

NLP Course - Assignment 3

Otmazgin, Shon Rassin, Royi
shon711@gmail.com isroei5700@gmail.com

December 10th, 2020

Report

1. (a) We employed the following filters:
 - Word frequency: 100 (mandatory)
 - Word frequency as attribute: 75 (mandatory)
 - #Attributes for each word: 100, 150, 200, 250, 300
 - a word is considered a content word if its tag is a noun, verb, adjective or an adverb and is not in our predefined stop-words list and punctuation list (can be found in *utils.py*).

(b)	Attributes per word	Window	Sentence	Dependency
	100	7944	7944	7943
	150	5690	5690	5659
	200	4917	4917	4917
	250	4222	4222	4222
	300	3697	3697	3697

Table 1: Words for similarity and #attributes for each word

(c)	Attributes per word	Window	Sentence	Dependency
	100	9529	9085	114420
	150	9451	8476	92243
	200	9257	7930	78314
	250	9057	7415	68431
	300	8806	6967	60643

Table 2: Attributes in computation and #attributes for each word

The quality of our similarities improved when we increased the threshold. For instance, we had the best results when we used 250-contexts per word (we shall expand on this later, as the results were evaluated based on AP and MAP). The results peaked for 250-contexts per word, and 300-contexts per word seems to be too many contexts for all methods.

2. From our examination: The sentence (middle) column indicates a similarity of words within a broad context, the whole sentence. For this column,

the most similar word to ‘gun’ is ‘cannon’, we can imagine sentences that will use them in similar context, albeit, probably not interchangeably: cannon blasts were fired at 9 PM / gun shots were heard at 9 PM. Here, the emphasis is on the overall context of the sentence.

The window (left) column indicates a similarity of words within a narrow context. In our case, a window of two. Naturally, many of the similarities are shared with the sentence-column, however, the emphasis is on the word’s window, and similar words are defined to be those that share similar contexts within a window and are not ranked by the context of the whole sentence, unlike the former definition. So here, ‘fire’ is ranked above ‘cannon’, because ‘gun’ and ‘fire’ share more similar windows than ‘gun’ and ‘cannon’.

The dependency (right) column seems to focus on words that share the same (general) function. For instance, the top words for ‘bus’ (excluding rail that is found in the top two of all three methods) are a form of public transportation, just like a bus: train, tram, ferry, taxi, and transit. Similarly, for “hotel”, we find: restaurant, resort, store, palace, castle, shop, theater, and casino. All of these places underscore service and customer-satisfaction (in a palace and a castle royalty is being serviced too).

NOTE: we’ve only attached the 100-features-smoothed version of our words to avoid ballooning the file with the 250-features-smoothed version too. If you find it necessary, please let us know and we shall send that too.

NOTE: top20 2nd order similarity in appendix 1

3. Words that top the context-list are those that are used in the vicinity of the target word. So if the target word is ‘gun’, and ‘anti-aircraft’ tops its sentence-context column, it means it is the most frequent word out of all the words that appear in the sentences with ‘gun’. That is not to say that it bears similar meaning to ‘gun’, (in contrast to the former type of similarity), but that it is used within the same context. This can be interpreted in similar fashion to the window-column while the dependency column indicates the effect the target-word has on another word in the sentence. For instance, for ‘gun’ as a target word, the feature (‘machine’, compmod, out) indicates that ‘machine’ is modified by ‘gun’. What type of machine? A machine-gun.

NOTE: we’ve only attached the 100-features-smoothed version of our words to avoid ballooning the file with the 250-features-smoothed version too. If you find it necessary, please let us know and we shall send that too.

NOTE: top20 1st order similarity in appendix 1

Table 3: Car AP with 100 attributes

4.	Judgement/method	Window	Sentence	Dependency
	topic	0.61	0.66	0.47
	semantic	0.41	0.38	0.36

Table 4: Car AP with 250 attributes

Judgement/method	Window	Sentence	Dependency
topic	0.68	0.64	0.61
semantic	0.59	0.33	0.61

Table 5: Piano AP with 100 attributes

Judgement/method	Window	Sentence	Dependency
topic	0.66	0.67	0.61
semantic	0.69	0.68	0.94

Table 6: Piano AP with 250 attributes

Judgement/method	Window	Sentence	Dependency
topic	0.66	0.75	0.61
semantic	0.58	0.60	0.91

Table 7: MAP with 100 attributes

Judgement/method	Window	Sentence	Dependency
topic	0.64	0.67	0.53
semantic	0.55	0.53	0.65

Table 8: MAP with 250 attributes

Judgement/method	Window	Sentence	Dependency
topic	0.67	0.70	0.61
semantic	0.59	0.46	0.76

In general, getting a high score on topic is easier than on semantic. A possible explanation for this is that ‘topic’ is a broader concept, whereas categorizing two words to be semantically related is a more specific relation. Also, in almost all cases ‘piano’ scored higher than ‘car’ on all metrics, we believe that it is partially due to the fact that there are much less appearances of ‘piano’ vs ‘car’. Nevertheless, since the manual judgements are based on a human, and are not clear cut (one person may find that a motorcycle is topically related to a car and someone else will not), it is entirely possible ‘piano’ did better because we lack understanding on it, and found many words to be somewhat related, however, a pianist would have not categorized it this way. This raises the question, what kind of system are we (hypothetically) building? One that returns results on ‘piano’ that a pianist will agree with, or a system for the average person? Ultimately, using many judges on the same list and scoring the system based on their average would probably indicate best on the system’s performance.

