

DEEP LEARNING IN (ACTION | NLP)

Sebastian Ebert
CIS, University of Munich

June 24, 2015

NLP

(NATURAL LANGUAGE PROCESSING)

SEARCH ENGINE

 bing Who's the president of Germany? 

Web Images Videos Maps News Explore

5,700,000 RESULTS Any time ▾



Joachim Gauck
Germany · President

SEARCH ENGINE

 bing Who is he married to? 

Web Images Videos Maps News Explore

15,900 RESULTS Any time ▾

Gerhild Gauck

(Since 1959)
Joachim Gauck · Spouse

SEARCH ENGINE

 bing How old is his wife? X

Web Images Videos Maps News Explore

7,470,000 RESULTS Any time ▾

Images of how old is his wife?

bing.com/images

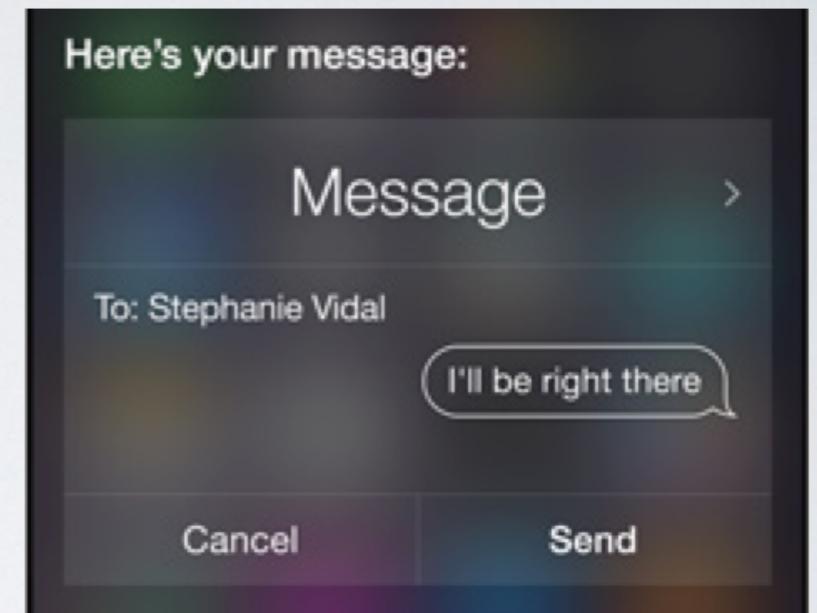


[See more images of how old is his wife?](#)

SIRI, CORTANA



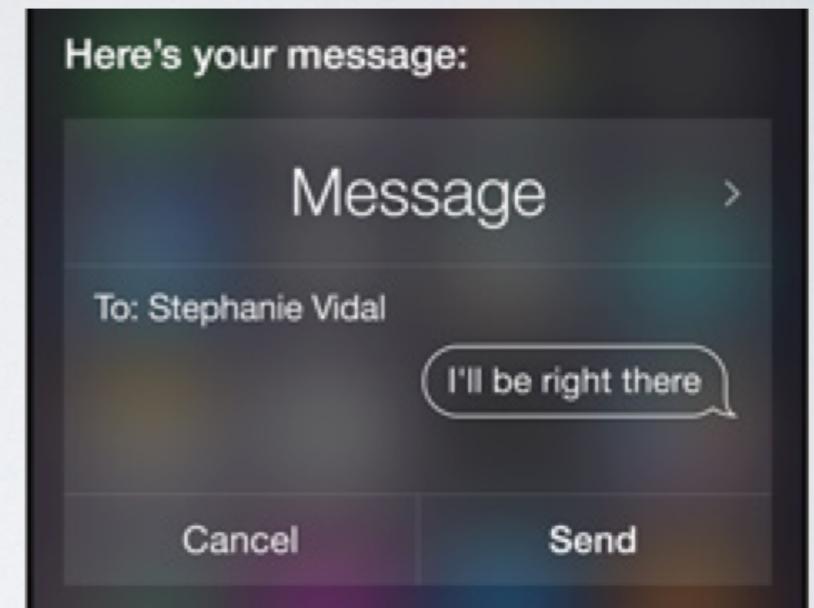
"Tell Stephanie 'I'll
be right there'"



SIRI, CORTANA



"Tell Stephanie 'I'll
be right there'"

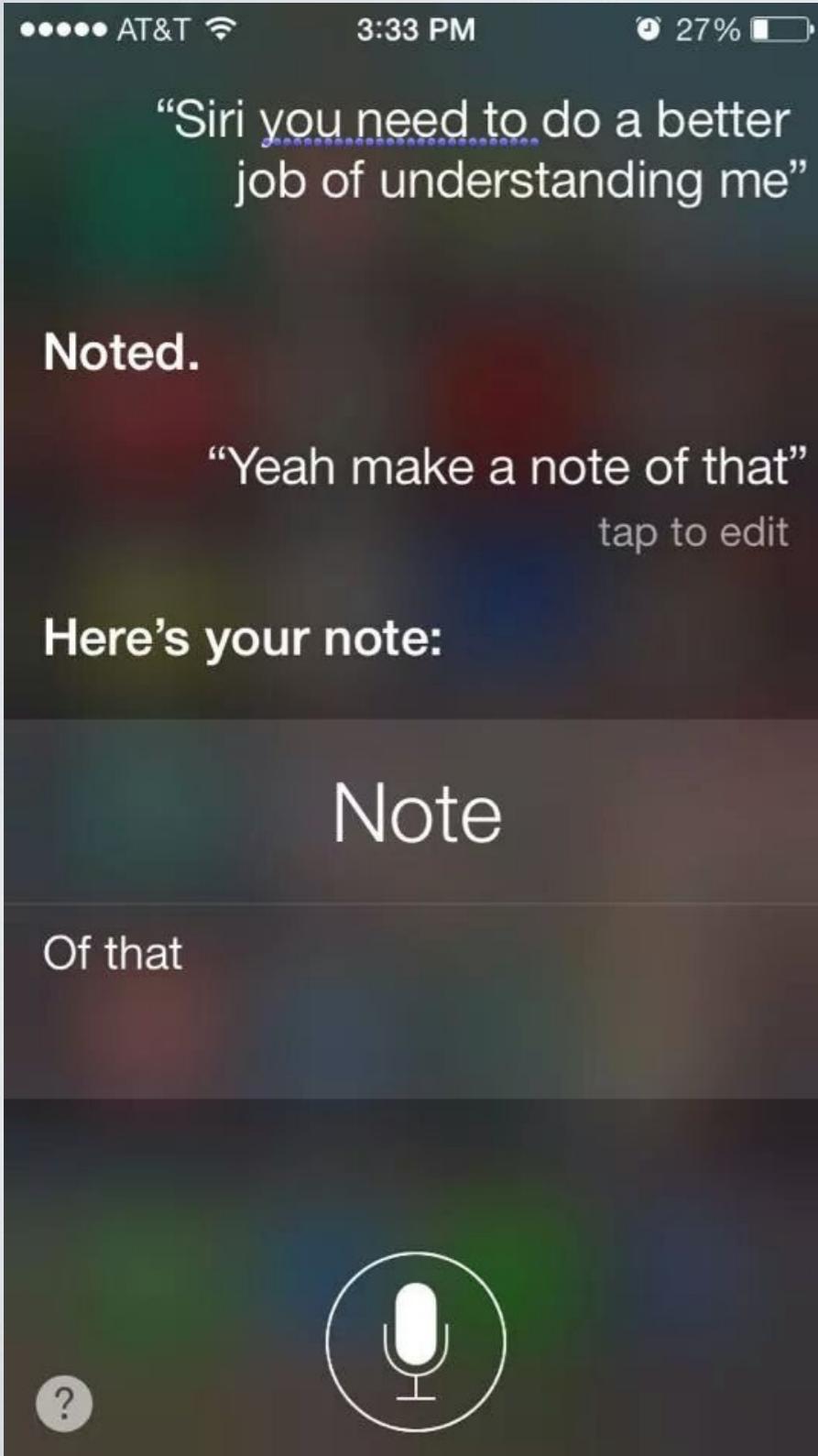
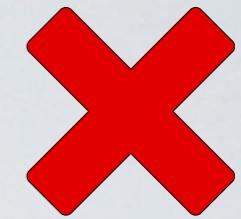


"Who won the
Mets game?"

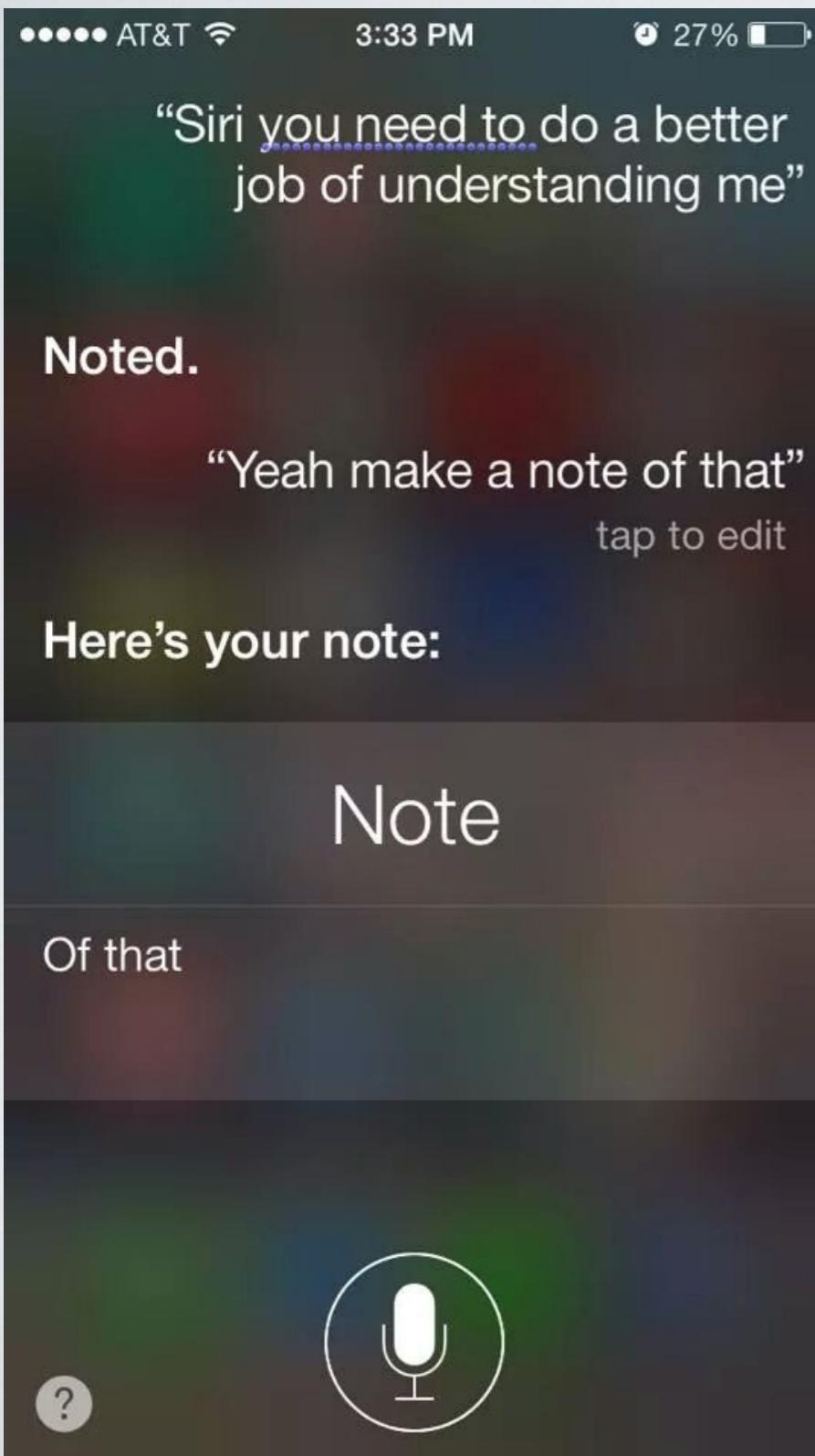
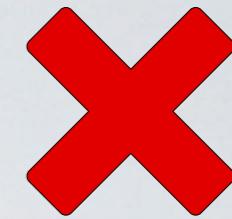


Here's the score for the last New York Mets game.			
Thu Apr 10 - FINAL			
Mets (4-5)	6		
Braves (5-4)		4	

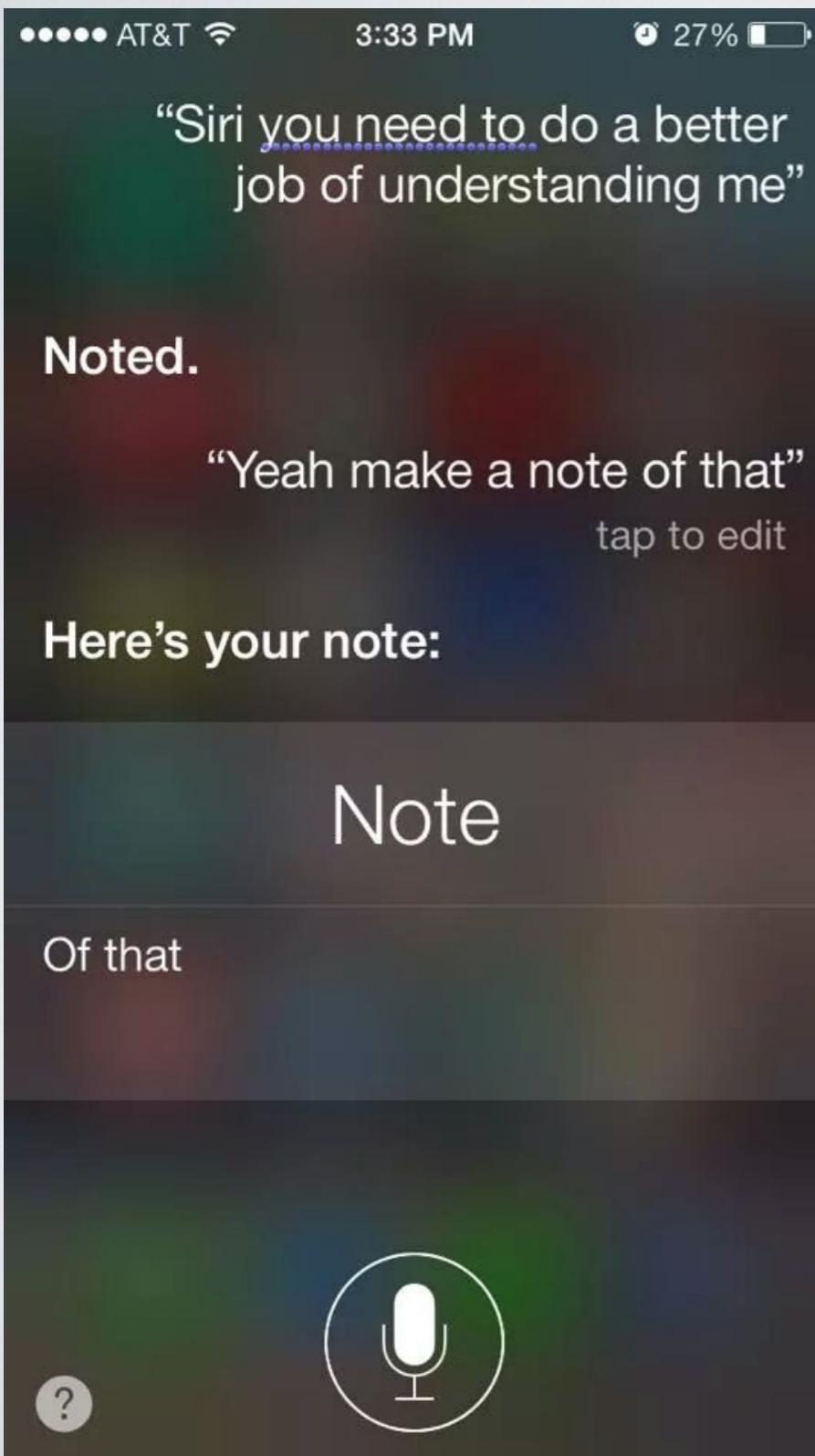
SIRI FAIL



SIRI FAIL



SIRI FAIL



MACHINE TRANSLATION

English Spanish German Detect language ▾  English Spanish German ▾ Translate

Wherever you are in the world, visiting a bookshop is always a treat.  Wo immer Sie in der Welt sind, den Besuch einer Buchhandlung ist immer ein Genuss.

es regnet wie aus Kübeln  it's raining cats and dogs

jedem das seine!  To each his own 

MACHINE TRANSLATION

English Spanish German Detect language ▾



English Spanish German ▾

Translate

Wherever you are in the world, visiting a bookshop is always a treat.



Wo immer Sie in der Welt sind, den Besuch einer Buchhandlung ist immer ein Genuss.

es regnet wie aus Kübeln



it's raining cats and dogs

jedem das seine



To each his own



Dudelsackpfeifenmachergesellenprüfung

Bagpipe pipe maker journeyman

Hast du den Boden sauber gemacht?



Did you do the floor clean?



Nimm mir bitte die Tasche, die ich in der rechten Hand halte, ab.



Please take my bag, which I hold in my right hand, from.

IBM WATSON



SENTIMENT ANALYSIS

RESTAURANT



Brittany C.

San Francisco, CA

501 friends

112 reviews

Elite '15



11/5/2014

2 check-ins

Great Korean place in Los Altos. The interior is spacious, clean, and well-decorated. The service is attentive, fast, and friendly. And of course, the food is really good, too:

interior

staff

location

food

overall

RESTAURANT



Brittany C.

