Report

Danit Yshaayahu 312434269 ; Refael Shaked Greenfeld

1. Statistics:
   a. We used the required threshold :
      i. Minimum occurrence of 75 times for a lemma in order to count it as a feature.
      ii. Minimum occurrence of 100 for a word in order to create for it a vector (if not it is not represented)
      iii. We also had a limit of 100 context feature (which is a lemma for the sentence and word window and a tuple of (direction, pos, lemma) for the Dependency window.
   b. Number of words considered for similarity:
      i. Word – 8218
      ii. Sentence – 8218
      iii. Dependency – 9777
   c. Number of features considered for similarity:
      i. Word – 9783
      ii. Sentence – 9784
      iii. Dependency – 96605

2. There are clear differences between the 3 context vectors. First, we can see that for the Window vector we mostly get in the top 20 similar words or a the general meaning of a word, for example ; $Hospital \rightarrow Facility$ or a type of the word $car \rightarrow b.m.w\ and\ ford$; $horse \rightarrow thoroughbred$ (a type of a racing horse breed) and finally we get also synonyms e.g. $car \rightarrow vehicle\ and\ automobile$.

   In the other hand, when we look at the words from the sentence we get more words in the **Domain of the word**. For example $gun \rightarrow soldier\ and\ enemy$ which although they are very different than gun but soldier uses a gun in order to kill the enemy. This also explain the reason we get a lot of different parts of speech for a car like drive (verb) fast (adverb) not like in the window when we mostly got nouns (as shown above). In the dependency part we anticipate an interesting phenomenon. We get wprds from a completely different domain, and seemingly not related, but actually it has the same meaning but in a different domain. For example, we get $car \rightarrow horse\ and\ acrrige$ which they are not a type of a car or have an engine, **but** they are both things that use to move people from place to place, another example is $hotel \rightarrow casino\ and\ shop\ and\ theater$ which again, seem like unrelated words (no one sleep in a casino, at least in purpose) bit they are all places people go to enjoy or spend their money.

So to conclude:

Window give us "type of"; generalization ; and synonyms, **Sentence** give us things from the domain and **Dependency** gives things that have the same use.

3. When examining the attributes with the highest PMI value for each context vector we can see that when we compare them with the words we got in item 2, the **Window** and **Sentence** attributes are much similar to each other, than in the second order similarity lists. When the Dependency one is standing out by the fact we get repetitive words but with different labels. We can explain that by the fact that the Sentence and the Window vectors are quite similar in the way the attributes are counted. The results itself are naturally different, we don't necessarily get similar words, but similar attributes for example, we get attributes like $fox \rightarrow 20th$, which makes sense to be with high PMI value, in another words the attributes with the highest PMI are more a descriptive criteria than a similar word.

   In addition, when comparing the results between the 20 similar words and 20 to context attributes for the dependency context vector we can see that the attributes are with repetitive words that come with different label (and from different directions) for example ($guitar \rightarrow compmod\_D\_bass$; $appos\_D\_bass$; $compmod\_D\_bass$ or $hotel \rightarrow conj\_D\_casino$; $compmod\_P\_casino$. What naturally push down the other attributes.

4. We can see that the dependency attend to return words that are more semantic related than the Window and Sentence one. And between the Window and Sentence we get pretty similar result with a slight more semantic relation in the window one, but not in a significantly way manner.

Semantic is marked in red
Topic is marked in blue

	car		
	window	sentence	dependency
0	truck	drive	vehicle
1	vehicle	driver	motorcycle
2	driver	truck	truck
3	automobile	race	automobile
4	drive	motor	driver
5	race	vehicle	aircraft
6	locomotive	ford	bus
7	ford	racing	locomotive
8	racing	run	auto