NOTE: Car and Piano annotations in appendix 3

5. (a) PMI Estimation: $\log \frac{p(w,att)}{p(w)p(att)} \approx \log \frac{\frac{\#(w,att)}{\#(*,*)}}{\frac{\#(w)}{\#(*,*)} \frac{\#(att)}{\#(*,*)}} = \log \frac{\#(w,att)\#(*,*)}{\#(w)\#(att)}$
- i. $\#(w,att)$ is the number of times w counted (occurred) with att .
 - ii. $\#(w)$ is the number of times w counted (occurred) with any att .
 - iii. $\#(att)$ is the number of times att counted (occurred) with any w .
 - iv. $\#(*,*)$ is the number of times any w counted (occurred) with any att .
 - v. We implemented the positive version of PMI called PPMI.
 $PPMI(w,att) = \max(0, PMI(w,att))$.
 - vi. Since for normal-PPMI we get an inflated value for low freq(x,y), the smoothed version of PPMI was implemented: $\log \frac{\#(w,att)\#(*,*)}{\#(w)^{\frac{3}{4}}\#(att)}$

The probabilities are computed based on the frequencies we collected from iterating over the data and are stored in the 'stats' dictionary (as advised in the assignment). Additionally, the dictionary distinguishes among 'window', 'sentence', and 'dependency' with a 'method' key, that ultimately leads to a co-occurrence-specific dictionary.

Relevant code: *WordsStats.fit*

The difference between the smoothed and the normal-PPMI, was somewhat negligible, with only the words in the bottom of the top-20 list swapped places, or some were replaced with new words. Since most of the changes, if any, occurred in the bottom of the list, they were somewhat insignificant. Regardless, we opted to use the smoothed version.

Relevant code: *WordSimilarities.get_PPMI*

- (b) Given word w and attribute att : first, we calculated $ppmi = PPMI(w, att)$. Then, we saved the result in two different hash tables (Python dictionary): $word_vecs[w][att] = ppmi$ and $att_vecs[att][w] = ppmi$. In addition we sum the $ppmi$ squared, to later compute w 's $l2$ norm. Relevant code: *WordSimilarities.fit*

Finally, we were able to make use of the efficient dot product pseudo-code seen in lecture 7, slide 52.

- i. We saved results only for $ppmi > 0$
- ii. We iterate only for w and att that occurred together in corpus after all filters.

Relevant code: *WordSimilarities.get_cosine_similarities*

6. (a) From our examination, the 20 most similar words (2nd-order similarity): The BOW (left) column indicates a similarity of words within their shared topic. So for 'car', we will get other words that are some type of vehicle, or are related to vehicles. We can see that

in action when we examine the words that top the column: ‘truck’, ‘automobile’, ‘vehicle’, a ‘motorbike’, and so on... However, the emphasis is not on the function of ‘car’, but the topic it is related to, and that is why some other words will make the cut like ‘driver’ and ‘front-engined’.

The Dep (right) column indicates a similarity of the words within their shared function. So once again, for ‘car’, the top words are the likes of: ‘truck’, ‘suv’, ‘vehicle’, ‘minivan’, and ‘cars’. In contrast to BOW, the emphasis is on the word’s function, and for ‘car’, that function is transportation, preferably on the road, by a four-wheel vehicle. That is why we see that this column is much more focused on different types of vehicles and does not include any engine types or other words that are centered around the topic of vehicles but are not actually a vehicle. Despite that the emphasis is on four-wheel vehicles on the road, other words make the cut too, and they still share the function of a car, words like: ‘motorcycle’, which is a vehicle on the road, but typically has two wheels, or a speedboat which is still a transportation tool, but in the water.

- (b) 10-top context attributes analysis: Distributional Semantics and W2V are similar in the way their contexts are ranked. The BOW (left) column’s top contexts are words that share a topic. For ‘car’ we get: ‘racing’, ‘mygale’ (which is a car manufacturing company for racing cars), ‘bmw’, ‘driver’, and so on... For the Dep (right) column, the top contexts are functions that are used with the word. For ‘car’, the most popular one is: `adpmod:byI_commute`. So in this case, commuting is not very related to ‘car’, however, it gives us a good idea of what the person who commutes does with the car he is referring to.
- (c) MAP and AP: Both BOW and Dependency return sensible lists, however, the dependency one is more nuanced. For instance, the Dependency column for ‘car’ accurately returned mostly road vehicles, with the exception of ‘speedboat’ and ‘wagon’. While ‘speedboat’ in essence shares the same functionality a ‘car’ does, a wagon does not. Nevertheless, ‘wagon’ is still a sensible value, especially for the last entry in the top-20 list. Wagons attached to horses are used to transport people and goods, and it is also a common slang for a car.
- (d) Comparison between distributional semantics and W2V: The distributional semantics we trained and the W2V we examined share the same properties: W2V’s BOW is the comparable version for ‘sentence’ and ‘window’ we have in distributional semantics, and the dependency-column plays a similar roles in both models. We can see that W2V returns results that are more relevant. The first difference between them is that W2V was trained on a larger portion of Wikipedia, and as a result, has an advantage over the distributional semantics we trained. Specifically, BOW and Dependency in

W2V return accurate words for both; topic and semantics. However, the columns they return have little overlap, and thus, the AP and MAP scores do not reach upwards of 90%. Furthermore, the results from W2V have less variance (that is an improvement) between ‘car’ and ‘piano’ in comparison to what we had with the distributional semantics; that is, ‘piano’ performed much better than ‘car’ when we analyzed them with distributional semantics and with W2V they perform more or less the same. Another W2V improvement that perhaps is not reflected in the results is that the semantics W2V returns are much more nuanced. For instance, for ‘piano’: harpsichord, fortepiano, and even marimba feel more semantically right than the similarities we got in the distributional semantics (and those that we still get at the top of the W2V lists), like ‘violin’.