San Francisco, CA

501 friends

112 reviews

Elite '15



11/5/2014

2 check-ins

Great Korean place in Los Altos. The interior is spacious, clean, and well-decorated. The service is attentive, fast, and friendly. And of course, the food is really good, too:

interior

staff

location

food

overall

RESTAURANT



Brittany C.

San Francisco, CA

501 friends

112 reviews

Elite '15



11/5/2014

2 check-ins

Great Korean place in Los Altos. The interior is spacious, clean, and well-decorated. The service is attentive, fast, and friendly. And of course, the food is really good, too:

interior

spacious, clean, well-decorated



staff

location

food

overall

RESTAURANT



Brittany C.

San Francisco, CA

501 friends

112 reviews

Elite '15



11/5/2014

2 check-ins

Great Korean place in Los Altos. The interior is spacious, clean, and well-decorated. The service is attentive, fast, and friendly. And of course, the food is really good, too:

interior

spacious, clean, well-decorated



staff

location

food

overall

RESTAURANT



Brittany C.

San Francisco, CA

501 friends

112 reviews

Elite '15



11/5/2014

2 check-ins

Great Korean place in Los Altos. The interior is spacious, clean, and well-decorated. The service is attentive, fast, and friendly. And of course, the food is really good, too:

interior

spacious, clean, well-decorated



staff

attentive, fast, friendly



location

food

overall

RESTAURANT



Brittany C.

San Francisco, CA

501 friends

112 reviews

Elite '15



11/5/2014

2 check-ins

Great Korean place in Los Altos. The interior is spacious, clean, and well-decorated. The service is attentive, fast, and friendly. And of course, the food is really good, too:

interior

spacious, clean, well-decorated



staff

attentive, fast, friendly

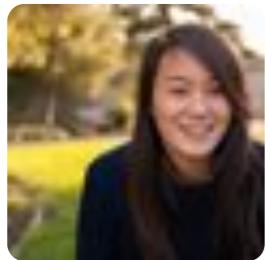


location

food

overall

RESTAURANT



Brittany C.

San Francisco, CA

501 friends

112 reviews

Elite '15



11/5/2014

2 check-ins

Great Korean place in Los Altos. The interior is spacious, clean, and well-decorated. The service is attentive, fast, and friendly. And of course, the food is really good, too:

interior

spacious, clean, well-decorated



staff

attentive, fast, friendly



location

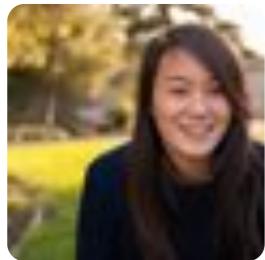
food

really good



overall

RESTAURANT



Brittany C.

San Francisco, CA

501 friends

112 reviews

Elite '15



11/5/2014

2 check-ins

Great Korean place in Los Altos. The interior is spacious, clean, and well-decorated. The service is attentive, fast, and friendly. And of course, the food is really good, too:

interior

spacious, clean, well-decorated



staff

attentive, fast, friendly



location

food

really good



overall



RESTAURANT



Brittany C.

San Francisco, CA

501 friends

112 reviews

Elite '15



11/5/2014

2 check-ins

Great Korean place in Los Altos. The interior is spacious, clean, and well-decorated. The service is attentive, fast, and friendly. And of course, the food is really good, too:

interior

spacious, clean, well-decorated



staff

attentive, fast, friendly



location

?

food

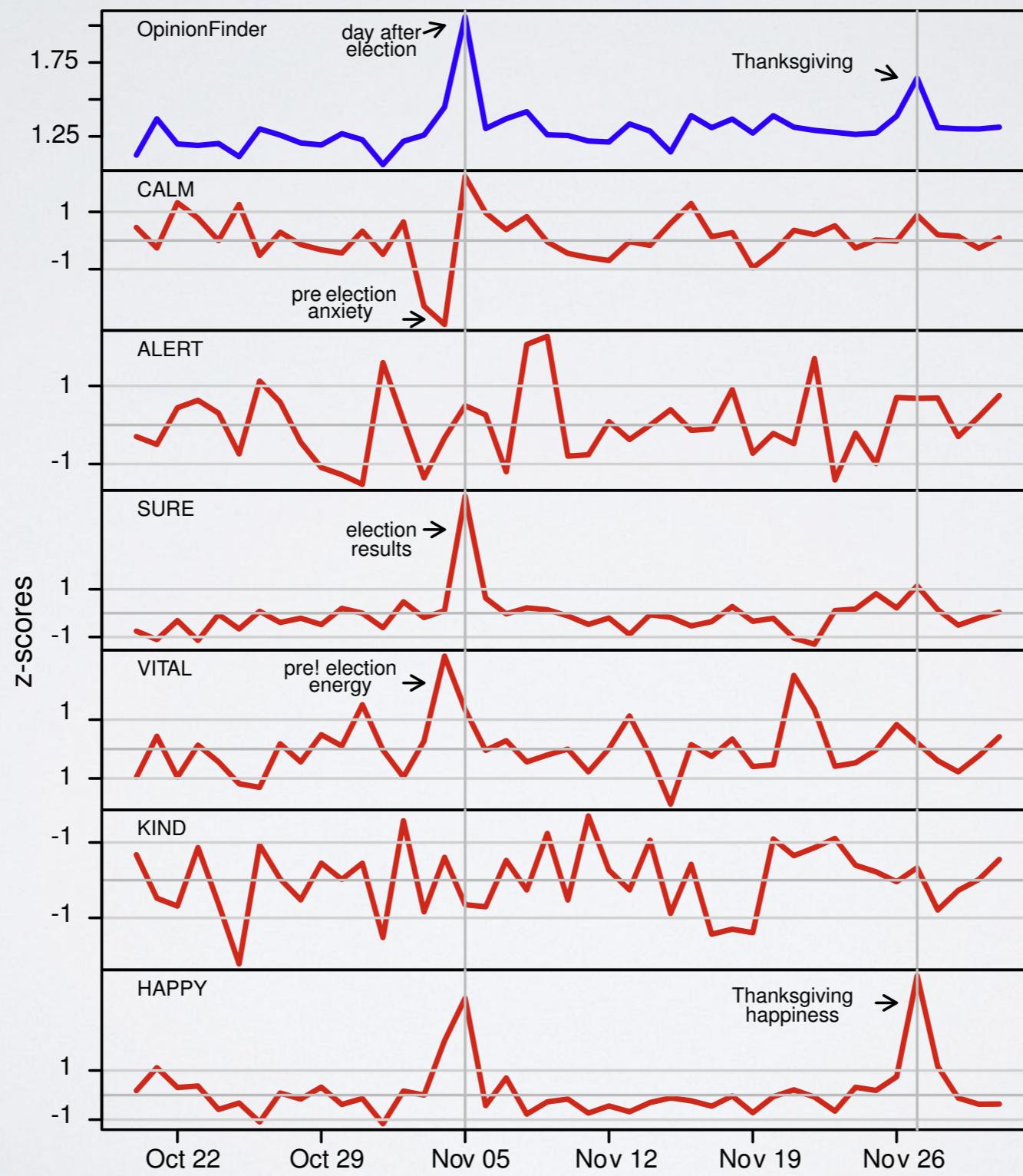
really good



overall

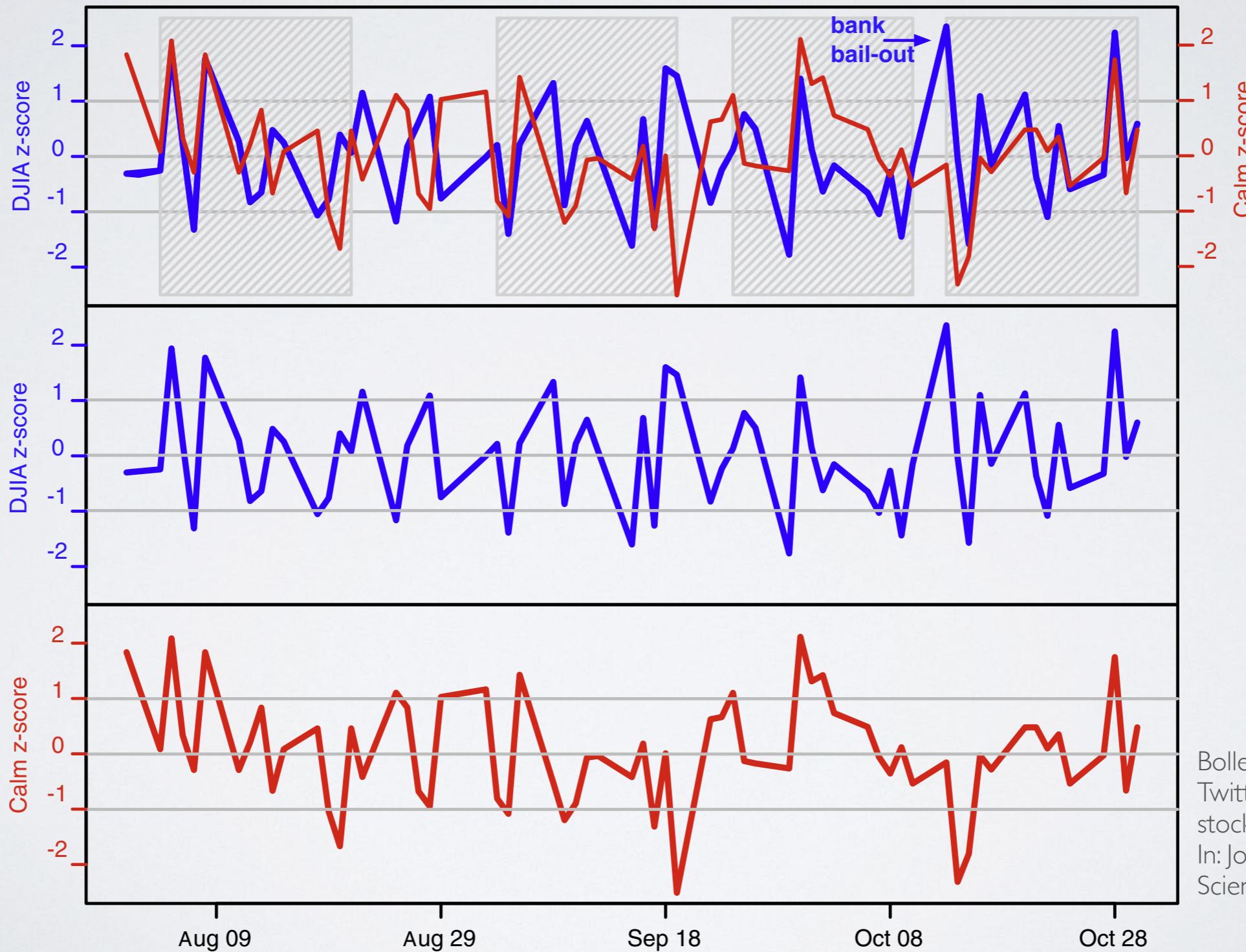


TWITTER EMOTIONS

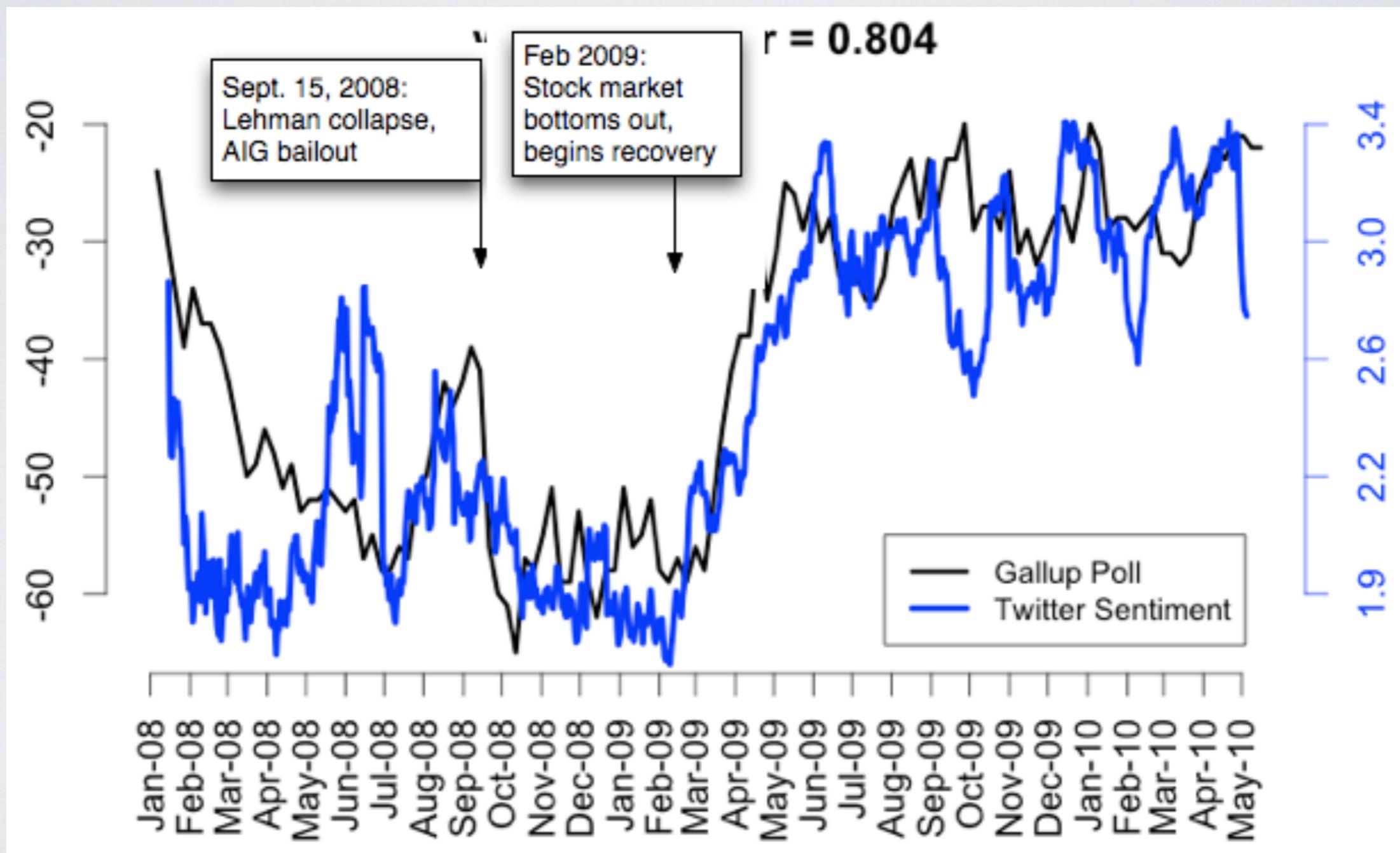


Bollen et al.
Twitter mood predicts the stock market
In: Journal of Computational Science 2
(2011)

TWITTER STOCK MARKET



TWITTER CONSUMER CONFIDENCE



O'Connor et al.

From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series

In: ICWSM 2010

MOVIES

- "Weak and ridiculous"

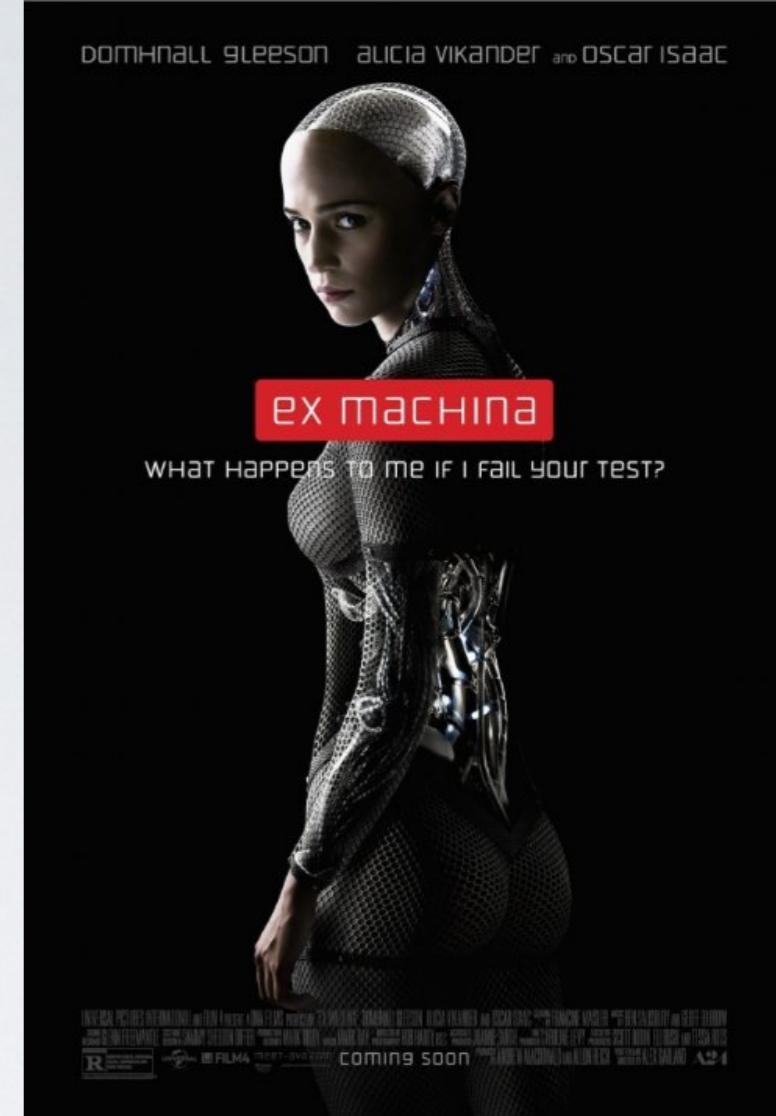


- "Too Slow. Too Nude. Interesting"



- "First of all Ava is absolutely stunning"





- "Weak and ridiculous" 😕
- "Too Slow. Too Nude. Interesting" 😐
- "First of all Ava is absolutely stunning" 😊

THE HARD PART

- misspelling: "Such a **wierd** book"
- negation: "**not** really **good**"
- domain dependency: "go read the book"
- social media text: "whr go sux? life is sooo beautiful!"