9	motorcycle	motorcycle	racing
10	motor	automobile	wagon
11	bus	set	bicycle
12	aircraft	auto	traffic
13	formula	toyota	station
14	bmw	fast	cruiser
15	boat	model	motor
16	prototype	chase	carriage
17	start	eventually	stock
18	model	drop	cyclist
19	train	home	horse

	window	Sentence	dependency	AVG
Semantically	9/20	7/20	13/20	0.48
Topically	11/20	13/20	7/20	0.52

piano

	window	sentence	dependency
0	violin	violin	clarinet
1	guitar	flute	violin
2	flute	cello	viola
3	solo	sonata	cello
4	cello	bass	guitar
5	concerto	percussion	flute
6	bass	concerto	bass
7	percussion	trumpet	keyboard
8	trumpet	saxophone	trumpet
9	keyboard	quartet	horn
10	saxophone	tenor	percussion
11	op	composition	saxophone
12	instrument	guitar	harp
13	drum	instrument	instrument
14	string	viola	drum
15	viola	soloist	vocal
16	vocal	orchestra	dance
17	sonata	string	orchestra
18	orchestra	op	composition
19	acoustic	trio	organ

	window	Sentence	dependency	**AVG**
Semantically	9/20	8/20	13/20	**0.5**
Topically	11/20	12/20	7/20	**0.5**

5. Write a brief description of your implementation:
   a. PMI:

   Instead of using the "Vanilla" version from class, we did some arithmetical modifications that created

   $$PMI(a,b) = \log\left(\frac{P(a,b)}{P(a)\cdot P(b)}\right) = \log\left(\frac{\frac{n(a,b)}{N}}{\frac{n(a)}{N}\cdot\frac{n(b)}{N}}\right) = \log\left(\frac{n(a,b)\cdot N}{n(a)\cdot n(b)}\right)$$

   The implementation is under the function $Vectorizer.pmi$ (Vectorizer is the abstract class of **all** the context vectors). And it being used in the creation of the vectors.
   b. Efficient Algorithm:

   First, we normalized all the vectors using l2 normalization (this part was performed in the vectors creation and not exactly relates to the algorithm), the normalization saved us the computation of cosine and we only had to preform dot product.

   For each pair of vectors we first found the mutual features. And then we multiply every corresponding feature and summed the result. The implementation is under $Vectorizer.efficient\_algo$ (Vectorizer is the abstract class of **all** the context vectors). And it being used in the $Vectorizer.get\_most\_similar$ which is the method that return the most similar vectors (surprise surprise) for a given word.

6. For part 2, When looking at the our lists and the w2v ones, we can see that the results are similar in their general meaning – the dependency give more semantic related words when the bow gives us more topic related ones like in the Sentence and the Window. But with the words themselves are slightly different because they have more rare words and also they are not in the lemma level but in their actual form. For example for fox all of the 5 different vectors we get tv networks (abc, cbs, etc.) but in w2v ones we will get also more unique version of them such as cbs-tv.
   When looking at the bow and context as a whole we can also say that they gives us more much more sematic related words than in out lists. (about 85% compared to about 50%)

   For part 3, we get more specific people or entities that define the word (high value attributes) in the w2v attributes (for example, Thompson and kreutzer for piano in the w2v-dep). Also, differently than our result we didn't notice repetitive words (with

different labels) in the deps w2v vectors, which could be explained by the richer vocabulary of them (which yield more context option).

For part 4, In w2v there are significantly more semantic words than topically words and also there are more semantic words in dependency feature than in window as we saw in the results of item 4. (See full results in Appendix 2 and in the files top20_w2v.txt and context_top20_w2v.txt)

```
                    car
            bow5            deps
0              cars            truck
1             truck             suv
2         automobile         vehicle
3            vehicle         minivan
4           motorbike            cars
5          motorcycle       speedboat
6              driver          racecar
7             minivan       automobile
8                 suv         motorcar
9               lorry            jeep
10            motorcar        limousine
11          mid-engined        minibus
12           limousine           lorry
13         front-engined          limo
14               moped       motorcycle
15           motorhome           bike
16        mercedes-benz       motorhome
17                bike          taxicab
18          rear-engined        roadster
19         three-wheeled          wagon
```