Table 9: Car AP

Judgement/method	BOW5	Dependency
topic	0.71	0.71
semantic	0.59	0.87

Table 10: Piano AP

Judgement/method	BOW5	Dependency
topic	0.74	0.74
semantic	0.67	0.83

Table 11: MAP

Judgement/method	BOW5	Dependency
topic	0.72	0.72
semantic	0.63	0.85

NOTE: word2vec top20 2nd order similarity in appendix 4

NOTE: word2vec top10 1nd order similarity in appendix 5

NOTE: word2vec Piano & Car annotations in appendix 6

Appendix 1

Top20 2nd order similarity

Table 12: Words similarity to car

2-word window	Sentence window	Dependency edge
truck	driver	vehicle
driver	drive	truck
vehicle	truck	driver
motor	vehicle	motorcycle
drive	race	locomotive
racing	motor	automobile
ford	ford	racing
formula	racing	aircraft
automobile	automobile	station
traffic	formula	auto
motorcycle	auto	horse
bus	crash	motor
race	toyota	stock
bicycle	nascar	traffic
aircraft	wheel	bicycle
auto	engine	bus
lap	motorcycle	boat
toyota	ferrari	item
stock	bmw	cyclist
toy	chassis	plane

Table 13: Words similarity to bus

2-word window	Sentence window	Dependency edge
rail	rail	train
passenger	commuter	rail
freight	transit	tram
transit	transportation	ferry
commuter	train	taxi
tram	station	transit
train	transport	railway
metro	passenger	road
terminal	line	vehicle
route	route	route
traffic	freight	traffic
line	operate	passenger
transport	metro	subway
station	tram	boat
ferry	terminal	freight
car	depot	cable
junction	traffic	truck
operate	airport	transport
airport	hub	car
taxi	connect	railroad

Table 14: Words similarity to hospital

2-word window	Sentence window	Dependency edge
clinic	medical	clinic
medical	clinic	school
patient	care	college
psychiatric	health	museum
nursing	medicine	library
care	surgeon	airport
health	surgery	station
facility	center	hall
ward	facility	prison
centre	doctor	center
library	physician	park
center	treatment	hotel
school	rehabilitation	office
rehabilitation	patient	stadium
surgery	dr	campus
surgeon	psychiatric	headquarters
physician	establishment	centre
sick	emergency	theatre
service	centre	church
treatment	nh	jail

Table 15: Words similarity to hotel

2-word window	Sentence window	Dependency edge
resort	restaurant	restaurant
restaurant	resort	resort
shop	inn	store
casino	shop	palace
owner	bar	castle
luxury	owner	shop
apartment	tourist	theater
store	pub	casino
pub	apartment	building
palace	store	estate
plaza	plaza	apartment
mall	chain	farm
residence	dining	inn
retail	supermarket	mill
cafe	retail	station
lobby	accommodation	facility
cottage	house	house
building	attraction	supermarket
shopping	purchase	complex
nightclub	street	hospital

Table 16: Words similarity to gun

2-word window	Sentence window	Dependency edge
fire	cannon	cannon
cannon	rifle	weapon
rifle	artillery	mortar
battery	weapon	pistol
artillery	fire	rifle
arm	bullet	artillery
mortar	assault	rocket
machine	armament	engine
bullet	mortar	sword
weapon	enemy	camera
pistol	ammunition	ammunition
turret	tank	tube
enemy	battery	firing
tank	sniper	missile
rocket	turret	battery
assault	trench	shot
ammunition	pistol	aircraft
shoot	soldier	tank
heavy	grenade	machine
shot	battalion	grenade

Table 17: Words similarity to bomb

2-word window	Sentence window	Dependency edge
bomber	bombing	torpedo
attack	raid	bombing
bombing	luftwaffe	missile
explosive	bomber	explosion
raid	injure	earthquake
torpedo	enemy	shell
plane	explode	bomber
aircraft	explosive	rocket
sink	airfield	weapon
fighter	explosion	fire
destroy	attack	air
luftwaffe	aircraft	ball
pilot	blast	grenade
explosion	weapon	bullet
crash	destroy	destroy
enemy	target	raid
kill	fire	charge
airfield	torpedo	chemical
allied	terrorist	flood
ship	kill	sink

Table 18: Words similarity to horse

2-word window	Sentence window	Dependency edge
ride	jockey	dog
bike	rid	car
breed	rider	motorcycle
carriage	racing	animal
thoroughbred	thoroughbred	thoroughbred
pig	ride	bicycle
sheep	regiment	bike
dog	5th	volunteer
guard	handicap	motor
regiment	artillery	bull
deer	cyclist	cavalry
riding	race	auto
rid	guard	person
jockey	breeder	man
cattle	cavalry	vehicle
bull	pig	cat
motorcycle	dog	infantry
bicycle	brigade	player
rider	detachment	truck
cart	breed	jockey

Table 19: Words similarity to fox

2-word window	Sentence window	Dependency edge
cbs	cbs	cbs
nbc	nbc	abc
abc	abc	nbc
network	affiliate	cbc
news	affiliation	paramount
broadcast	network	bbc
cnn	programming	cable
programming	broadcast	radio
anchor	broadcasting	television
switch	news	smith
television	channel	tv
pb	anchor	entertainment
radio	television	pb
bbc	espn	espn
show	show	itv
tv	newscast	sport
channel	cnn	hudson
broadcasting	switch	anderson
espn	kid	network
cbc	entertainment	shaw

Table 20: Words similarity to table

2-word window	Sentence window	Dependency edge
bottom	row	page
top	column	list
list	list	category
row	contain	watchlist
column	bottom	system
contain	key	picture
template	heading	content
remove	following	infobox
reference	template	scene
rather	top	history
article	element	column
entry	place	section
header	database	diagram
page	header	poll
heading	example	map
content	point	top
box	text	season
second	content	ranking
instead	finish	tower
player	second	image