THE "HARD" PART

THE "HARD" PART

meaning	example	sentiment
1 firm, stiff	hard floor	neu
2 difficult	hard question	neg
3 intensely	work hard	neu
4 intense	hard look	neu
5 unkind	hard man	neg
6 definitely true	hard truth	neu
7 hard-rock-type music	hard beats	neu
8 opposite of soft transition	hard edge	neu
9 negative phrase	hard drugs	neg
10 neutral phrase	hard disk	neu

WHAT WE WANT

"First of all Ava is absolutely stunning"

WHAT WE WANT

"First of all Ava is **absolutely stunning**"

- context sensitive (cf. "hard")

WHAT WE WANT

"First of all Ava is **absolutely stunning**"

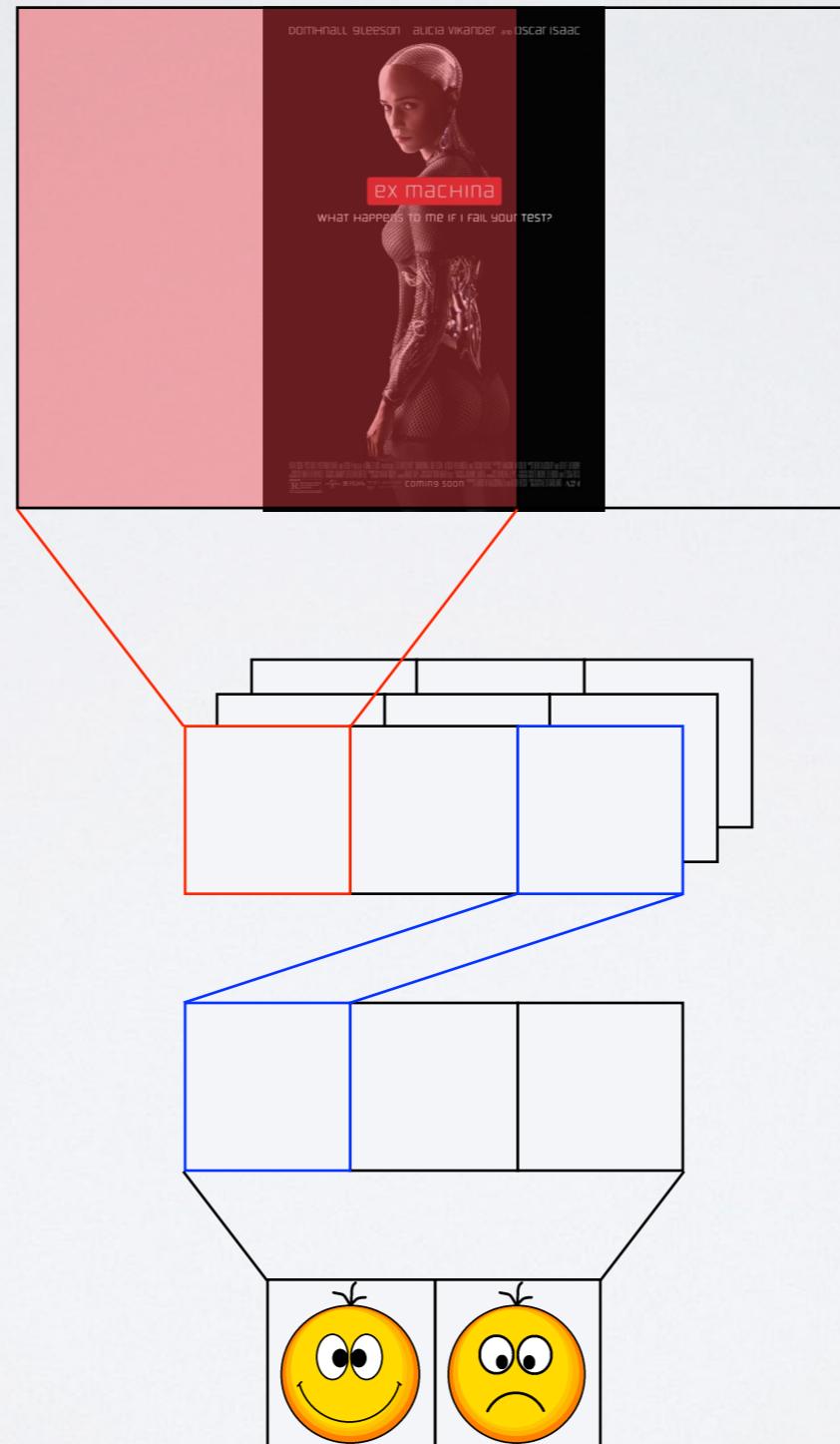
- context sensitive (cf. "hard")
- length independent

WHAT WE WANT

"First of all Ava is **absolutely stunning**"

- context sensitive (cf. "hard")
- length independent
- only some words are important

CNN



input layer

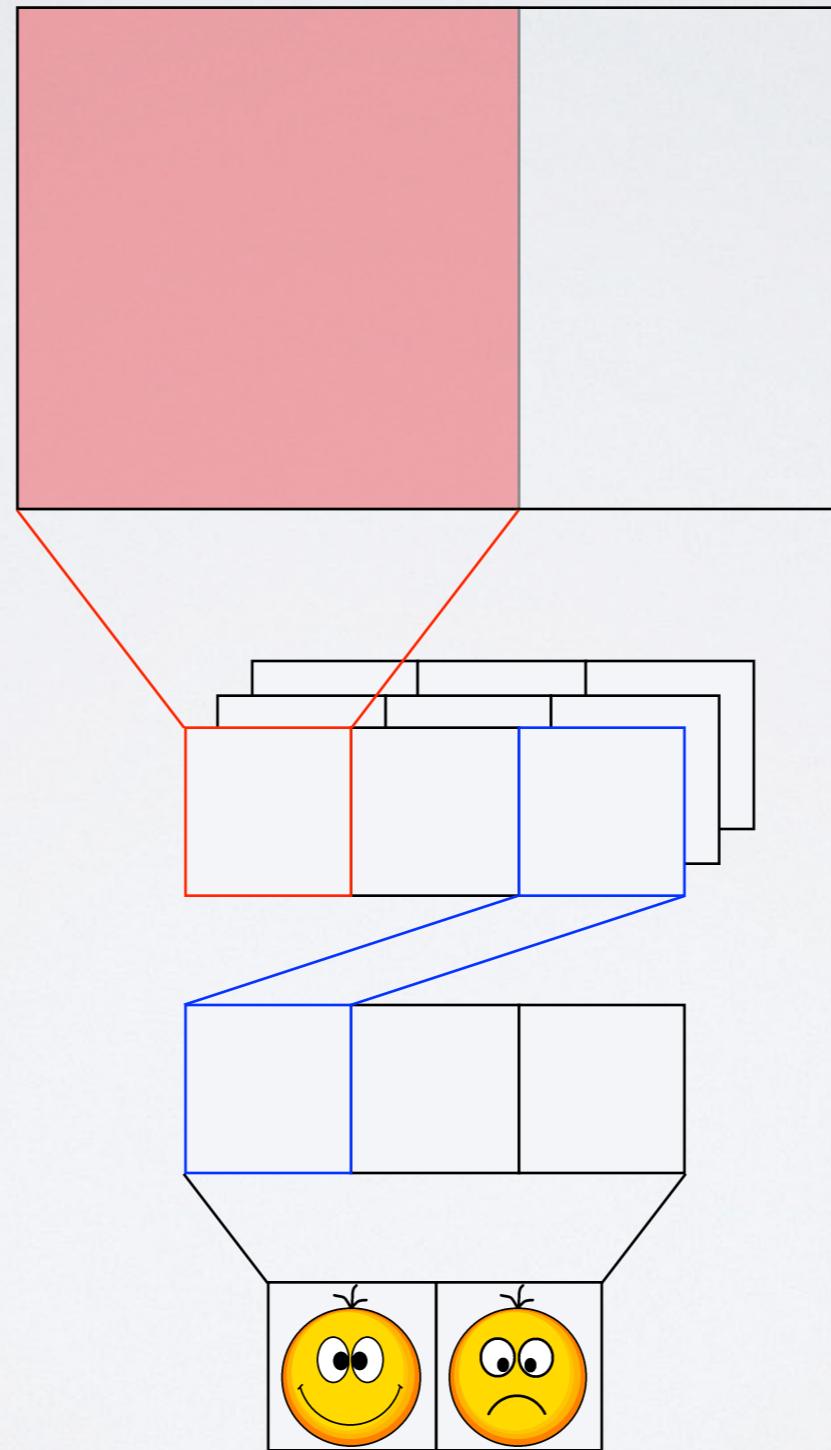
convolution layer

pooling layer

softmax layer

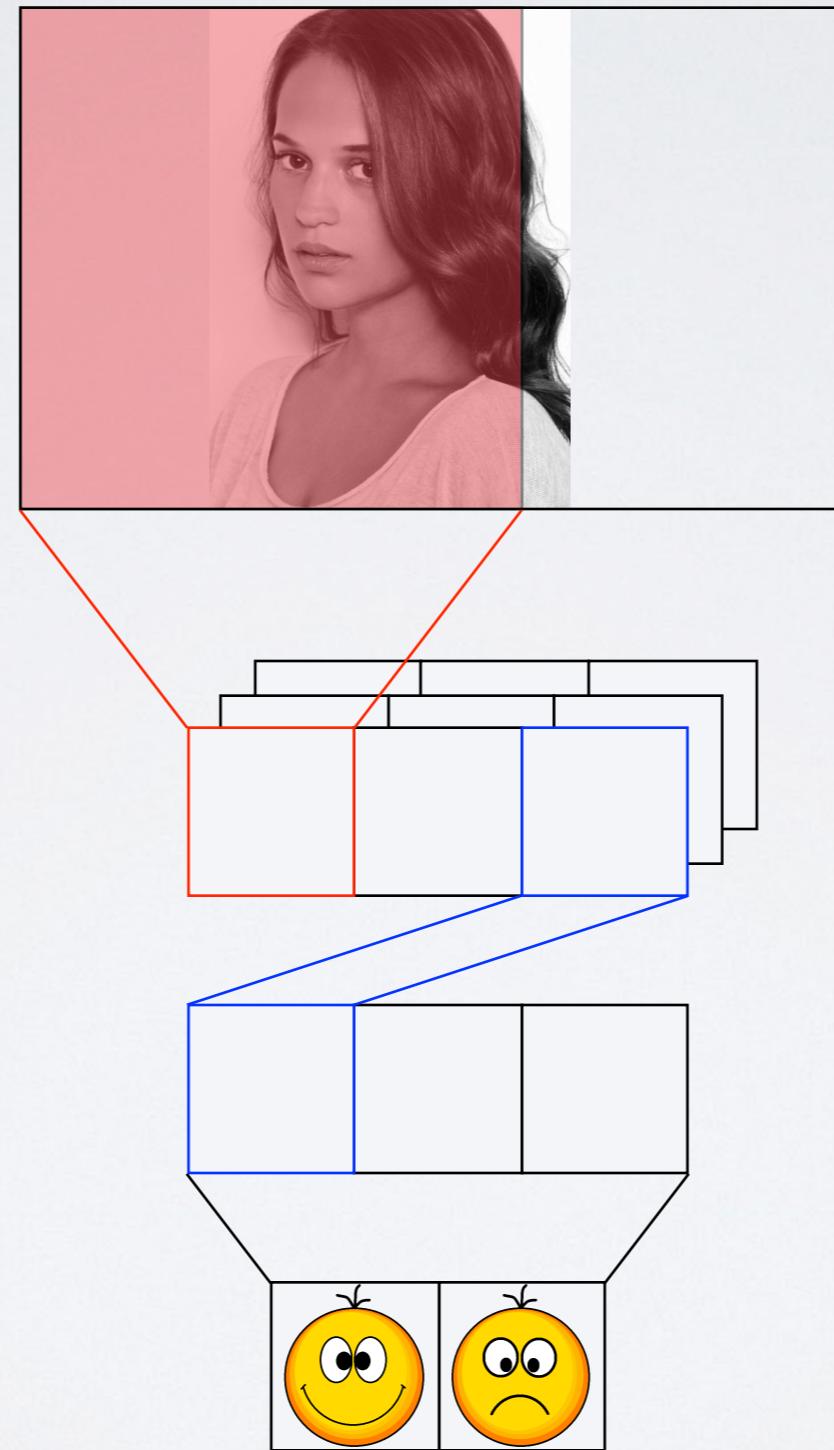
CNN FOR TEXT

"First of all
Ava is
absolutely
stunning"



CNN FOR TEXT

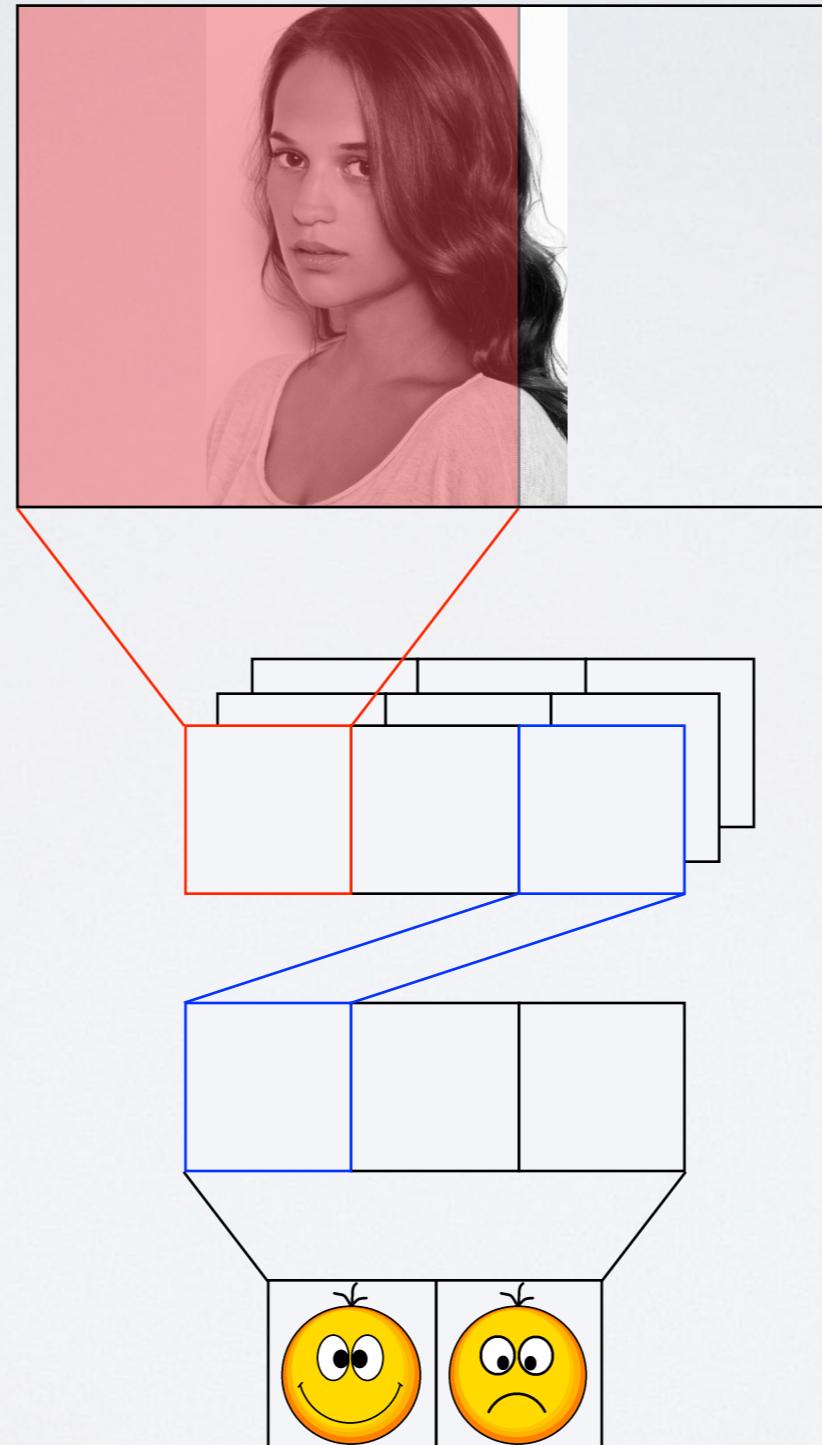
"First of all
Ava is
absolutely
stunning"



