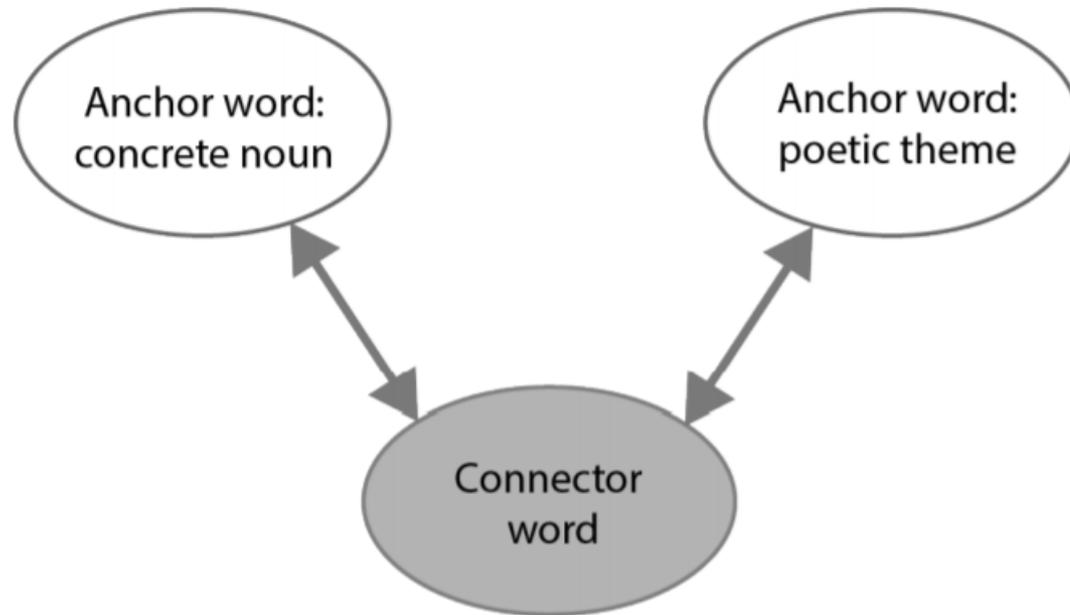# Intersecting word vectors to take figurative language into new heights

A paper by Andrea Gagliano & Emily Paul & Kyle Booten & Marti A. Hearst

So what's the plan?

- Motivation and purpose

- Previous work

- Some background

- The new approach

- Conclusions and future work

What are we trying to do?

Figure 1: Connector word drawing together the two semantic spaces of the anchor words.

Previous work

- A metaphor is a "mapping" between two distinct semantic spaces
  (Lakoff and Turner, 1989)

- Using adjectives as an indicator for metaphorical relationship between two terms
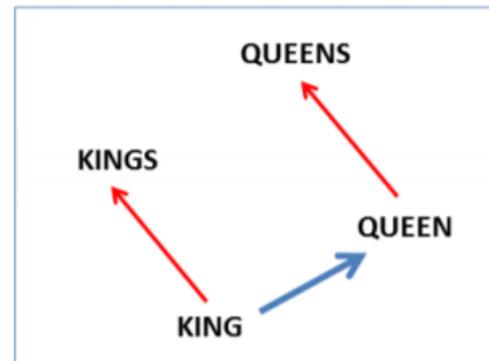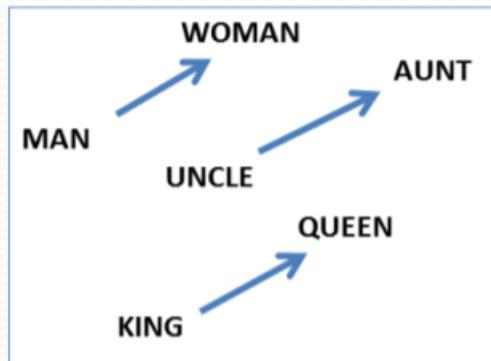 (Veale and Hao 2007, 2008)

Word2vec

"You shall know a word by the company it keeps"
  –John R. Firth

- How does it work with words?

- How does it work with phrases?

- Syntactic and semantic relationships

Word2vec

- Syntactic and semantic relationships

Word2vec

- How do we measure the similarity of two words?

  We are using the *cosine similarity* formula (on the board)

How will we measure our results?

- Quantitative

- Qualitative

Let's get started

In the beginning…

Concrete nouns		
bed	ear	finger
horse	sand	hair
bell	grass	rock
book	rose	breast
ship	blood	window
wing	girl	snow
wood	ring	body
room	wine	ground
mouth	garden	stone
storm	brain	flame
town	wave	shadow
silver	mist	line
stream	dawn	path
dust	breath	king
color	spring	darkness
side	nation	race
state		

Table 2: Pool of concrete nouns used in the selection of anchor pairs.