Table 21: Words similarity to bowl

2-word window	Sentence window	Dependency edge
wicket	victory	cup
bowler	bowler	wrestling
inning	batsman	league
pro	pace	super
super	consecutive	playoff
batsman	super	baseball
bat	inning	soccer
league	wicket	all-ireland
title	score	tennis
season	steelers	nfl
occasional	win	all-star
match	game	ncaa
consecutive	tournament	junior
final	ncaa	football
win	playoff	olympic
ncaa	colt	championship
competition	bat	plate
score	patriot	eurovision
victory	cup	cricket
overall	cowboy	rugby

Table 22: Words similarity to guitar

2-word window	Sentence window	Dependency edge
bass	bass	drum
drum	drum	bass
instrument	keyboard	piano
keyboard	vocal	vocal
vocal	acoustic	keyboard
piano	instrument	flute
acoustic	percussion	cello
string	solo	saxophone
solo	piano	instrument
flute	rhythm	trumpet
vocalist	guitarist	violin
ensemble	flute	percussion
percussion	band	tenor
violin	vocalist	viola
rhythm	backing	string
saxophone	saxophone	organ
melody	drummer	solo
trumpet	musician	guitarist
cello	trumpet	horn
tune	string	music

Table 23: Words similarity to piano

2-word window	Sentence window	Dependency edge
violin	violin	violin
flute	flute	viola
cello	sonata	guitar
concerto	cello	cello
viola	concerto	bass
sonata	percussion	flute
op	bass	keyboard
guitar	trumpet	percussion
string	instrument	horn
trumpet	op	drum
bass	viola	saxophone
saxophone	composition	trumpet
solo	saxophone	instrument
keyboard	quartet	vocal
instrument	tenor	orchestra
quartet	horn	organ
soloist	trio	choir
percussion	guitar	dance
composition	string	solo
ensemble	keyboard	music

Appendix 2

Top20 1nd order similarity

Table 24: Highest attribute's PPMI of car

2-word window	Sentence window	Dependency edge
touring	truck	parking, compmod, in
accident	driver	wash, compmod, in
parking	accident	car, conj, out
driver	racing	truck, conj, out
truck	lap	race, dobj, in
crash	car	drive, partmod, out
racing	ford	concept, compmod, out
formula	crash	accident, compmod, in
manufacturer	formula	race, amod, out
steal	motor	race, partmod, out
ford	drive	armored, amod, out
motor	stock	car, conj, in
drive	wheel	ferry, compmod, in
cable	race	bomb, compmod, in
stock	vehicle	crash, compmod, in
bomb	passenger	drive, dobj, in
buy	speed	fit, nsubjpass, in
concept	engine	hit, adpmod:by, in
race	front	strike, adpmod:by, in
passenger	sport	damage, dobj, in

Table 25: Highest attribute's PPMI of bus

2-word window	Sentence window	Dependency edge
taxi	depot	tram, conj, in
depot	transit	subway, conj, in
terminal	bus	route, nsubj, in
subway	terminal	rout, nsubj, in
tram	operator	stop, compmod, in
interchange	taxi	terminus, compmod, in
transit	interchange	serial, compmod, out
commuter	stop	rail, conj, in
operator	transportation	bus, conj, in
shuttle	truck	paint, nsubjpass, in
shelter	transport	interchange, compmod, in
frequent	metro	accessible, adpmod:by, in
truck	terminus	rail, conj, out
stop	express	stand, compmod, in
parking	regular	bus, conj, out
route	connect	passenger, compmod, in
rapid	passenger	shelter, compmod, in
lane	route	hybrid, amod, out
terminus	travel	terminal, compmod, in
metro	transfer	travel, partmod, out

Table 26: Highest attribute’s PPMI of hospital

2-word window	Sentence window	Dependency edge
psychiatric	psychiatric	doctor, adpmod:at, in
bed	bed	teaching, amod, out
clinic	clinic	mercy, compmod, out
teaching	hospital	rush, adpmod:to, in
dr	teaching	hospital, appos, out
ward	dr	hospital, conj, in
rush	surgery	stay, compmod, in
recover	trust	hospital, appos, in
hospital	patient	clinic, conj, out
admit	doctor	hospital, conj, out
patient	care	psychiatric, amod, out
memorial	medical	bed, compmod, in
treat	memorial	transport, adpmod:to, in
mental	emergency	teaching, compmod, out
doctor	cancer	open, rmod, out
emergency	treatment	child, adpmod:for, out
medical	heart	hopkins, compmod, out
stay	medicine	eye, compmod, out
treatment	facility	visit, adpmod:in, in
cancer	private	take, adpmod:to, in

Table 27: Highest attribute’s PPMI of hotel

2-word window	Sentence window	Dependency edge
hilton	hotel	resort, conj, out
casino	luxury	hotel, appos, in
lobby	resort	lobby, compmod, in
plaza	plaza	neutral, compmod, out
luxury	casino	restaurant, adpmod:at, in
resort	inn	hotel, appos, out
hotel	restaurant	casino, conj, out
restaurant	stay	hotel, conj, out
lodge	chain	plaza, compmod, out
tourism	tourist	resort, compmod, out
chain	palace	vega, adpmod:in, out
stay	guest	luxury, compmod, out
room	owner	hotel, conj, in
palace	purchase	stay, adpmod:at, in
convert	room	restaurant, conj, out
tourist	shop	palace, compmod, out
check	bar	operate, adpmod:a, in
nearby	tower	convert, adpmod:into, in
purchase	spring	windsor, compmod, out
bar	grand	milk, compmod, out