CNN FOR TEXT

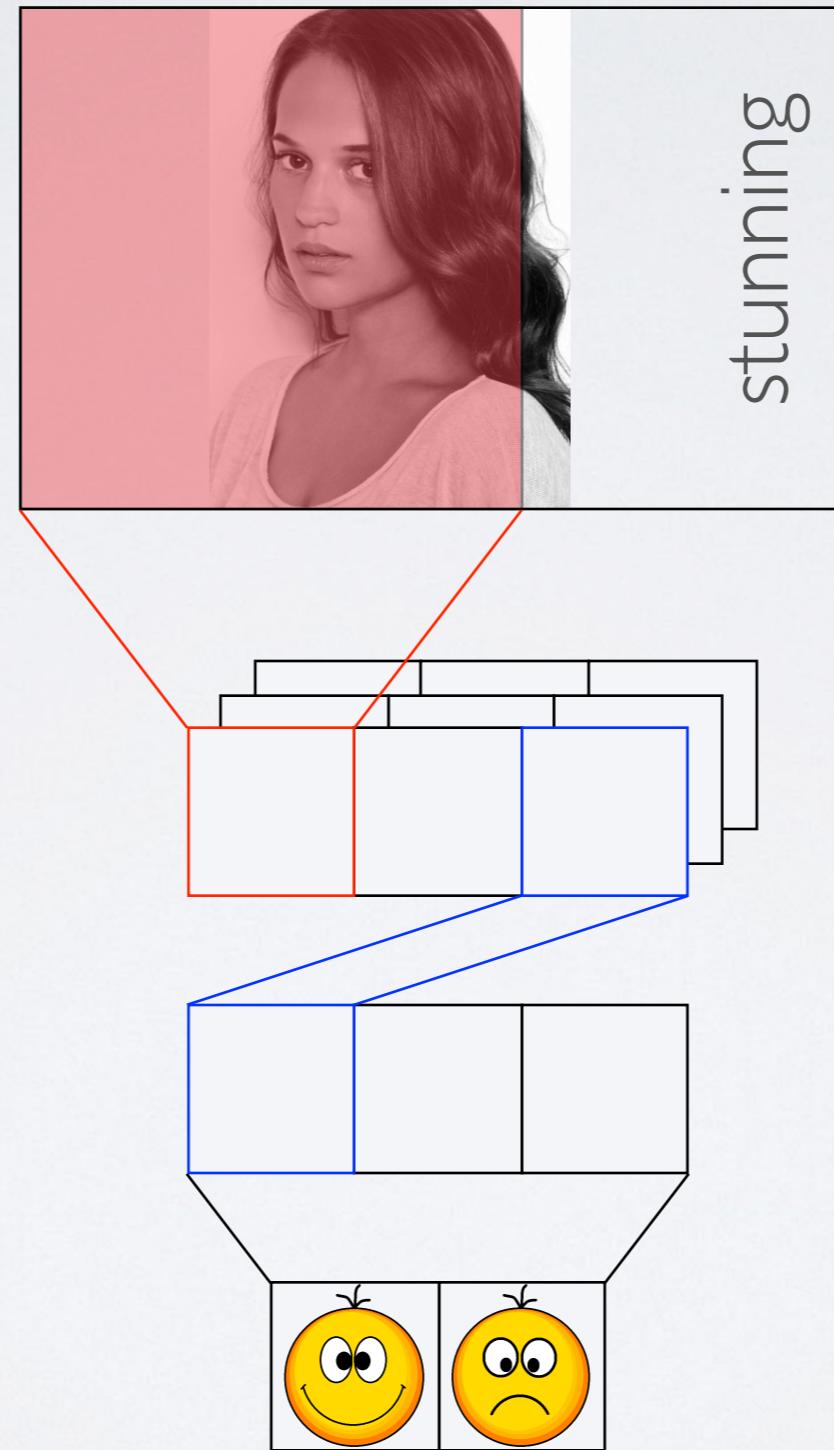
"First of all
Ava is
absolutely
stunning"

stunning



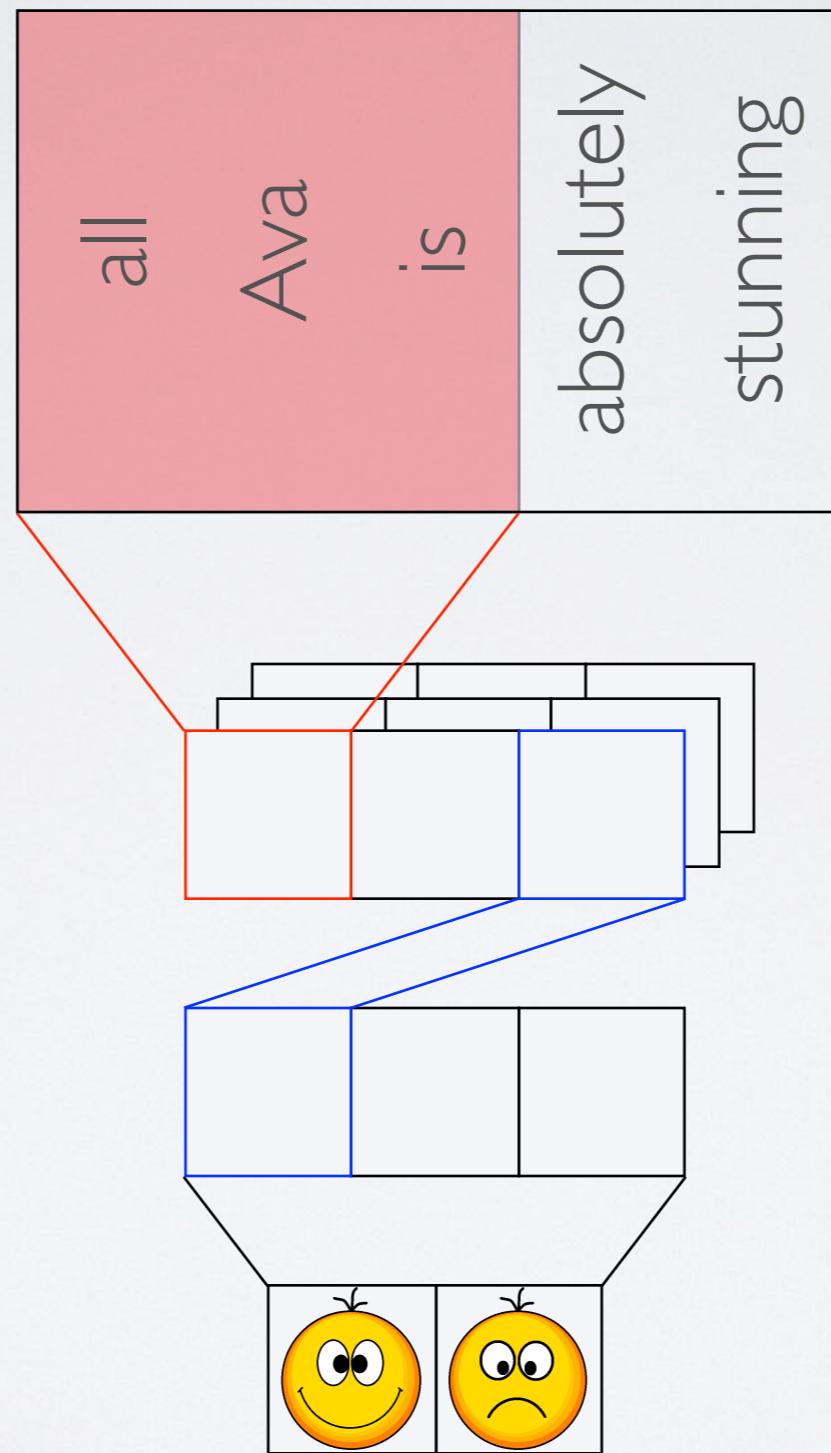
CNN FOR TEXT

"First of all
Ava is
absolutely
stunning"



CNN FOR TEXT

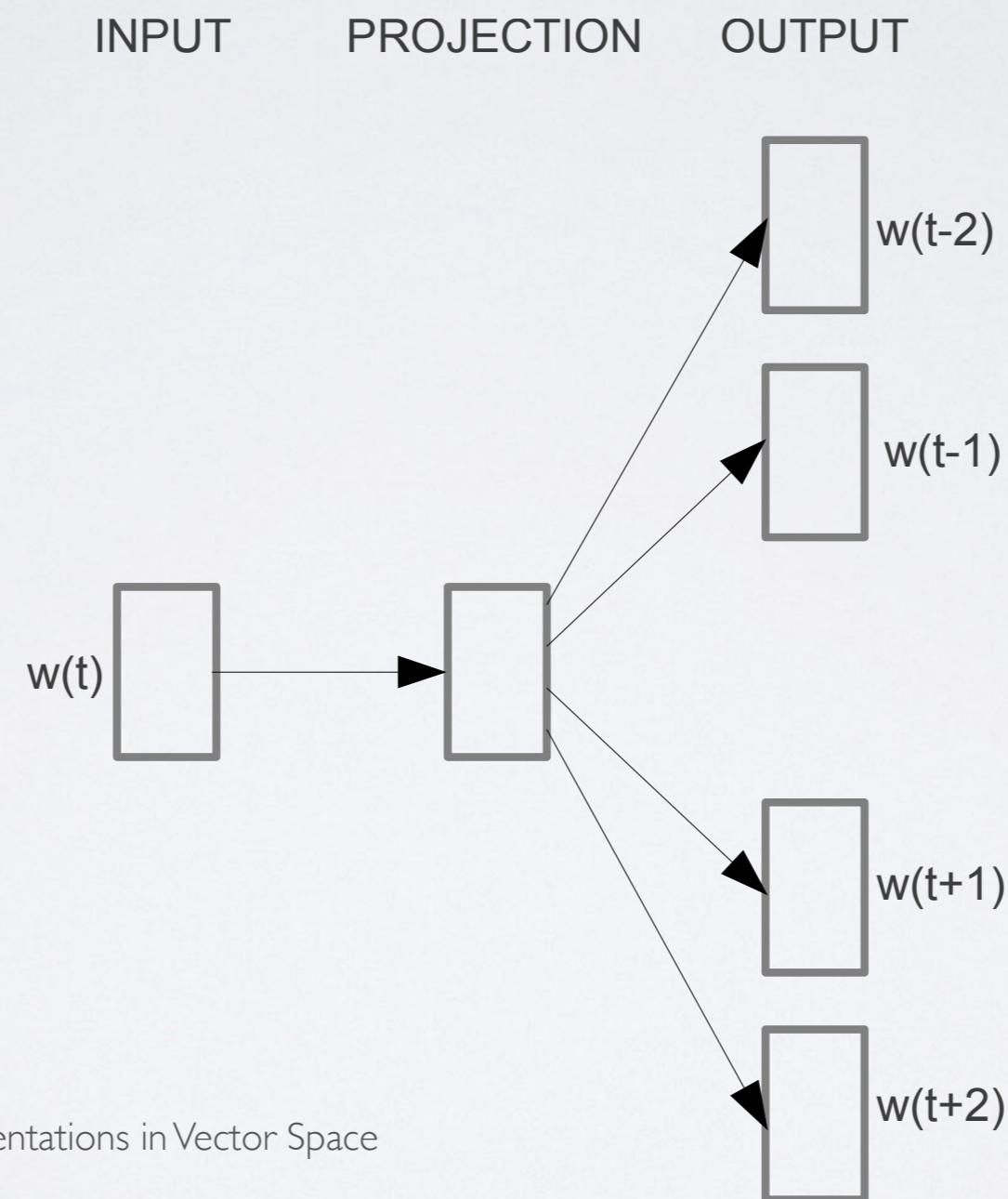
"First of all
Ava is
absolutely
stunning"



WORD REPRESENTATIONS

CONTINUOUS SKIP-GRAM

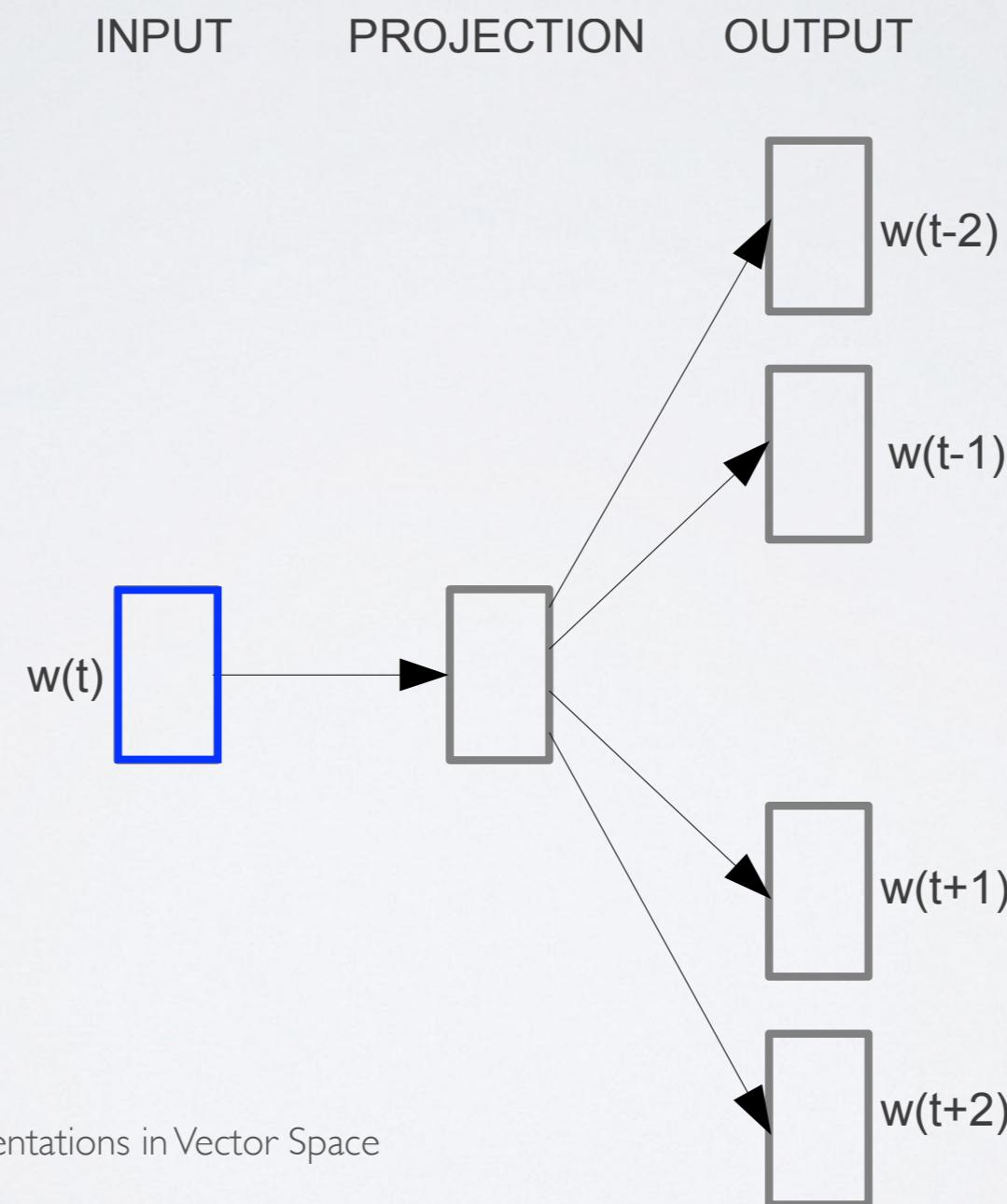
"First of all
Ava is
absolutely
stunning"



Mikolov et al.
Efficient Estimation of Word Representations in Vector Space
In: ICLR 2013

CONTINUOUS SKIP-GRAM

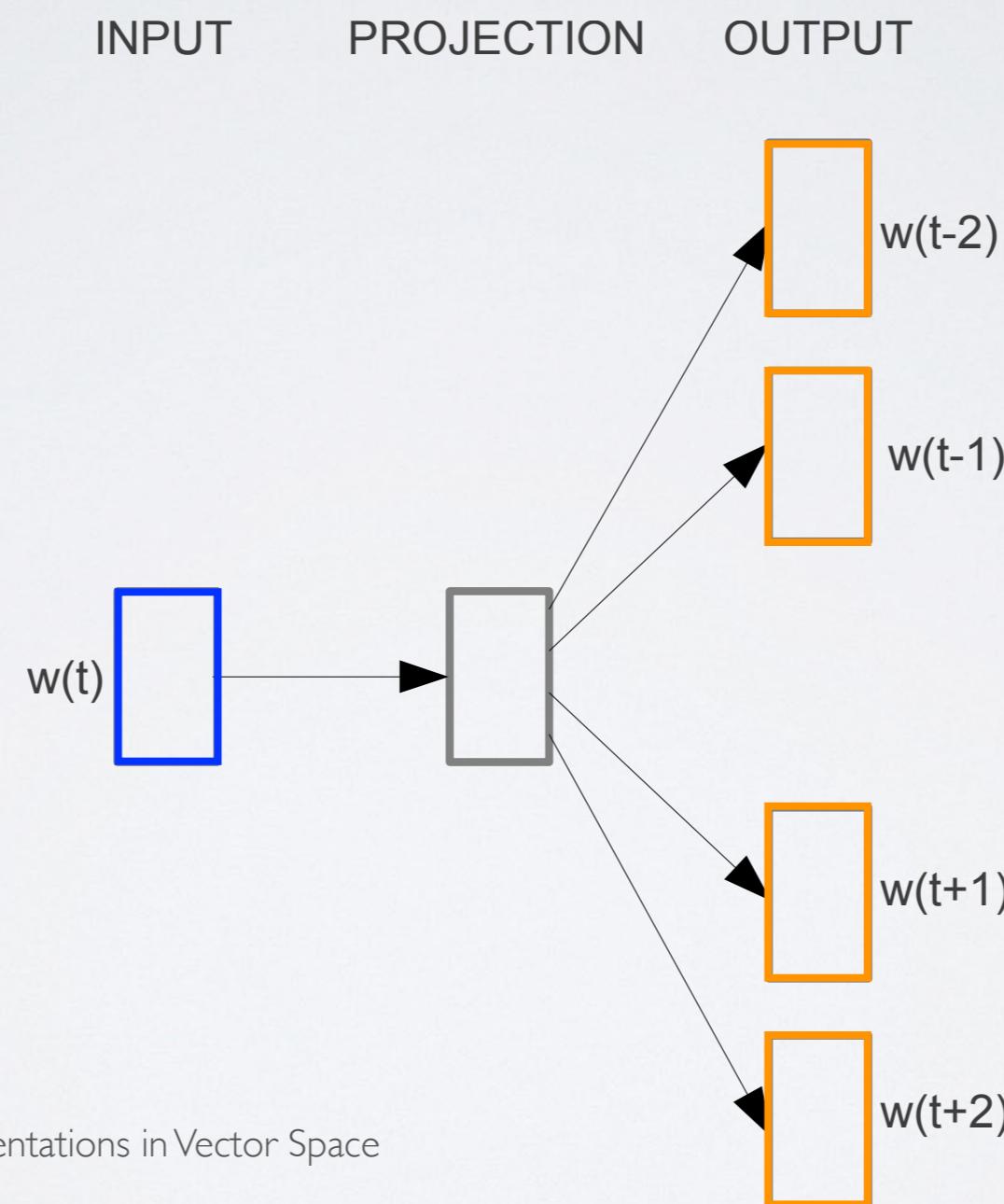
"First of all
Ava is
absolutely
stunning"