	Bow5	dependency	AVG
Semantically	12/20	19/20	0.775
Topically	8/20	1/20	0.225

```
                  piano
            bow5            deps
0             violin           violin
1              cello            cello
2         harpsichord      harpsichord
3            clarinet        saxophone
4               viola         clarinet
5               flute           guitar
6             bassoon         trombone
7          violoncello         mandolin
8                oboe        vibraphone
9            concerto          marimba
10          saxophone        accordion
```

		accordion	pianoforte
11		accordion	pianoforte
12		harp	bassoon
13		trombone	fortepiano
14		sonatas	violoncello
15		trumpet	trumpet
16		mandolin	harmonica
17		pianoforte	clavinet
18		vibraphone	clavichord
19		concertos	euphonium

	Bow5	dependency	AVG
Semantically	17/20	19/20	0.9
Topically	3/20	1/20	0.1

Second order top 20 results (Window Sentence Dependency)

_____

	car		
	window	sentence	dependency
0	truck	drive	vehicle
1	vehicle	driver	motorcycle
2	driver	truck	truck
3	automobile	race	automobile
4	drive	motor	driver
5	race	vehicle	aircraft
6	locomotive	ford	bus
7	ford	racing	locomotive
8	racing	run	auto
9	motorcycle	motorcycle	racing
10	motor	automobile	wagon
11	bus	set	bicycle
12	aircraft	auto	traffic
13	formula	toyota	station
14	bmw	fast	cruiser
15	boat	model	motor
16	prototype	chase	carriage
17	start	eventually	stock
18	model	drop	cyclist
19	train	home	horse

	bus		
	window	sentence	dependency
0	tram	rail	train
1	rail	commuter	tram
2	train	line	rail
3	line	transit	taxi
4	passenger	station	traffic
5	railway	train	subway
6	commuter	operate	car
7	metro	route	railway
8	airport	metro	commuter
9	transport	traffic	automobile
10	operate	railway	freight
11	ferry	tram	transport
12	station	transport	vehicle

13	freight	transportation	auto
14	transit	connect	transit
15	terminal	passenger	road
16	taxi	hub	ferry
17	junction	airport	truck
18	road	freight	passenger
19	operator	terminus	railroad

hospital

	window	sentence	dependency
0	centre	clinic	clinic
1	center	medical	college
2	facility	center	school
3	headquarters	facility	campus
4	clinic	care	library
5	park	centre	hall
6	office	city	museum
7	school	establishment	institution
8	campus	health	stadium
9	museum	house	station
10	library	st	church
11	medical	home	cemetery
12	residence	office	university
13	community	campus	airport
14	care	medicine	center
15	hall	school	chapel
16	town	nearby	centre
17	city	establish	academy
18	airport	community	park
19	home	residence	hotel

hotel

	window	sentence	dependency
0	restaurant	restaurant	resort
1	residence	shop	restaurant
2	house	home	store
3	resort	tourist	shop
4	building	street	casino
5	palace	house	supermarket
6	store	garden	theater
7	apartment	store	theatre

8	shop	city	mill
9	plaza	location	apartment
10	inn	building	building
11	castle	hall	castle
12	tower	centre	residence
13	farm	cottage	station
14	manhattan	build	house
15	estate	palace	farm
16	park	inn	inn
17	mall	nearby	estate
18	mansion	resort	hospital
19	mill	main	cottage

gun

	window	sentence	dependency
0	cannon	cannon	cannon
1	fire	rifle	weapon
2	weapon	enemy	mortar
3	tank	weapon	rifle
4	artillery	fire	pistol
5	machine	artillery	artillery
6	rifle	arm	battery
7	aircraft	assault	shotgun
8	mortar	tank	rocket
9	pistol	battery	firing
10	enemy	ammunition	ammunition
11	arm	heavy	engine
12	bullet	mortar	armament
13	heavy	armament	grenade
14	battery	attack	tank
15	load	trench	bullet
16	armor	capture	machine
17	barrel	bullet	aircraft
18	ammunition	soldier	knife
19	turret	pistol	shot

bomb

	window	sentence	dependency
0	bomber	raid	bombing
1	aircraft	attack	torpedo
2	attack	enemy	weapon