Table 28: Highest attribute’s PPMI of gun

2-word window	Sentence window	Dependency edge
anti-aircraft	anti-aircraft	rifle, amod, out
turret	armament	deck, compmod, out
machine	turret	aim, dobj, in
ammunition	ammunition	barrel, compmod, in
barrel	machine	machine, compmod, out
mortar	barrel	man, dobj, in
mm	gun	fire, rcmmod, out
battery	rose	battery, adpmod:of, in
rose	battery	armstrong, compmod, out
n	rifle	mm, compmod, out
deck	shot	gun, conj, out
rifle	deck	arm, adpmod:with, in
gun	fit	ammunition, conj, out
shot	assault	shoot, adpmod:with, in
shield	artillery	sword, conj, in
instal	n	mount, nsubjpass, in
mount	shoot	jump, dobj, in
lewis	weapon	mount, nsubj, in
artillery	mount	rifle, conj, in
jump	heavy	mortar, conj, out

Table 29: Highest attribute’s PPMI of bomb

2-word window	Sentence window	Dependency edge
explode	explode	luftwaffe, nsubj, out
luftwaffe	luftwaffe	pipe, amod, out
petrol	ira	ally, adpmod:by, out
disposal	explosion	explode, nsubj, in
atomic	bombing	plant, partmod, out
blast	explosive	explode, compmod, in
ira	atomic	aim, nsubjpass, in
explosive	raid	wave, adpmod:in, out
raid	bomb	cluster, compmod, out
conventional	drop	raid, dobj, out
explosion	injure	blast, compmod, in
hydrogen	bomber	explosion, compmod, in
drop	target	aircraft, adpmod:by, out
pipe	plane	atomic, amod, out
raf	nuclear	ira, compmod, out
terrorist	sink	drop, dobj, in
suicide	destroy	disposal, compmod, in
squad	least	hydrogen, compmod, out
allied	damage	alert, compmod, in
bomb	weapon	raid, amod, in

Table 30: Highest attribute’s PPMI of horse

2-word window	Sentence window	Dependency edge
harness	mounted	thoroughbred, amod, out
trojan	thoroughbred	crazy, compmod, out
crazy	rid	bull, conj, in
thoroughbred	jockey	race, adpmod:for, in
mule	horse	rid, partmod, out
pony	rider	ride, dobj, in
rid	riding	wagon, conj, out
carriage	trailer	man, adpmod:on, in
jockey	ride	rider, conj, out
stable	cattle	troop, adpmod:of, in
riding	breed	rid, dobj, in
ride	stake	ride, conj, out
rider	racing	breeding, compmod, in
racing	wild	judge, nsubjpass, in
cattle	cavalry	race, partmod, out
sheep	artillery	tram, compmod, in
breed	farm	age, partmod, out
wild	brigade	fall, adpmod:from, in
pull	trail	trailer, compmod, in
steal	regiment	thoroughbred, compmod, out

Table 31: Highest attribute’s PPMI of fox

2-word window	Sentence window	Dependency edge
hound	hound	morning, adpmod:on, in
sac	fox	beaver, conj, in
coyote	affiliation	coyote, conj, in
affiliation	kid	televise, adpmod:on, in
affiliate	affiliate	hare, conj, in
cnn	nbc	sac, conj, in
hunting	switch	abc, conj, in
kid	abc	hunting, compmod, in
net	cbs	ohio, compmod, in
matthew	programming	subspecies, adpmod:of, in
fox	20th	cat, conj, out
programming	entertainment	kit, compmod, out
nbc	news	net, compmod, in
news	sport	affiliation, compmod, in
hunt	channel	twentieth, compmod, out
switch	morning	gray, amod, out
abc	picture	kid, compmod, in
twentieth	originally	squirrel, compmod, in
20th	network	mask, compmod, in
distribute	soccer	sport, compmod, in

Table 32: Highest attribute’s PPMI of table

2-word window	Sentence window	Dependency edge
periodic	table	following, compmod, out
picnic	row	periodic, compmod, out
tennis	gospel	list, adpmod:below, out
canon	column	picnic, amod, out
table	database	tennis, compmod, in
column	tennis	truth, compmod, out
row	following	table, conj, in
database	key	column, adpmod:of, in
coffee	element	sit, adpmod:at, in
bottom	format	picnic, compmod, out
following	template	medal, compmod, out
chair	content	content, adpmod:of, out
salt	text	table, conj, out
content	contain	salt, compmod, in
knight	top	periodic, amod, out
indicate	finish	rotary, amod, out
entry	round	chair, conj, in
simply	medal	compare, nsubj, in
key	section	update, dobj, in
element	turn	row, adpmod:of, in

Table 33: Highest attribute’s PPMI of bowl

2-word window	Sentence window	Dependency edge
quiz	quiz	batsman, rcm, in
batsman	rose	take, adpm:without, out
super	pace	bowl, conj, in
right-handed	bowl	quiz, compmod, out
subdivision	steelers	bowl, conj, out
bermuda	right-handed	select, adpm:to, in
rose	super	air, adpm:during, in
alley	batsman	alley, amod, in
pro	pro	green, dobj, out
lawn	bowler	rose, compmod, out
bowler	sugar	subdivision, compmod, in
hawaii	consecutive	ball, rcm, in
cotton	orange	invitation, adpm:to, in
sugar	wicket	hand, nsubjpass, in
invitation	cotton	pace, dobj, out
dust	formerly	bermuda, compmod, out
liberty	inning	game, appos, out
bowl	nfl	pro, compmod, out
orange	rename	barrow, compmod, in
wicket	victory	super, compmod, out

Table 34: Highest attribute’s PPMI of guitar

2-word window	Sentence window	Dependency edge
acoustic	gibson	vocal, appos, in
amplifier	acoustic	vocal, appos, out
vintage	line-up	flourish, compmod, in
gibson	amplifier	intro, compmod, in
rhythm	playing	tune, dobj, in
bass	backing	gibson, compmod, out
synthesizer	rhythm	bass, appos, out
keyboard	percussion	vocal, conj, out
percussion	keyboard	keyboard, conj, out
vocal	signature	amplifier, compmod, in
backing	bass	rhythm, compmod, out
tenor	drum	vocal, conj, in
hero	saxophone	acoustic, amod, out
harmony	vocal	solo, conj, out
electric	guitar	bass, compmod, out
playing	hero	bass, conj, out
drum	electric	consist, adpm:on, in
tune	string	percussion, conj, out
signature	piano	drum, conj, out
string	violin	solo, compmod, in