Mikolov et al.
Efficient Estimation of Word Representations in Vector Space
In: ICLR 2013

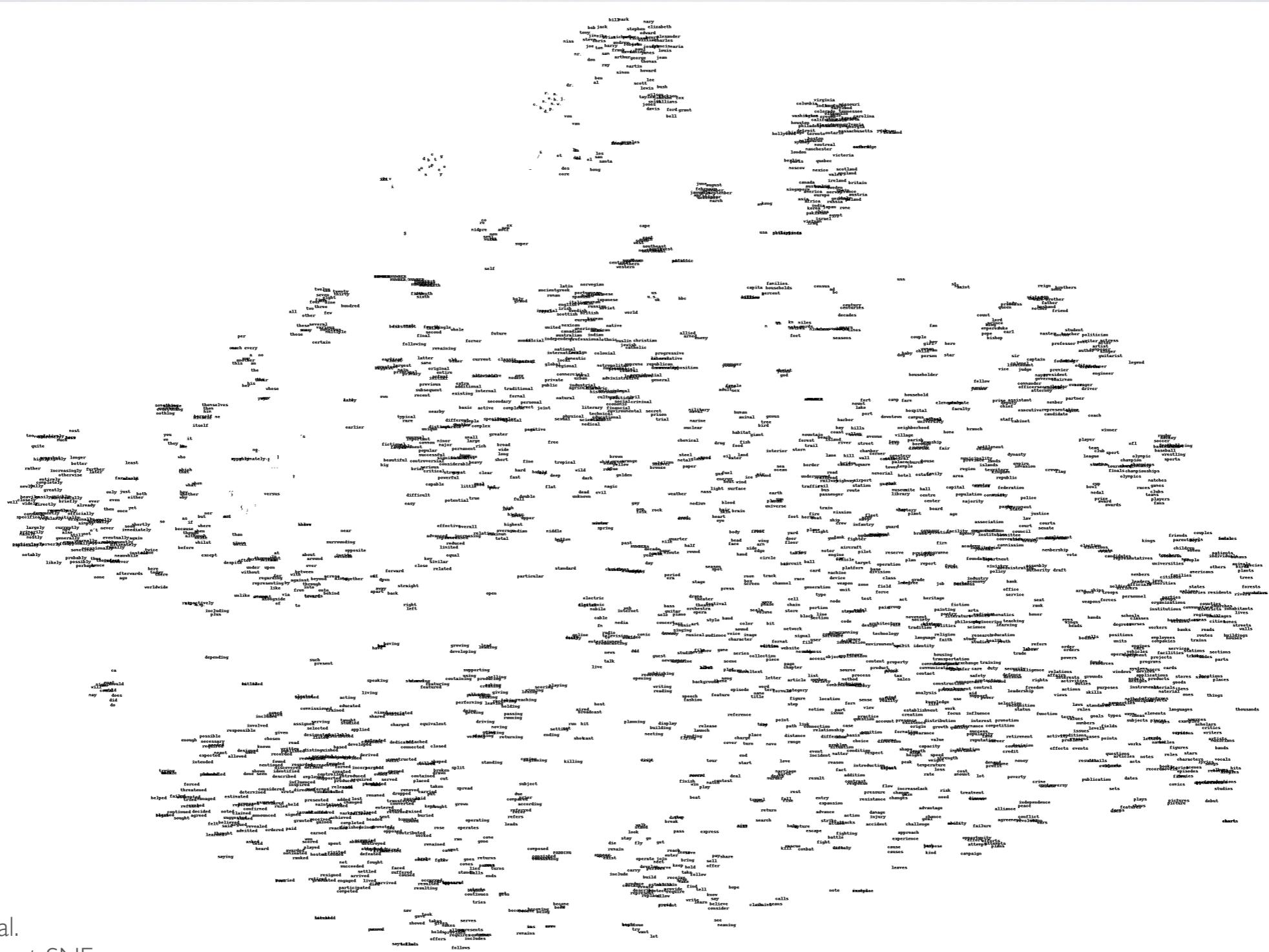
CONTINUOUS SKIP-GRAM

"First of all
Ava is
absolutely
stunning"



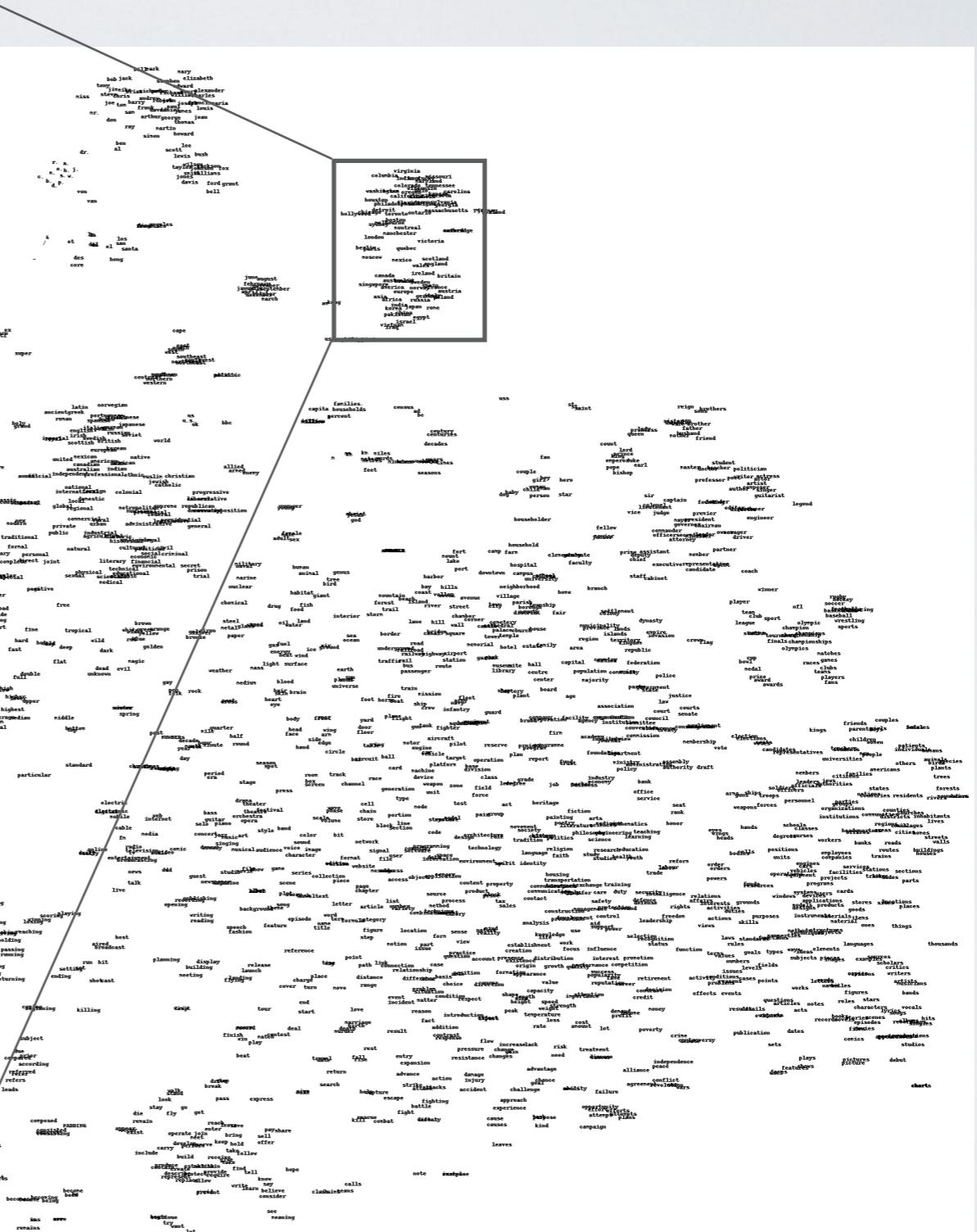
Mikolov et al.
Efficient Estimation of Word Representations in Vector Space
In: ICLR 2013

WORD CLUSTERS

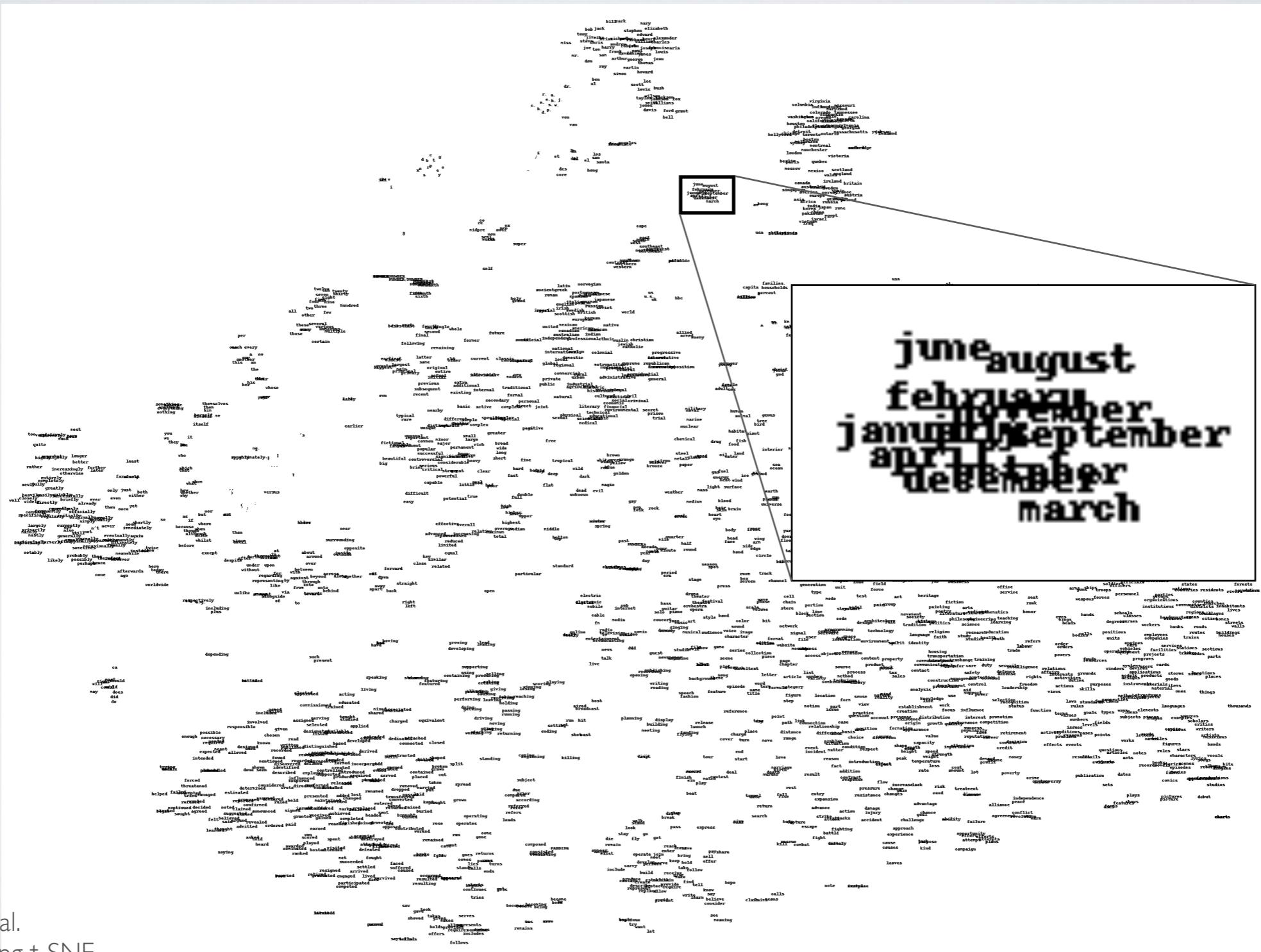


WORD CLUSTERS

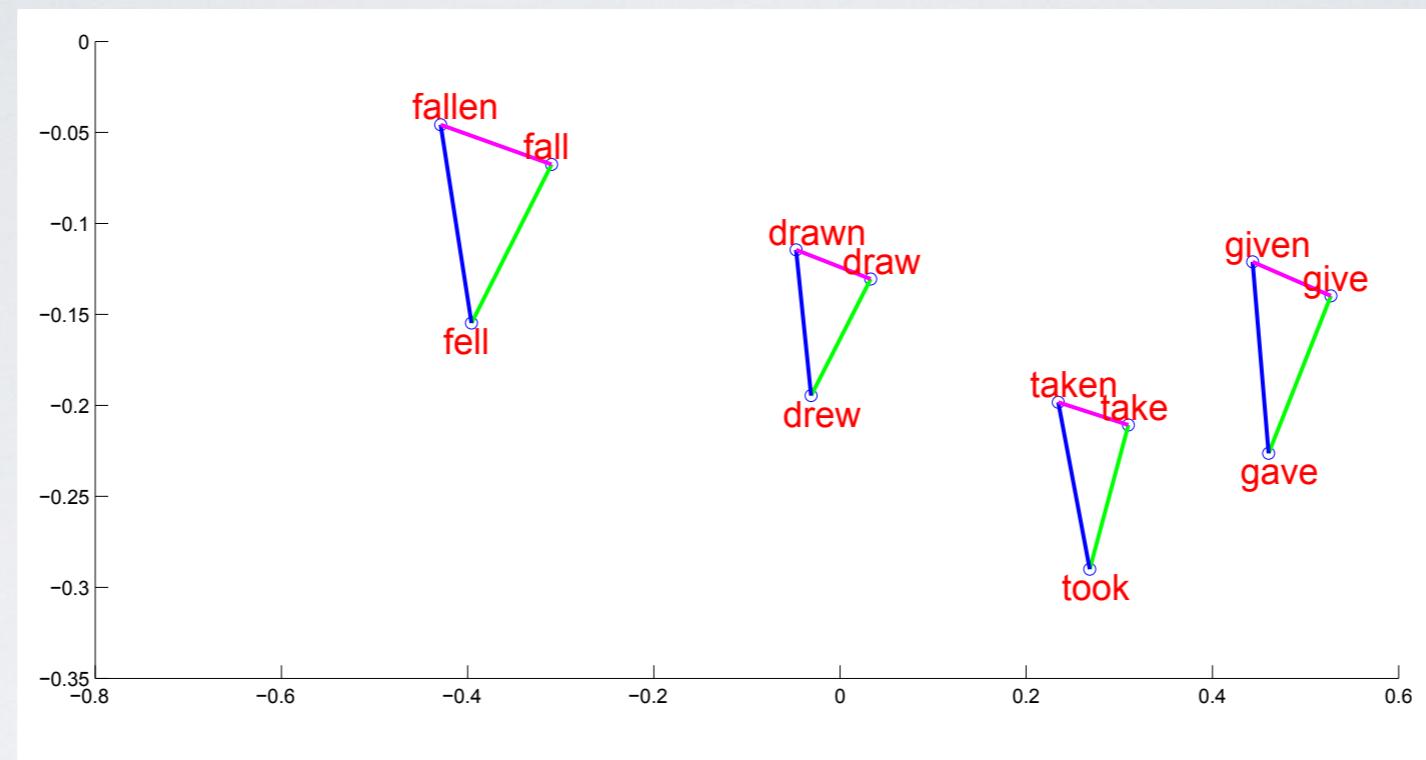
virginia
 columbia missouri
 indiana kansas
 maryland
 colorado tennessee
 washington oregon illinois carolina
 houston californiakansas
 philadelphia pennsylvania
 detroit toronto ontario massachusetts
 hollywood boston sydney
 montreal manchester
 london cambridge
 beijing quebec
 moscow mexico scotland
 wales england
 ireland britain
 australia sweden
 singapore spain
 america norway france
 europe austria
 asia germany
 africa russia
 india japan rome
 pakistan egypt
 israel vietnam