Poetic themes		
loss	melancholy	anger
animals	calmness	compassion
confusion	death	envy
faith	fear	forgiveness
freedom	friendship	god
grace	gratitude	grief
hate	hope	immortality
jealousy	joy	life
mothers	nature	peace
people	religion	remembrance
love	sadness	silence
smiling	songs	spirituality
spring	suffering	truth
unity	vanity	war
water	wind	bitterness
consciousness	happiness	earth
soul	surrender	violence

Table 3: Pool of poetic themes used in the selection of anchor pairs.

Selecting anchor words

Selecting anchor words

Concrete nouns		
bed	ear	finger
horse	sand	hair
bell	grass	rock
book	rose	breast
ship	blood	window
wing	girl	snow
wood	ring	body
room	wine	ground
mouth	garden	storm
storm	brain	flame
town	wave	shadow
silver	mist	line
stream	dawn	path
dust	breath	king
color	spring	darkness
side	nation	race
state		

Table 2: Pool of concrete nouns used in the selection of anchor pairs.

Poetic themes		
loss	melancholy	anger
animals	calmness	compassion
confusion	death	envy
faith	fear	forgiveness
freedom	friendship	god
grace	gratitude	grief
hate	hope	immortality
jealousy	joy	life
mothers	nature	peace
people	religion	remembrance
love	sadness	silence
smiling	songs	spirituality
spring	suffering	truth
unity	vanity	war
water	wind	bitterness
consciousness	happiness	earth
soul	surrender	violence

Table 3: Pool of poetic themes used in the selection of anchor pairs.

Selecting anchor words

- Setting a threshold: *similarity* < 0.4

Addition model

Addition model

- We create a new vector, a:
  $$vec(a) = vec(c) + vec(t)$$

Addition model

- We create a new vector, a:
  $$vec(a) = vec(c) + vec(t)$$

- Find a set A containing n words closest to $vec(a)$

Addition model

Top 10 words from word2vec addition for storm + surrendering
surrendered
hurricane
storms
snowstorm
rainstorm
tornado
blizzard
typhoon
twister
squall

Table 4: Top 10 words retrieved when adding anchor words *storm* and *surrendering* using word2vec addition.

Addition model

Top 10 words from word2vec addition for storm + surrendering
surrendered
hurricane
storms
snowstorm
rainstorm
tornado
blizzard
typhoon
twister
squall

Table 4: Top 10 words retrieved when adding anchor words *storm* and *surrendering* using word2vec addition.

Not good enough...

Intersection model

- Find two sets:

  -C containing n words closest to vec(c)
  -T containing n words closest to vec(t)

- Look at the intersection set I = C∩T

Intersection model

- Find two sets:

  -C containing n words closest to vec(c)
  -T containing n words closest to vec(t)

- Look at the intersection set I = C∩T

*choosing n = 1000

Let's separate the wheat from the chaff

- Observation: $A \backslash I \neq \varnothing \neq I \backslash A$

- Create two new sets as follows:
  -Unique to intersection: $U_I = I \backslash A$
  -Unique to addition: $U_A = A \backslash I$

Unique to Intersection $U_I = I \setminus A$	Unique to Addition $U_A = A \setminus I$
onslaught	squall
stranding	tornado
blowing	typhoon
dissipating	snowstorm
battering	flooding
game	rainstorm
breastworks	deluge
regrouped	downpour
batter	blizzard
dissipated	ike
outburst	twister
pounding	hurricane
submerging	rain
pounded	
barrage	
regrouping	
stalemate	

Table 5: Connector words for *storm* and *surrendering* retrieved from the words unique to I and the words unique to A.

Quantitative observations

Quantitative observations

Unique to Addition $U_I = I \setminus A$	Similarity to noun *storm*	Similarity to theme *surrendering*
onslaught	0.30	0.20
stranding	0.27	0.28
blowing	0.24	0.29
dissipating	0.23	0.22
battering	0.29	0.24
game	0.19	0.25
breastworks	0.19	0.20
regrouped	0.19	0.31
batter	0.22	0.25
dissipated	0.24	0.21
outburst	0.21	0.20
pounding	0.20	0.26
submerging	0.26	0.23
pounded	0.24	0.32
barrage	0.25	0.20
regrouping	0.19	0.31
stalemate	0.19	0.21
Average spread between similarity scores: 0.05		

Table 6: Similarity scores between connector words found in U_I to anchor words *storm* and *surrendering*. The average spread between the scores of 0.05 indicates the small band of similarity the words exist in, showing the balanced similarity the connector word has with each of the anchor words.