Table 35: Highest attribute’s PPMI of piano

2-word window	Sentence window	Dependency edge
opus	opus	harp, conj, in
sonata	sonata	saxophone, conj, in
concerto	op	cum, compmod, in
recital	cello	aid, amod, out
op	viola	percussion, conj, in
viola	trio	viola, conj, in
trio	lesson	cello, conj, in
cello	concerto	op, conj, out
lesson	violin	compose, adpmod:for, in
synthesizer	clarinet	recital, compmod, in
violin	arrangement	train, adpmod:on, in
clarinet	pianist	trio, adpmod:for, in
trumpet	piano	ph, compmod, out
harp	trumpet	flute, conj, out
saxophone	saxophone	roll, dobj, out
suite	flute	voice, conj, in
arrangement	organ	organ, conj, in
flute	composition	violin, conj, in
organ	horn	violin, conj, out
arrange	acoustic	lesson, compmod, in

Appendix 3

Car & Piano annotations

Table 36: Car’s annotations with 100 attributes

2-word window	topic	semantic	Sentence	topic	semantic	Dependency	topic	semantic
truck	+	+	driver	+	-	vehicle	+	+
driver	+	-	drive	+	-	truck	+	+
vehicle	+	+	truck	+	+	driver	+	-
motor	+	-	vehicle	+	+	motorcycle	+	-
drive	+	-	race	+	-	locomotive	+	-
racing	+	-	motor	+	-	automobile	+	+
ford	+	+	ford	+	+	racing	+	-
formula	+	-	racing	+	-	aircraft	+	-
automobile	+	+	automobile	+	+	station	-	-
traffic	+	-	formula	+	-	auto	+	+
motorcycle	-	-	auto	+	+	horse	+	-
bus	+	+	crash	+	-	motor	+	-
race	+	-	toyota	+	+	stock	-	-
bicycle	+	-	nascar	+	-	traffic	+	-
aircraft	+	-	wheel	+	-	bicycle	+	-
auto	+	+	engine	+	-	bus	+	+
lap	+	-	motorcycle	-	-	boat	+	-
toyota	+	+	ferrari	+	+	item	-	-
stock	-	-	bmw	+	+	cyclist	-	-
toy	-	-	chassis	-	-	plane	+	-

Table 37: Piano’s annotations with 100 attributes

2-word window	topic	semantic	Sentence	topic	semantic	Dependency	topic	semantic
violin	+	+	violin	+	+	violin	+	+
flute	+	+	flute	+	+	viola	+	+
cello	+	+	sonata	+	-	guitar	+	+
concerto	+	-	cello	+	+	cello	+	+
viola	+	+	concerto	+	-	bass	+	+
sonata	+	-	percussion	+	+	flute	+	+
op	+	-	bass	+	+	keyboard	+	+
guitar	+	+	trumpet	+	+	percussion	+	+
string	+	-	instrument	+	+	horn	-	-
trumpet	+	+	op	+	-	drum	+	-
bass	+	+	viola	+	+	saxophone	+	-
saxophone	+	-	composition	+	-	trumpet	+	+
solo	-	-	saxophone	+	+	instrument	+	+
keyboard	+	+	quartet	+	-	vocal	+	-
instrument	+	+	tenor	+	-	orchestra	+	-
quartet	+	-	horn	-	-	organ	+	+
soloist	+	-	trio	+	-	choir	+	-
percussion	+	+	guitar	+	+	dance	+	-
composition	+	-	string	+	-	solo	-	-
ensemble	+	-	keyboard	+	+	music	+	-

Appendix 4

Table 38: Top 20 similar words for: car

BoW-5	Dependency edge
cars	truck
truck	suv
automobile	vehicle
vehicle	minivan
motorbike	cars
motorcycle	speedboat
driver	racecar
minivan	automobile
suv	motorcar
lorry	jeep
motorcar	limousine
mid-engined	minibus
limousine	lorry
front-engined	limo
moped	motorcycle
motorhome	bike
mercedes-benz	motorhome
bike	taxicab
rear-engined	roadster
three-wheeled	wagon

Table 39: Top 20 similar words for: bus

BoW-5	Dependency edge
buses	minibus
tram	tram
metrobus	buses
intercity	jeepney
busses	taxicab
fixed-route	motorcoach
minibus	taxi
inter-city	trolleybus
ksrtc	lorry
commuter	truck
apsrtc	metrobus
msrtc	streetcar
inter-urban	busses
dial-a-ride	ferryboat
mini-bus	trolley
light-rail	tramcar
rail	railcar
transit	railmotor
trolleybus	intercityexpress
limited-stop	train

Table 40: Top 20 similar words for: hospital

BoW-5	Dependency edge
clinic	sanatorium
hospitals	hospice
infirmiry	sanitorium
hospice	hospitals
lying-in	sanitarium
dispensary	clinic
polyclinic	infirmiry
sanatorium	polyclinic
convalescent	dispensary
mulago	orphanage
addenbrooke	poorhouse
bethlem	almshouse
psychiatric	workhouse
maudsley	institutet
siriraj	leprosarium
sanitarium	rikshospitalet
in-patient	heliport
incurables	gaol
orthopaedic	guesthouse
westmead	motherhouse