WORD CLUSTERS

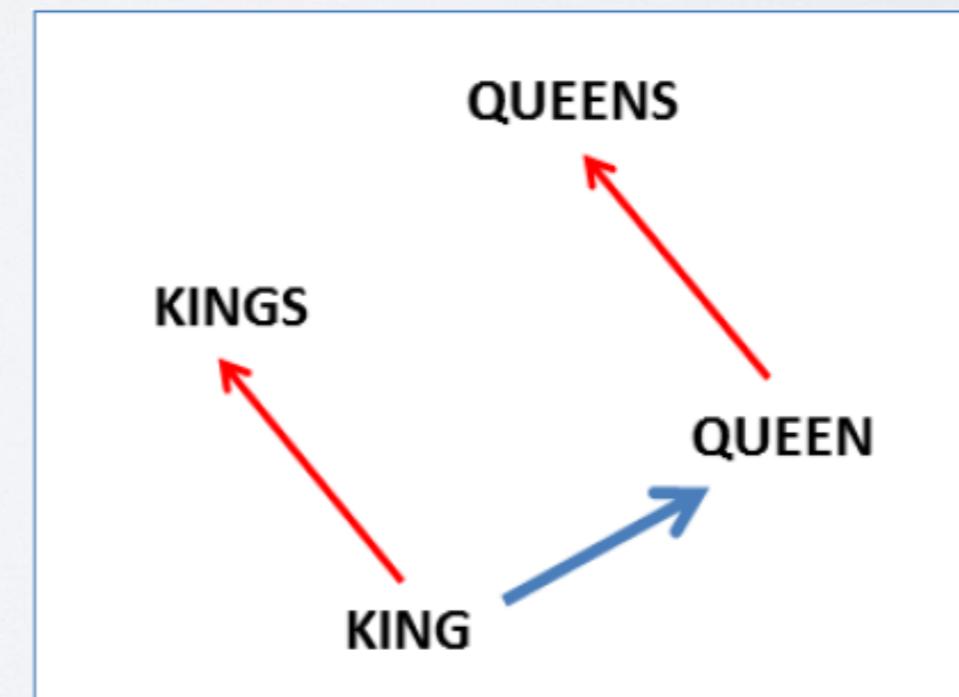
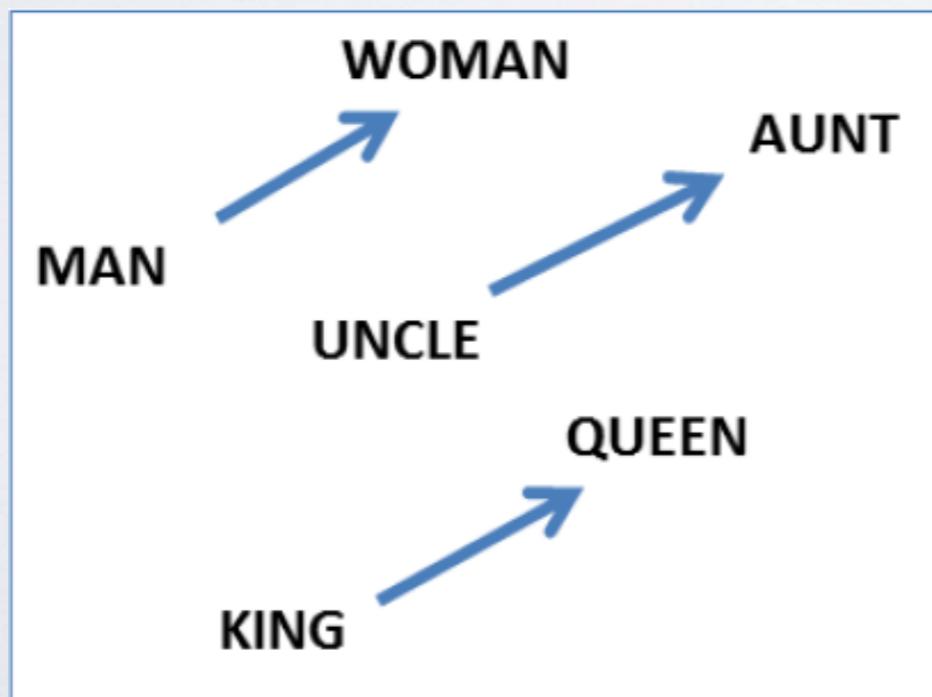
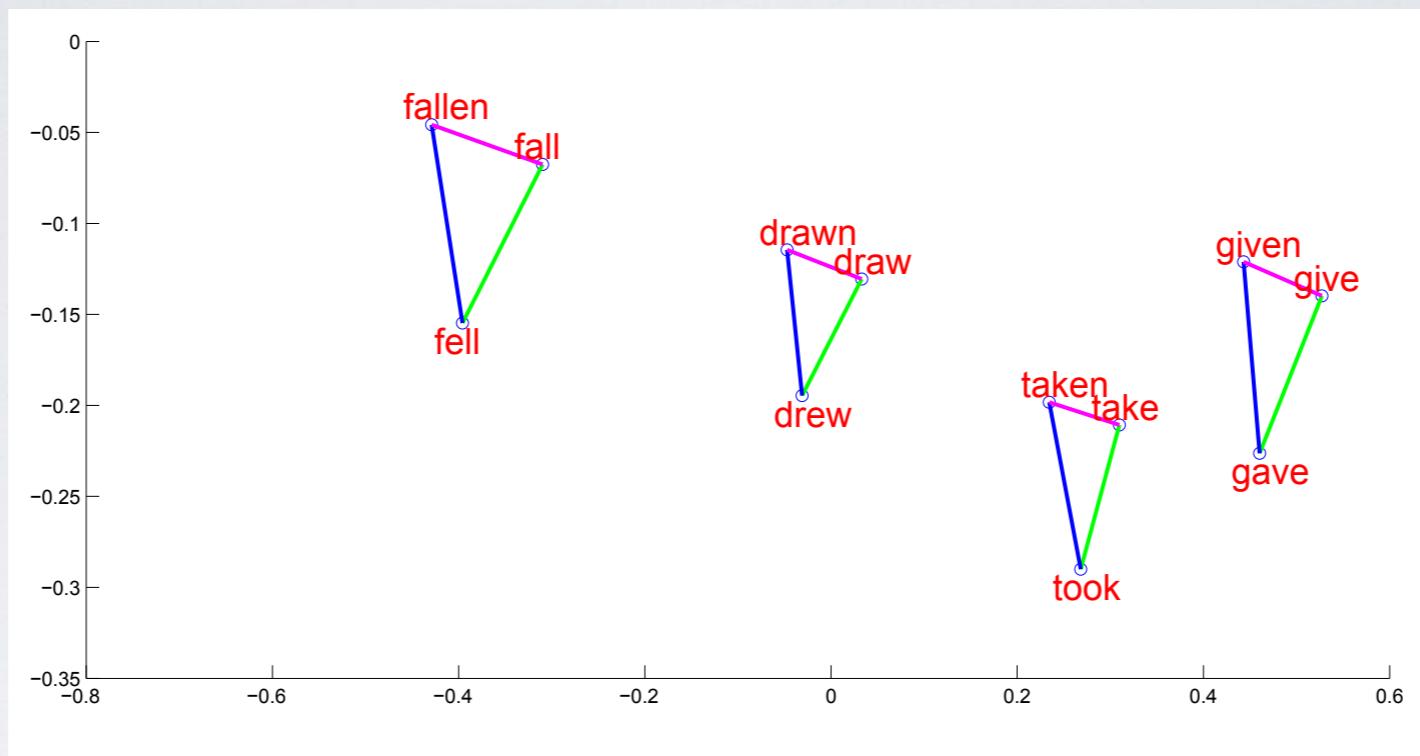


REGULARITY

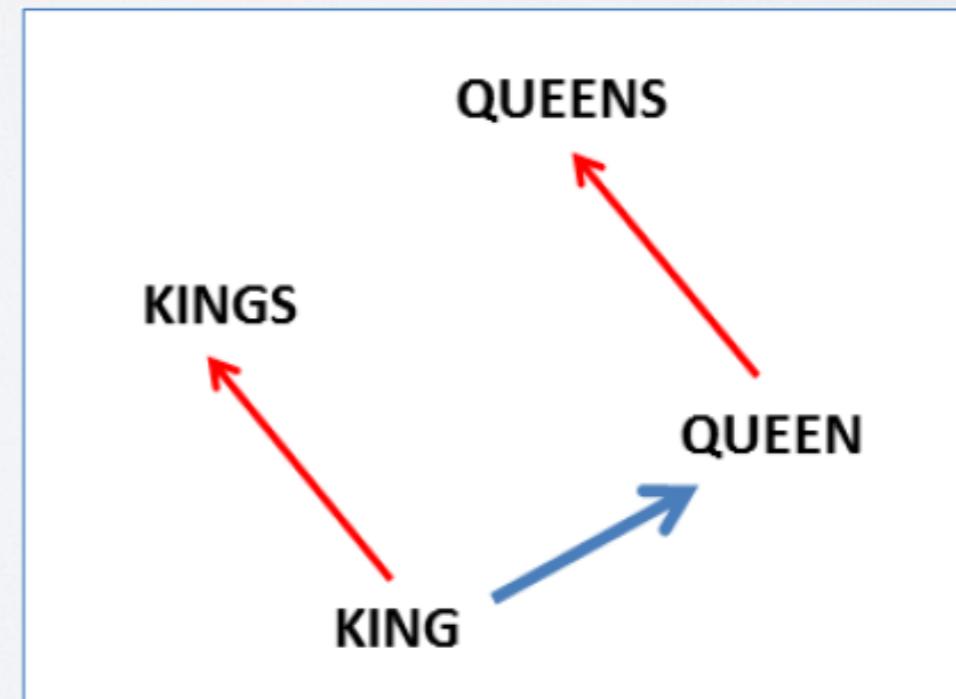


Mikolov et al.
Linguistic Regularities in Continuous
Space Word Representations
In: NAACL-HLT 2013