Quantitative observations

Unique to Intersection $U_A = I \setminus A$	Similarity to noun *storm*	Similarity to theme *surrendering*
squall	0.63	-0.03
tornado	0.64	-0.02
typhoon	0.62	-0.01
snowstorm	0.64	0.01
flooding	0.57	0.01
rainstorm	0.57	0.07
deluge	0.50	0.08
downpour	0.52	0.08
blizzard	0.61	0.00
ike	0.58	0.02
twister	0.62	-0.01
hurricane	0.73	0.04
rain	0.46	0.10
Average spread between similarity scores: 0.56		

Table 7: Similarity scores between connector words found in U_A to anchor words *storm* and *surrendering*. The average spread between the scores of 0.56 shows the wide range of similarity scores.

Quantitative observations

Anchor word pairs	Range of avg. sim. from words in U_I to anchor words	Range of avg. sim. from words in U_A to anchor words
flame & caring	0.22 – 0.30	0.13 – 0.58
color & earthly	0.28 – 0.32	0.17 – 0.55
hair & anguish	0.27 – 0.33	0.14 – 0.66
flame & killing	0.23 – 0.26	0.09 – 0.55
mouth & comp.	0.25 – 0.29	0.16 – 0.54
storm & surr.	0.21 – 0.26	0.03 – 0.69
ring & mankind	0.21 – 0.34	0.11 – 0.57
hair & envied	0.27 – 0.31	0.17 – 0.58
book & liberties	0.23 – 0.29	0.15 – 0.54
town & grieving	0.24 – 0.28	0.14 – 0.54

Table 8: The low end of the ranges is the average of the minimum similarity scores across all the connector words to each of the words in the anchor word pair. The upper end of the ranges is the average of the maximums. A smaller range means that the anchor words have more balanced similarity to the connec-

Qualitative observations

- Constructing a dataset of sentences using crowd-sourced workers (Mechanical Turk)

- The task:
  complete a template sentence of the form:
  "[connector word] connects [concrete noun] and [poetic theme] because…"

- "*Barrage* connects *storm* and *surrendering* because…"

An example

Unique to Intersection	Unique to Addition
affection	compassion
friendship	**torch**
spirit	selfless
passion	considerate
soul	kindness
brotherhood	compassionate
aloneness	loving
love	**devotion**
cook	
undying	

Table 11: Figurative ties between *flame* and *caring*. Bolded words were selected by Mechanical Turk workers as the best word to create the figurative tie.

An example

- "**Cook** connects <u>caring</u> and <u>flame</u> because it is related to flame as flames are used in cooking and cooking can be a symbol of caring for someone with good food."

- "**Torch** connects <u>caring</u> and <u>flame</u> because when someone cares about someone else it's often said they are carrying a torch for them, while the visual of a torch itself tends to have a flame atop it."

Qualitative observations

- Our goal: blend the distinct semantic spaces of the two anchor words to create figurative relationships.

- Synonym-based relationships

- Relationships blending distinct semantic spaces

Qualitative observations

- Synonym-based relationships

- "**Torch** connects <u>caring</u> and <u>flame</u> because when someone cares about someone else it's often said they are carrying a torch for them, while the visual of a torch itself tends to have a flame atop it."

- Similarity score **torch**-<u>caring</u>: 0.06
  Similarity score **torch**-<u>flame</u>: 0.67

Qualitative observations

- Relationships blending distinct semantic spaces

- "**Cook** connects <u>caring</u> and <u>flame</u> because it is related to flame as flames are used in cooking and cooking can be a symbol of caring for someone with good food."

- Similarity score **cook**-<u>caring</u>: 0.26
  Similarity score **cook**-<u>flame</u>: 0.22

Discussing the results

- Unbalanced cosine similarity- leads to a synonymous relationships.

- Balanced cosine similarity- blends the two distinct semantic spaces of the anchor word and creates a new shared semantic space.

Discussing the results

- Unbalanced cosine similarity- leads to a synonymous relationships.
  *mostly the words from the set U_A

- Balanced cosine similarity- blends the two distinct semantic spaces of the anchor word and creates a new shared semantic space.
  *mostly the words from the set U_I

Future work

- Test this hypothesis directly

- Checking other bands of similarity outside 0.25-0.30

- Poetry generation

- Tools to assist creative writing

Thank you!