Table 41: Top 20 similar words for: hotel

BoW-5	Dependency edge
motel	motel
restaurant	hotels
doubletree	casino
sheraton	restaurant
hotels	inn
ritz-carlton	guesthouse
sofitel	tavern
westin	cafe
ramada	ritz-carlton
casino	nightclub
kempinski	travelodge
mansion	pizzeria
inn	roadhouse
cafe	boardinghouse
tavern	café
apartments	condo
boutique	brewpub
nightclub	sheraton
marriott	steakhouse
travelodge	brasserie

Table 42: Top 20 similar words for: gun

BoW-5	Dependency edge
guns	guns
cannon	handgun
howitzer	machinegun
sub-machine	howitzer
flamethrower	pistol
belt-fed	rifle
37mm	shotgun
smoothbore	firearm
pistol	cannon
shkas	musket
105mm	crossbow
40mm	autocannon
gatling	phaser
recoilless	flamethrower
76mm	revolver
3-inch	carbine
rifle	machine-gun
88mm	weapon
large-caliber	carronade
autocannons	pounder

Table 43: Top 20 similar words for: bomb

BoW-5	Dependency edge
bombs	bombs
detonated	firebomb
detonates	landmine
detonate	car-bomb
booby-trap	grenade
detonating	torpedo
firebomb	ied
car-bomb	warhead
exploded	bomber
detonation	bomblets
warhead	missile
500-pound	nuke
b61	detonator
laser-guided	booby-trap
blast	kamikaze
explosives	munition
detonations	explosives
landmine	machinegun
tallboy	a-bomb
thermonuclear	firebombs

Table 44: Top 20 similar words for: horse

BoW-5	Dependency edge
horses	horses
standardbred	goat
saddlebred	dog
gelding	stallion
thoroughbred	mule
stallion	bronc
racehorses	cow
dog	unicycle
riderless	greyhound
gaited	bareback
bronc	camel
percheron	appaloosa
pony	saddlebred
trotting	colt
harness	zebu
chariot	donkey
appaloosa	sidesaddle
sulky	racehorse
racehorse	elephant
greyhound	pony

Table 45: Top 20 similar words for: fox

BoW-5	Dependency edge
abc	daystar
cbs	nbc
nbc	byutv
wxyz-tv	kron
msnbc	wolf
ctv	cbs-tv
wsvn	familynet
familynet	abc
wttg	wccb
wjbk	oln
wfxt	wjar
espn	hdnet
wofl	telefutura
cnn	woodchuck
oln	nbc-tv
blitzer	soapnet
nesn	cinemax
wesh	wdiv
wb	mundofox
wgn-tv	coyote

Table 46: Top 20 similar words for: table

BoW-5	Dependency edge
tables	tables
sortable	leaderboard
wikitable	sideboard
look-up	chessboard
foosball	textbox
toc	taskbar
bulleted	gameboard
ping-pong	worksheet
billiard	tray
table-tennis	viewport
textbox	dais
tray	flowchart
lookup	playfield
wikitables	mantelpiece
brackets	stepladder
header	cladogram
footer	letterbox
tabular	windowsill
menu	bookcase
carom	wikitable

Table 47: Top 20 similar words for: bowl

BoW-5	Dependency edge
xliii	bowls
xlii	superbowl
xliv	arenabowl
bowls	wcws
xlvi	wnit
tostitos	nlcs
xli	arenacup
xxxviii	postseason
xl	nit
xxxv	xliii
xxxix	llws
xxxvii	beanpot
xlvi	xl
xxxvi	triplemanfa
xxxiv	alcs
bluebonnet	nlds
gator	cup
xxviii	kvalserien
xxxii	tourney
xxxi	cws

Table 48: Top 20 similar words for: guitar

BoW-5	Dependency edge
harmonica	saxophone
mandolin	bass
bass	mandolin
drums	harmonica
guitars	accordion
keyboards	trombone
accordion	violin
banjo	banjo
saxophone	guitars
12-string	cello
ukulele	piano
trombone	vibraphone
fiddle	sax
autoharp	trumpet
melodica	autoharp
percussion	clarinet
vibraphone	sitar
tambourine	fiddle
vocals	drums
fretless	marimba

Table 49: top 20 similar words for: piano

BoW-5	Dependency edge
violin	violin
cello	cello
harpsichord	harpsichord
clarinet	saxophone
viola	clarinet
flute	guitar
bassoon	trombone
violoncello	mandolin
oboe	vibraphone
concerto	marimba
saxophone	accordion
accordion	pianoforte
harp	bassoon
trombone	fortepiano
sonatas	violoncello
trumpet	trumpet
mandolin	harmonica
pianoforte	clavinet
vibraphone	clavichord
concertos	euphonium

word2vec - Top20 2nd order similarity

Appendix 5

word2vec - Top10 1nd order similarity

Table 50: Top 10 attributes for: car

BoW-5	Dependency edge
car	adpmod:byI_commute
racing	amod_street-legal
mygale	conj_hovercraft
bmw	amod_newly-designed
driver	amod_liter
motor	amod_late-model
cars	poss_brink
dealership	compmod_m1918
parked	adpmod:fromI_tossed
rear-drive	compmod_high-wing

Table 51: Top 10 attributes for: bus

BoW-5	Dependency edge
bus	adpmod:byI_commute
buses	amod_east-bound
ksrtc	conj_hovercraft
samtrans	adpmod:onI_ridership
seabus	compmod_airlink
smrt	compmod_operates
connexxion	conj_busses
inter-city	compmod_xpt
intercity	compmod_yrt
msrtc	dojI_onboard

Table 52: Top 10 attributes for: hospital

BoW-5	Dependency edge
hospital	compmod_siriraj
bethlem	compmod_safdarjung
moorfields	adpmod:of_nuova
psychiatric	compmod_strangeways
hospitals	amod_maximum-security
infirmary	compmod_armley
foundling	compmod_combermere
siriraj	compmod_fresnes
maudsley	compmod_eastview
westmead	compmod_desloge