REGULARITY

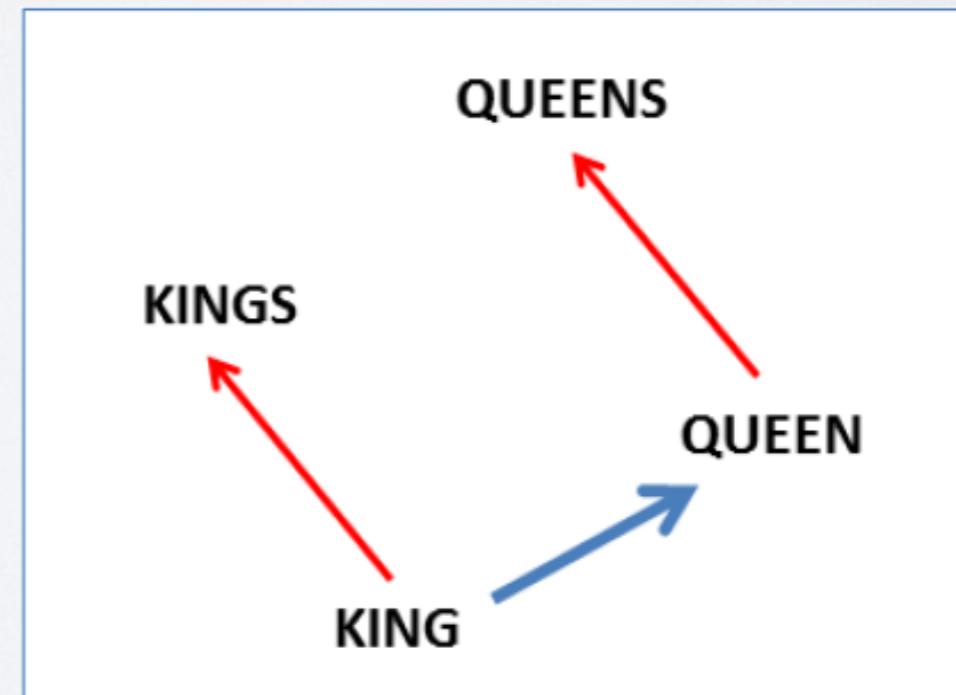


VECTOR ALGEBRA



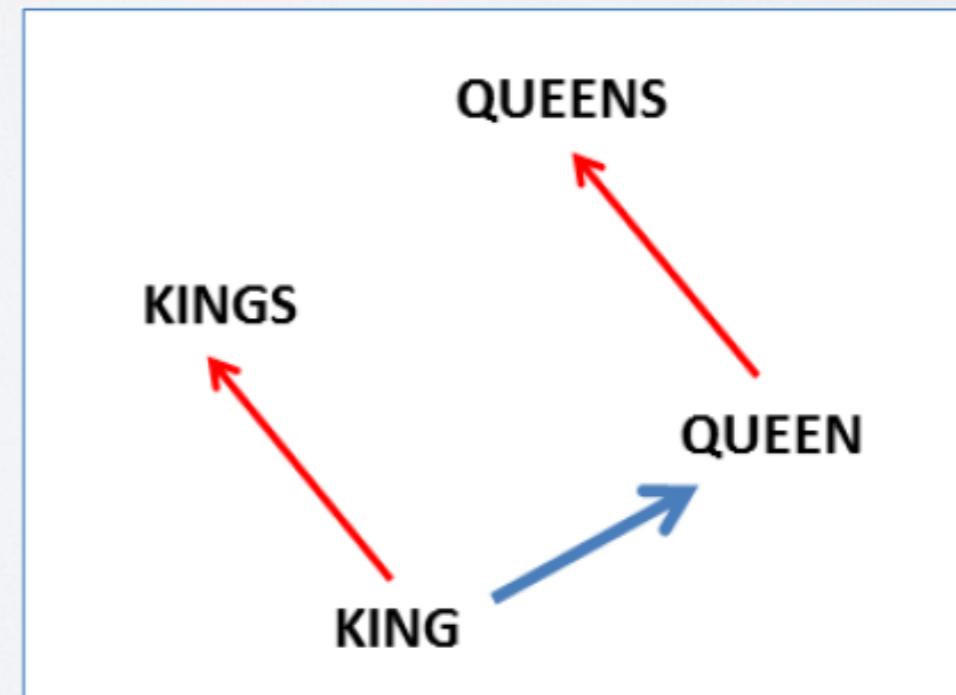
VECTOR ALGEBRA

bigger - big + cold =



VECTOR ALGEBRA

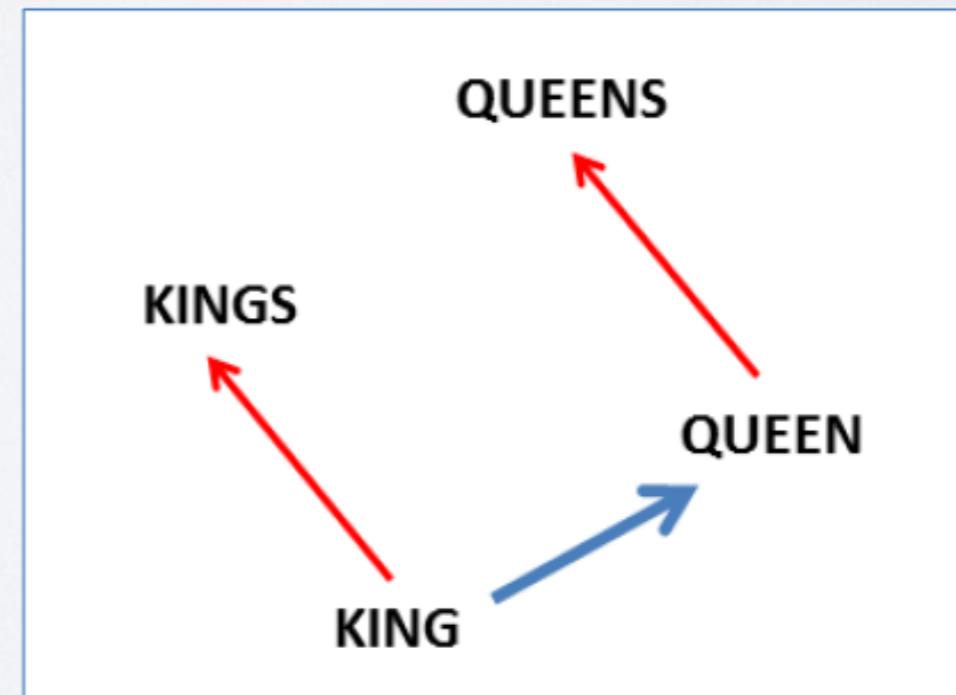
bigger - big + cold = colder



VECTOR ALGEBRA

bigger - big + cold = colder

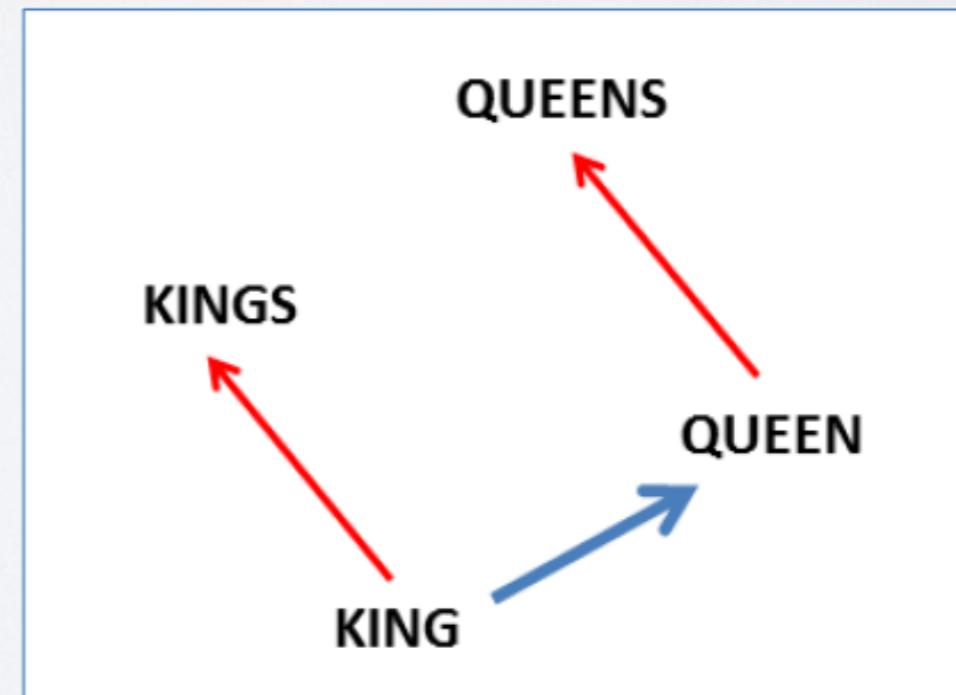
Paris - France + Italy =



VECTOR ALGEBRA

bigger - big + cold = colder

Paris - France + Italy = Rome

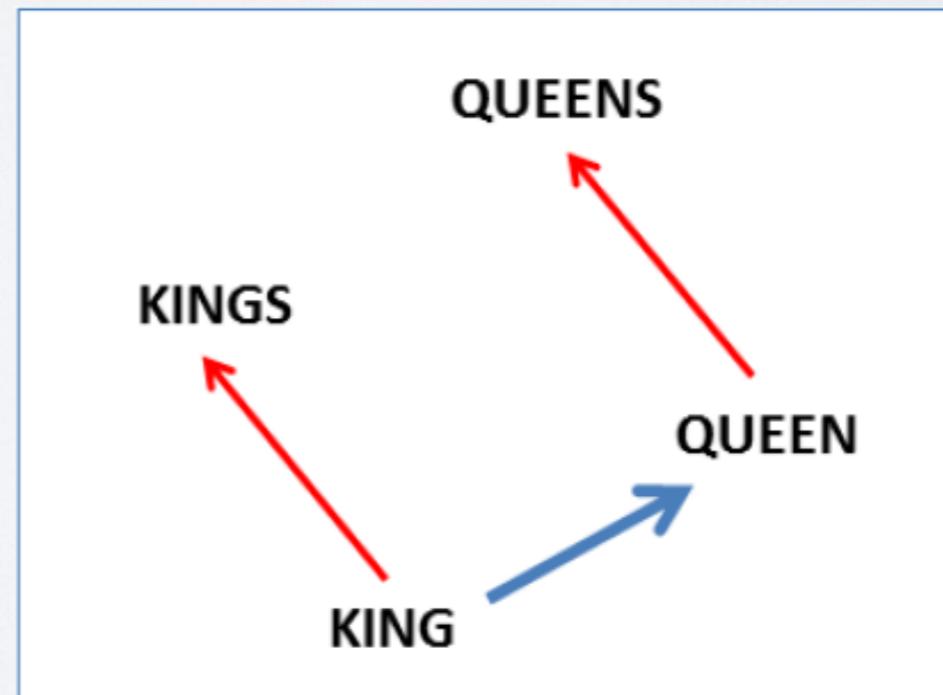


VECTOR ALGEBRA

bigger - big + cold = colder

Paris - France + Italy = Rome

Windows - Microsoft + Google =

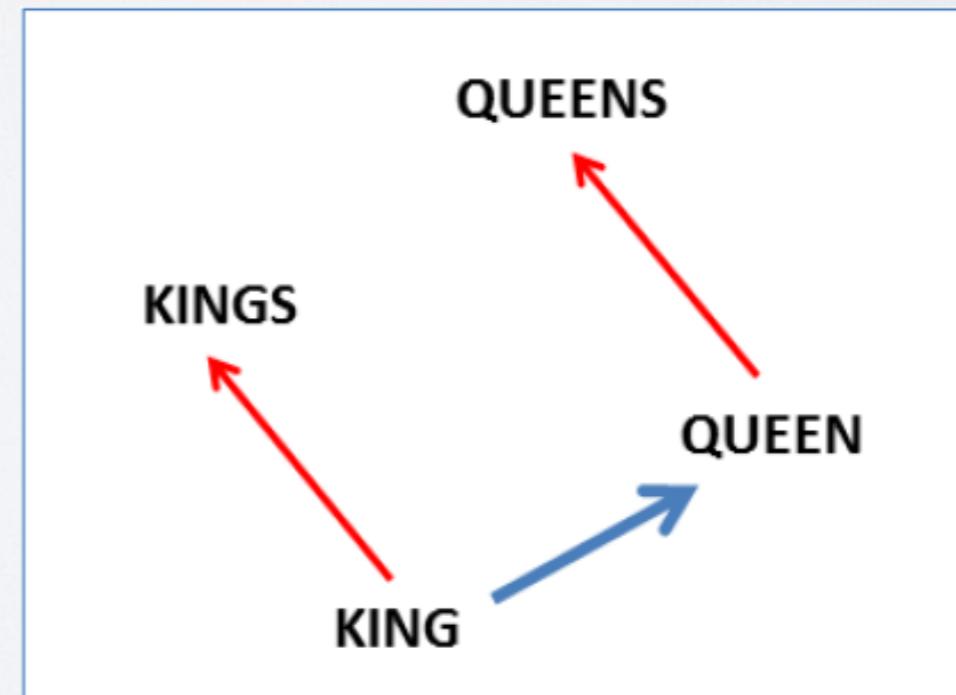


VECTOR ALGEBRA

bigger - big + cold = colder

Paris - France + Italy = Rome

Windows - Microsoft + Google = Android



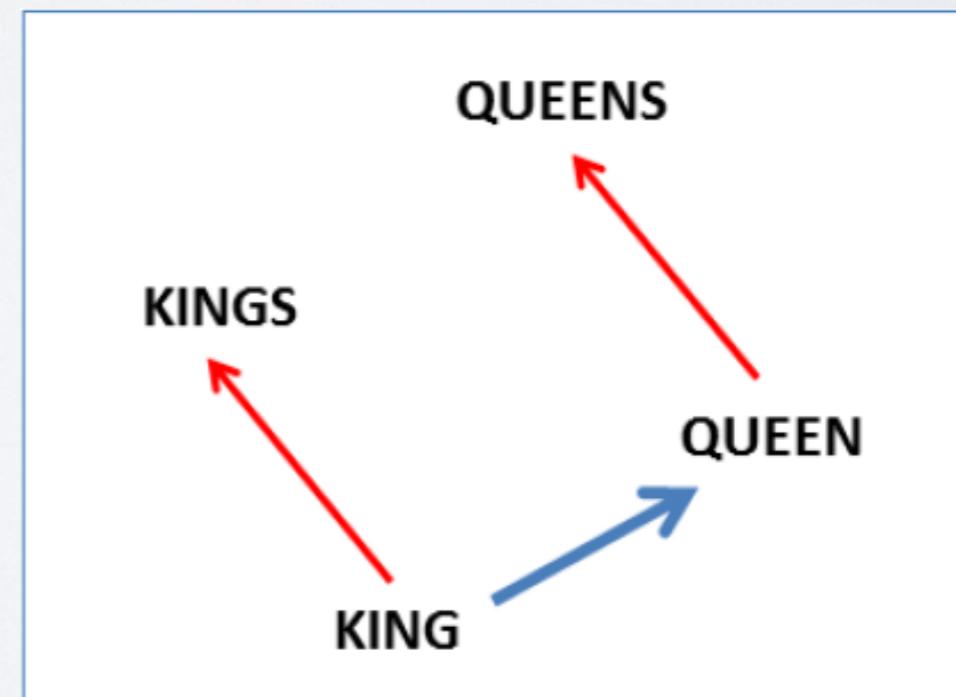
VECTOR ALGEBRA

bigger - big + cold = colder

Paris - France + Italy = Rome

Windows - Microsoft + Google = Android

sushi - Japan + Germany =



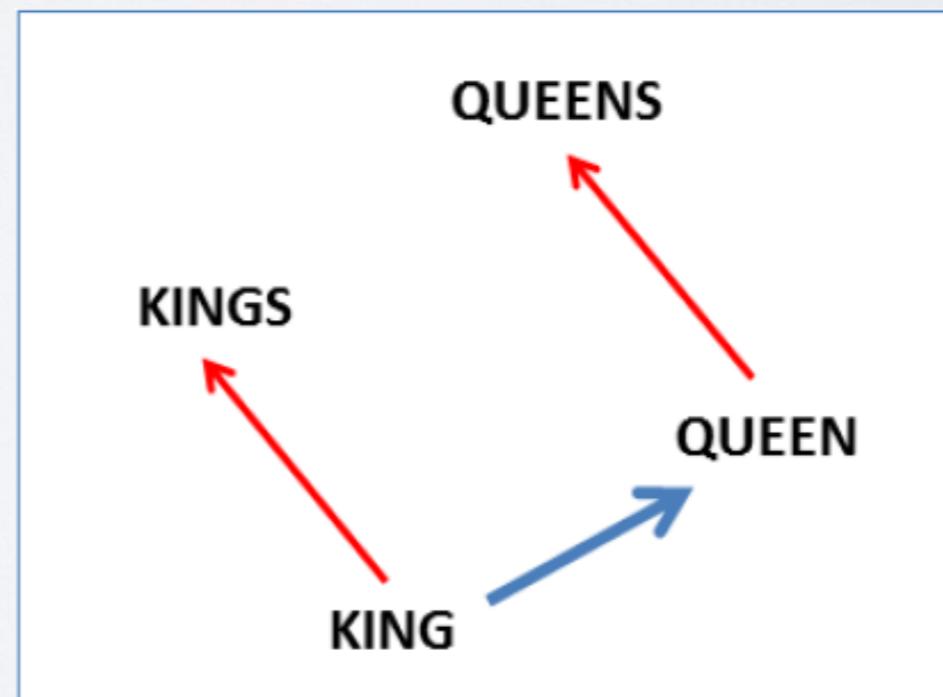
VECTOR ALGEBRA

bigger - big + cold = colder

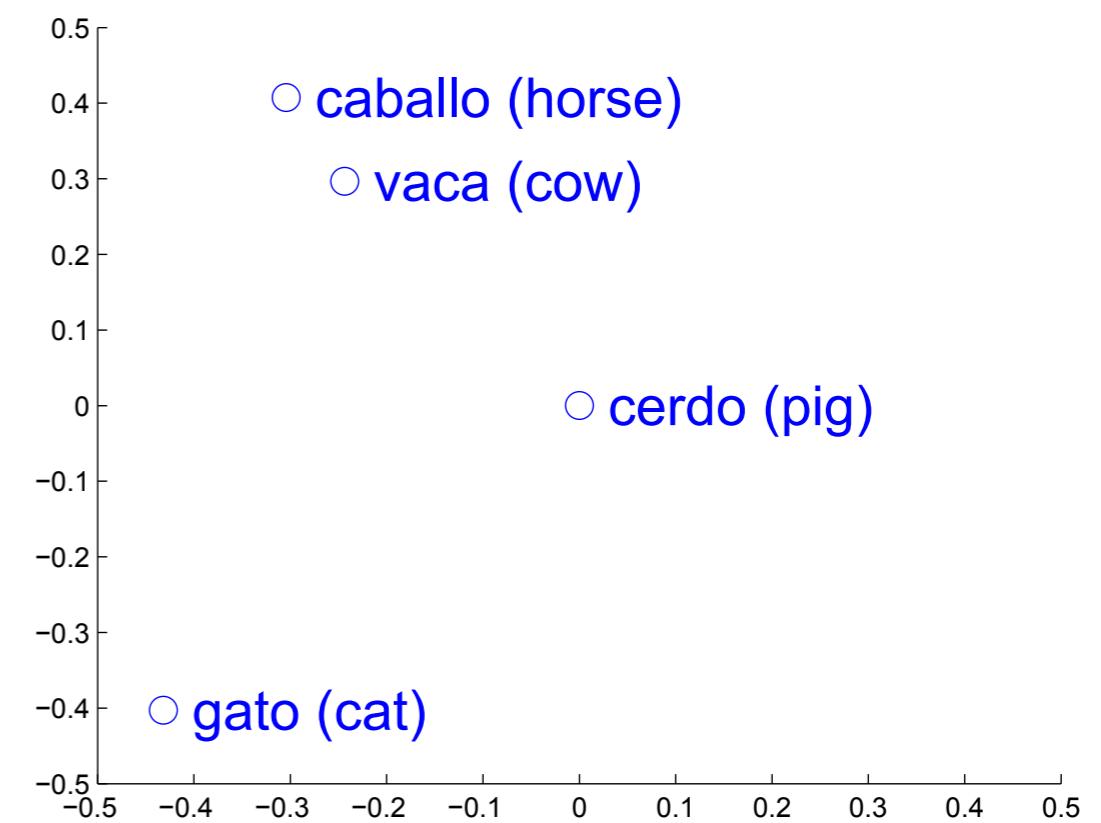
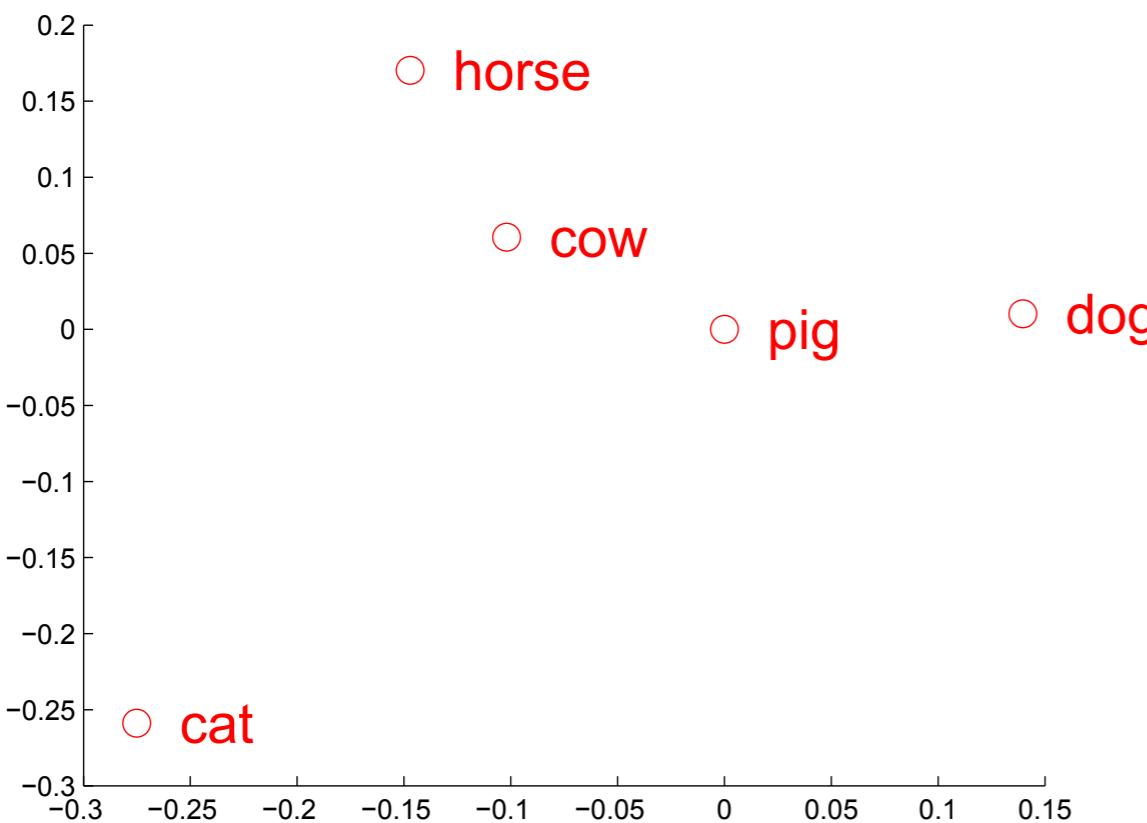
Paris - France + Italy = Rome

Windows - Microsoft + Google = Android

sushi - Japan + Germany = Bratwurst

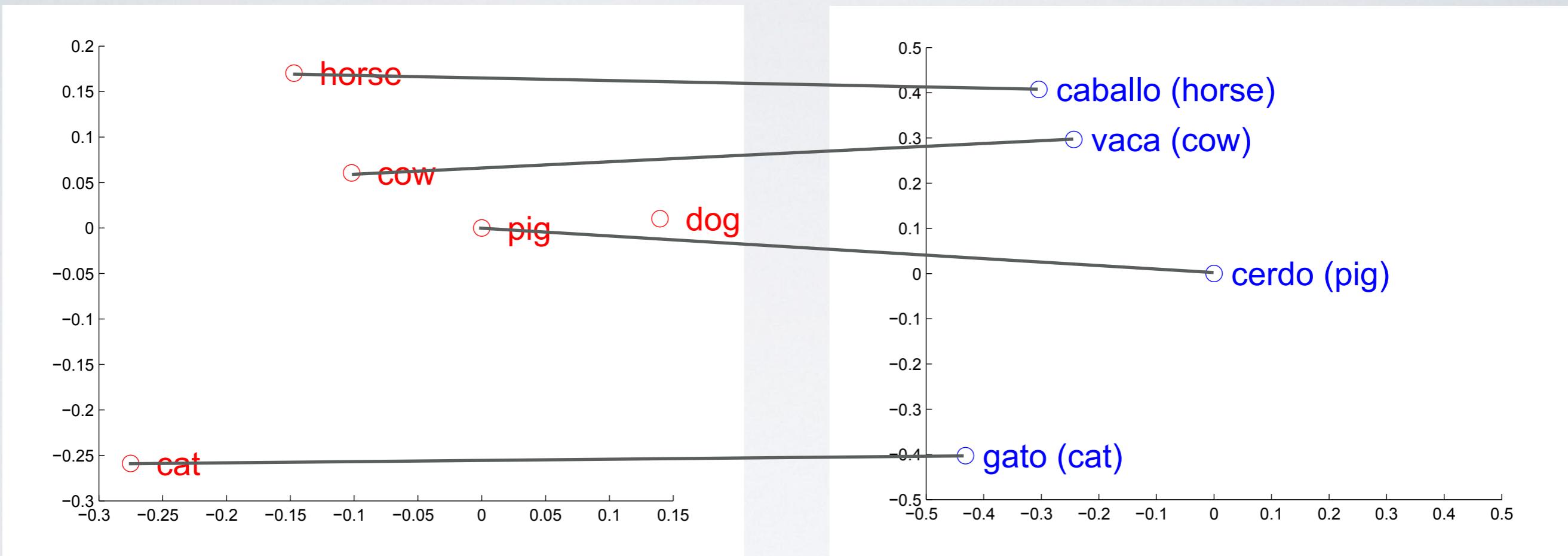


TRANSLATION



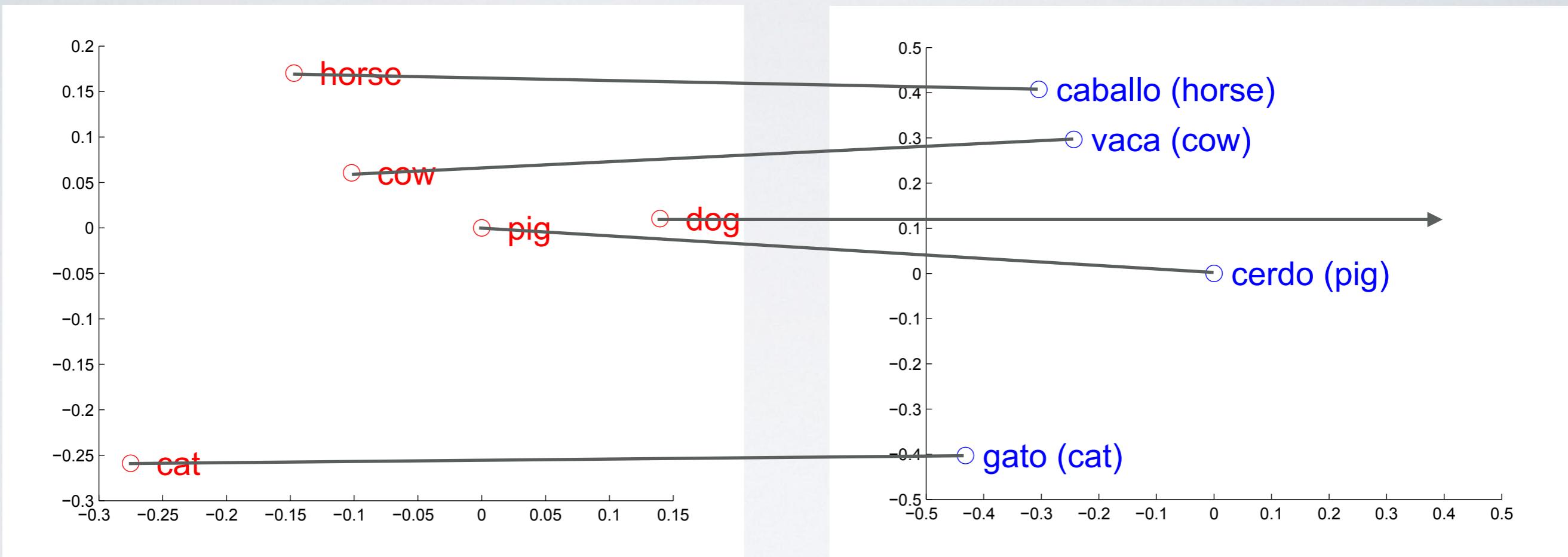
Mikolov et al.
Exploiting Similarities among Languages for Machine Translation
In: CoRR 2013

