 746–751, 2013.