Table 53: Top 10 attributes for: hotel

BoW-5	Dependency edge
hotel	compmod_nymphenburg
hotels	compmod_dolmabahçe
westin	compmod_whitwell
radisson	conj_guesthouses
kempinski	compmod_ravenscourt
sofitel	compmod_hanworth
ramada	compmod_siriraj
biltmore	compmod_strangeways
ritz-carlton	compmod_safdarjung
sheraton	compmod_armley

Table 54: Top 10 attributes for: gun

BoW-5	Dependency edge
gun	num_88mm
guns	adpmod:withI_rearmed
submachine	compmod_m1918
machine	compmodI_fellatio
gatling	num_60mm
sub-machine	adpmod:ofI_hilt
howitzer	compmod_karabiner
rifle	compmod_t-55
40mm	adpmod:forI_cartridges
11-inch	compmodI_cwt

Table 55: Top 10 attributes for: bomb

BoW-5	Dependency edge
bomb	num_88mm
bombs	adpmod:on_pentagon
detonated	compmodI_splashes
detonates	adpmod:byI_unharmd
detonate	adpmod:withI_rearmed
atomic	compmod_single-car
exploded	compmod_m1918
detonating	compmod_phishing
bomber	compmodI_fellatio
bombing	rcmod_bursts

Table 56: Top 10 attributes for: horse

BoW-5	Dependency edge
horse	amod_gaited
horses	nsubjpassI_spooked
standardbred	compmodI_drovers
thoroughbred	dobjI_shoe
trotting	adpmod:byI_commute
riding	rcmod_roams
racing	poss_quixote
ridden	compmod_ch-53e
galloping	appos_pony
bred	compmod_poitevin

Table 57: Top 10 attributes for: fox

BoW-5	Dependency edge
fox	amod_crab-eating
vulpes	appos_canadensis
news	compmodI_subchannels
movietone	conj_mattel
owned-and-operated	compmod_jacki
nbc	conjI_raimi
cbs	conjI_corgan
terrier	conjI_leda
affiliate	conj_bluebird
cnn	conj_nolte

Table 58: Top 10 attributes for: table

BoW-5	Dependency edge
table	adpmod:inI_rows
sortable	adpmod:offI_knocks
tables	adpmod:onI_seventh
tennis	amod_ten-foot
lookup	compmod_abydos
billiards	amod_25-metre
foosball	adpmod:intoI_face
periodic	adpmod:belowI_added
billiard	adpmod:throughI_throw
toc	compmod_three-judge

Table 59: Top 10 attributes for: bowl

BoW-5	Dependency edge
bowl	adpmod:against_auburn
xli	conjI_preseason
bowls	adpmod:against_oklahoma
xliv	adpmod:inI_2-1
xxxviii	conj_play-offs
xlili	compmod_pan-pacific
xlvi	adpmod:forI_stadiums
xlvi	adpmod:atI_clinched
xlvi	adpmod:intoI_win
super	amod_25-metre

Table 60: Top 10 attributes for: guitar

BoW-5	Dependency edge
guitar	conjI_bongos
bass	amod_end-blown
guitars	conj_back-up
drums	adpmod:onI_harris
keyboards	adpmod:forI_adagio
nyckelharpa	adpmod:onI_thompson
glockenspiel	appos_mandolin
6-string	dep_keyboard
tambourine	adpmod:ofl_virtuoso
banjo	adpmod:onI_foster

Table 61: Top 10 attributes for: piano

BoW-5	Dependency edge
piano	adpmod:forI_adagio
violin	amod_end-blown
sonata	conjI_bongos
cello	adpmod:onI_harris
concerto	adpmod:ofl_virtuoso
op	adpmod:onI_thompson
harpsichord	conj_back-up
concertos	compmod_kreutzer
viola	compmod_vomeronasal
violoncello	adpmod:onI_supplemented

Appendix 6

word2vec - Car & Piano annotations

Table 62: word2vec Car’s annotations

BOW5	topic	semantic	Dependency	topic	semantic
cars	+	+	truck	+	+
truck	+	+	suv	+	+
automobile	+	+	vehicle	+	+
vehicle	+	+	minivan	+	+
motorbike	+	+	cars	+	+
motorcycle	+	+	speedboat	+	+
driver	+	-	racecar	+	+
minivan	+	+	automobile	+	+
suv	+	+	motorcar	+	+
lorry	+	+	jeep	+	+
motorcar	+	+	limousine	+	+
mid-engined	+	-	minibus	+	+
limousine	+	+	lorry	+	+
front-engined	+	-	limo	+	+
moped	+	+	motorcycle	+	+
motorhome	+	+	bike	+	+
mercedes-benz	+	+	motorhome	+	+
bike	+	+	taxicab	+	+
rear-engined	+	-	roadster	+	+
three-wheeled	+	-	wagon	+	+

Table 63: word2vec Piano’s annotations

BOW5	topic	semantic	Dependency	topic	semantic
violin	+	+	violin	+	+
cello	+	+	cello	+	+
harpsichord	+	+	harpsichord	+	+
clarinet	+	+	saxophone	+	+
viola	+	+	clarinet	+	+
flute	+	+	guitar	+	+
bassoon	+	+	trombone	+	+
violoncello	+	+	mandolin	+	+
oboe	+	+	vibraphone	+	+
concerto	+	-	marimba	+	+
saxophone	+	+	accordion	+	+
accordion	+	+	pianoforte	+	+
harp	+	+	bassoon	+	+
trombone	+	+	fortepiano	+	+
sonatas	+	-	violoncello	+	+
trumpet	+	+	trumpet	+	+
mandolin	+	+	harmonica	+	+
pianoforte	+	+	clavinet	+	+
vibraphone	+	+	clavichord	+	+
concertos	+	-	euphonium	+	+