TRANSLATION



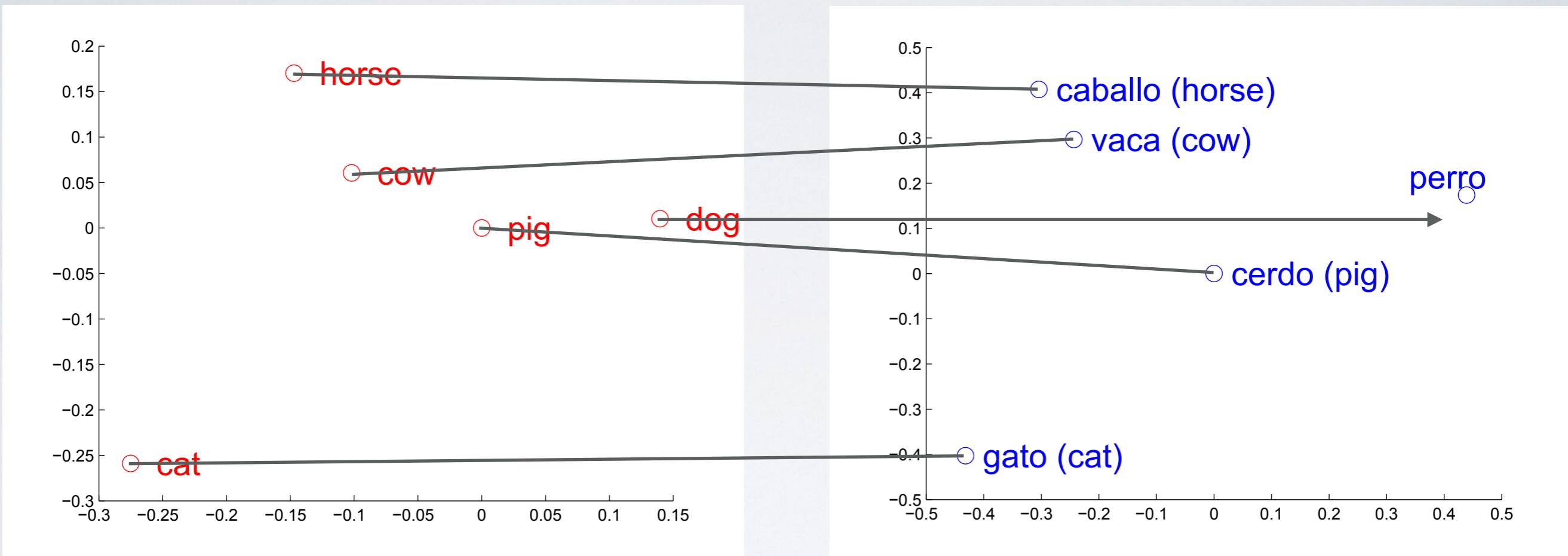
Mikolov et al.
Exploiting Similarities among Languages for Machine Translation
In: CoRR 2013

TRANSLATION



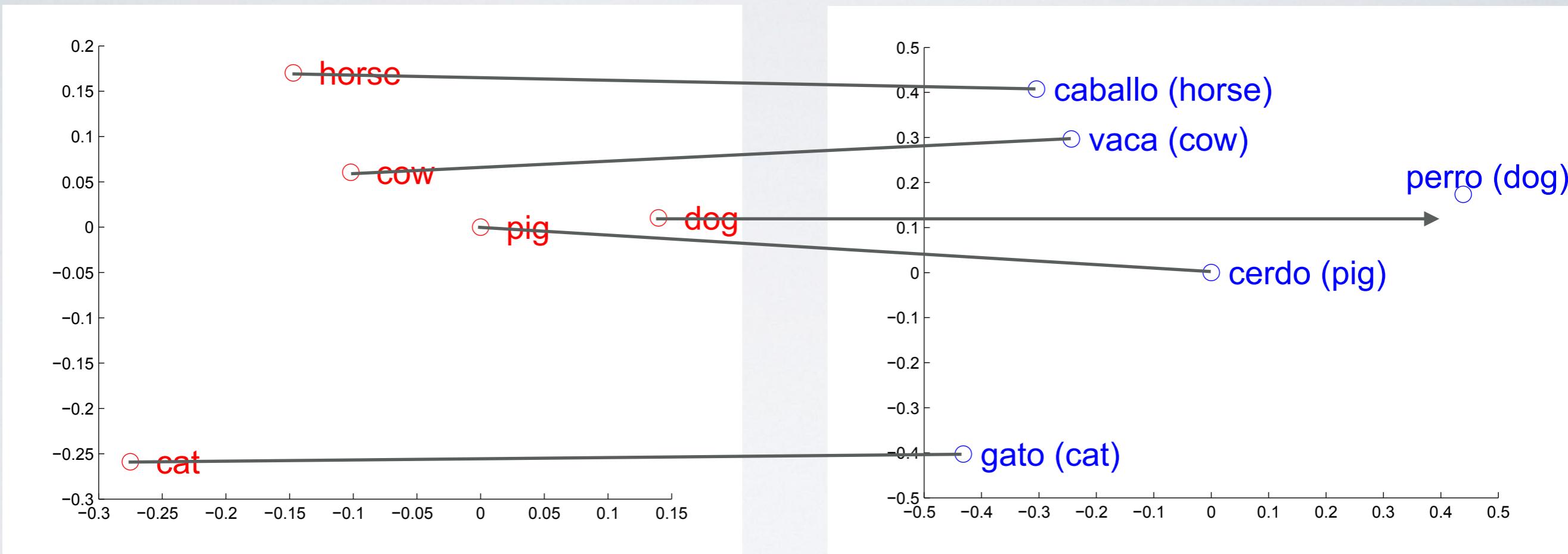
Mikolov et al.
Exploiting Similarities among Languages for Machine Translation
In: CoRR 2013

TRANSLATION



Mikolov et al.
Exploiting Similarities among Languages for Machine Translation
In: CoRR 2013

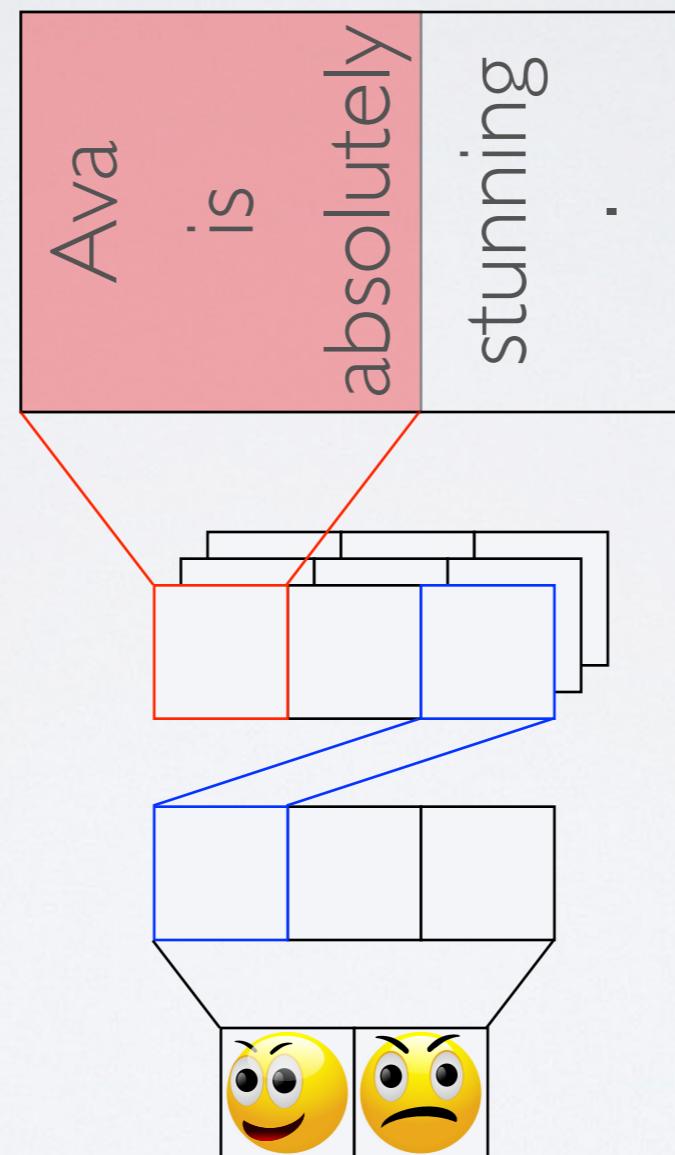
TRANSLATION



Mikolov et al.
Exploiting Similarities among Languages for Machine Translation
In: CoRR 2013

CNN FOR TEXT

"First of all
Ava is
absolutely
stunning"



Ebert,Vu, Schütze:
CIS-positive: Combining Convolutional Neural Networks
and SVMs for Sentiment Analysis in Twitter
In: SemEval 2015

Ebert et al.
hidden for blind review
In: EMNLP 2015 (submitted)

WHERE ARE WE?

ensemble system	64.84
CNN with add. training data	64.59
our CNN	64.46

Ebert et al.
hidden for blind review
In: EMNLP 2015 (submitted)

WHERE ARE WE?

- "cant wait to see the ipad hd [.]" (positive)
- "i have hspa tomorrow why the fuck am i awake ?" (negative)
- "<user> are you coming to the fair on sunday" (neutral)



ensemble system	64.84
CNN with add. training data	64.59
our CNN	64.46

Ebert et al.
hidden for blind review
In: EMNLP 2015 (submitted)

WHERE ARE WE?

- "cant wait to see the ipad hd [.]" (positive)
- "i have hspa tomorrow why the fuck am i awake ?" (negative)
- "<user> are you coming to the fair on sunday" (neutral)



- "the spurs may have won the battle , but not the war !" (positive, negative)
- "golf tomorrow [!] meet at the forum at 2pm [!]" (neutral, positive)
- "ren fair with owain tomorrow . fluff yes" (positive, neutral)



ensemble system	64.84
CNN with add. training data	64.59
our CNN	64.46

Ebert et al.
hidden for blind review
In: EMNLP 2015 (submitted)

CONCLUSION

CONCLUSION

NLP

- NLP is cool
- lots of unsolved problems
- deep learning in many areas
 - speech recognition
 - machine translation
 - word representation learning

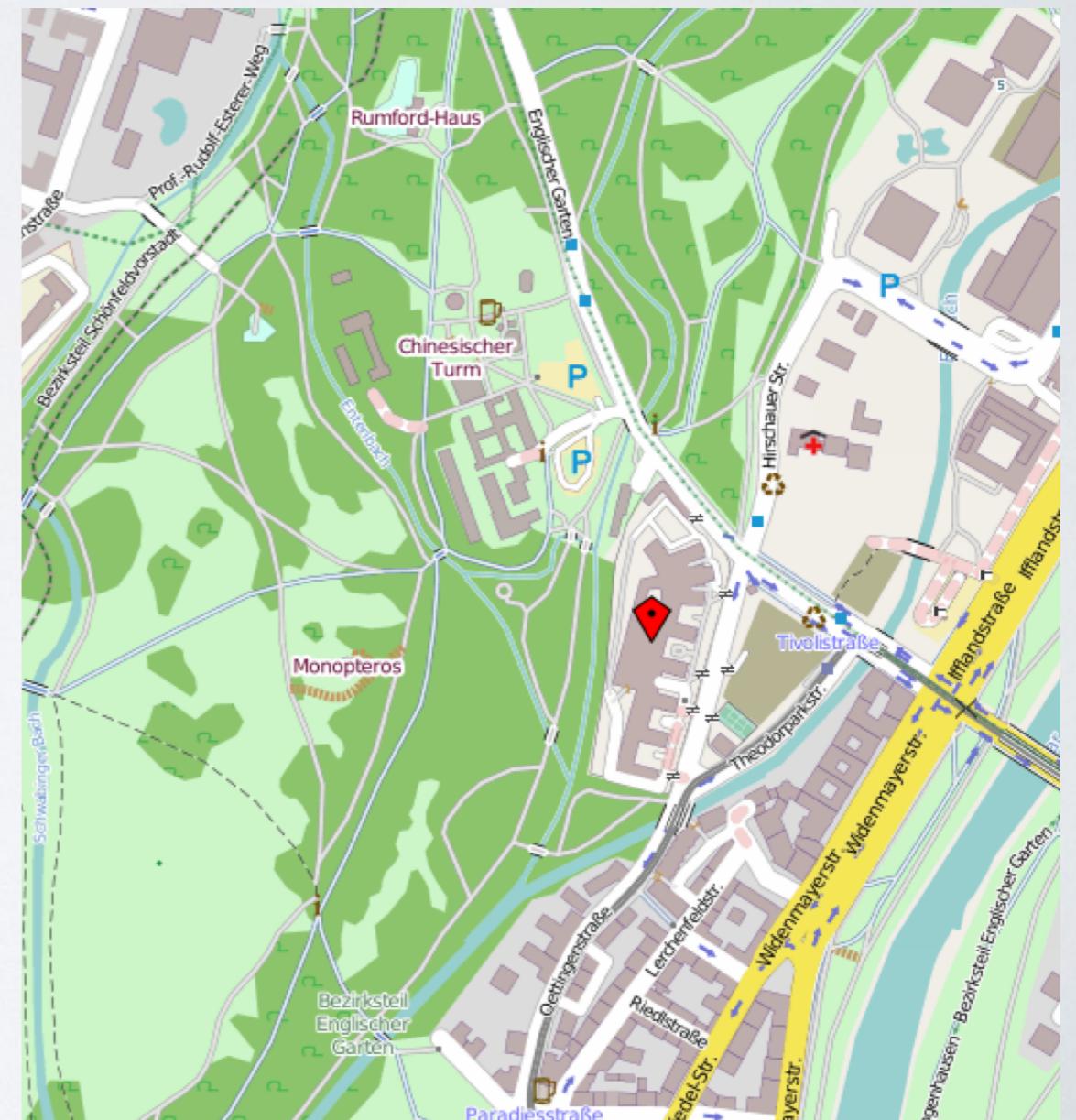
CONCLUSION

SENTIMENT ANALYSIS

- private, commercial applications
- challenging (e.g., "hard")
- CNN for polarity classification

NLP @ CIS

- word- / sentence representation learning
- language modeling
- speech recognition
- information extraction
- paraphrase classification
- knowledge-base systems
- machine translation
- sentiment analysis
- and more



Thanks for being such a

(great|awesome|patient|interested)

audience!

Sebastian Ebert

ebert@cis.lmu.de

<http://www.cis.uni-muenchen.de/ebert/>

IMAGE SOURCES

- Siri
 - www.apple.com/ios/siri
 - <http://mashable.com/2014/08/18/siri-fails/>
- Cortana
 - <https://www.microsoft.com/en-us/mobile/campaign-cortana/>
 - <http://i.guim.co.uk/static/w-700/h--/q-95/sys-images/Guardian/Pix/pictures/2014/4/14/1397491352713/52d8c136-20e3-4b38-9644-33ca23c93356-1020x612.jpeg>
- Watson: <http://techproessentials.com/wp-content/uploads/2015/02/img-video-jeopardy.jpg>
- TNSE: <https://lvdmaaten.github.io/tsne/>
- cross icon: android-ui-utils.googlecode.com
- checkmark icon: <https://commons.wikimedia.org/wiki/File:Check.svg>