3	bombing	bombing	grenade
4	plane	destroy	missile
5	destroy	bomber	shell
6	explosive	airfield	tree
7	torpedo	fly	rocket
8	fighter	explode	bomber
9	raid	force	attack
10	explosion	aircraft	air
11	pilot	explosion	explosive
12	sink	fighter	explosion
13	kill	fire	engage
14	submarine	pilot	bombay
15	fire	shoot	commando
16	crash	explosive	charge
17	enemy	kill	strike
18	target	strike	sink
19	damage	luftwaffe	chemical

horse

	window	sentence	dependency
0	dog	race	pony
1	ride	rider	dog
2	car	time	motorcycle
3	guard	racing	car
4	regiment	men	bicycle
5	man	dog	thoroughbred
6	pig	guard	bike
7	racing	thoroughbred	animal
8	breed	place	volunteer
9	cart	woman	bull
10	soldier	seven	truck
11	bird	field	camel
12	wild	great	racing
13	light	lead	canoe
14	rider	know	cat
15	motorcycle	regiment	infantry
16	bike	rid	mare
17	home	white	motor
18	thoroughbred	turn	jockey
19	rid	world	goat

fox

	window	sentence	dependency
0	abc	cbs	cbs
1	nbc	nbc	nbc
2	cbs	abc	cbc
3	network	broadcast	abc
4	television	television	coyote
5	tv	network	bbc
6	anchor	channel	pb
7	radio	affiliate	paramount
8	broadcast	tv	radio
9	entertainment	broadcasting	sport
10	pb	news	cable
11	news	programming	deer
12	host	host	sb
13	movie	anchor	j
14	mtv	espn	television
15	bbc	sport	itv
16	affiliate	radio	cnn
17	broadcasting	entertainment	hudson
18	switch	series	anderson
19	channel	affiliation	shaw

table

	window	sentence	dependency
0	list	time	system
1	text	example	chair
2	contain	able	watchlist
3	note	place	category
4	template	way	page
5	section	list	file
6	reference	point	entry
7	usually	set	match
8	category	etc	knockout
9	point	combine	spelling
10	example	category	footnote
11	mean	instance	fixture
12	number	determine	badminton
13	page	instead	box
14	paragraph	like	history
15	suggest	come	account

16	article	individual	pillar
17	content	draw	section
18	source	row	engine
19	instead	previous	spot

bowl

	window	sentence	dependency
0	tournament	season	cup
1	team	victory	league
2	season	game	playoff
3	championship	consecutive	wrestling
4	ncaa	team	super
5	match	tournament	season
6	cup	final	championship
7	win	cup	all-star
8	playoff	ncaa	ncaa
9	trophy	super	baseball
10	victory	playoff	eurovision
11	wicket	match	tournament
12	nfl	win	football
13	junior	score	tennis
14	consecutive	nfl	golf
15	squad	beat	all-ireland
16	competition	fifth	world
17	league	finish	rugby
18	conference	fourth	trophy
19	title	all-time	soccer

guitar

	window	sentence	dependency
0	bass	drum	drum
1	drum	bass	bass
2	piano	vocal	piano
3	instrument	keyboard	vocal
4	vocal	acoustic	keyboard
5	keyboard	guitarist	cello
6	solo	band	flute
7	percussion	rhythm	clarinet
8	flute	drummer	violin
9	acoustic	instrument	saxophone
10	saxophone	percussion	trumpet

11	vocalist	musician	instrument
12	rhythm	vocalist	synthesizer
13	violin	saxophone	percussion
14	guitarist	solo	viola
15	trumpet	backing	tenor
16	string	piano	harp
17	sing	tenor	whistle
18	lyric	trio	horn
19	ensemble	flute	string

piano

	window	sentence	dependency
0	violin	violin	clarinet
1	guitar	flute	violin
2	flute	cello	viola
3	solo	sonata	cello
4	cello	bass	guitar
5	concerto	percussion	flute
6	bass	concerto	bass
7	percussion	trumpet	keyboard
8	trumpet	saxophone	trumpet
9	keyboard	quartet	horn
10	saxophone	tenor	percussion
11	op	composition	saxophone
12	instrument	guitar	harp
13	drum	instrument	instrument
14	string	viola	drum
15	viola	soloist	vocal
16	vocal	orchestra	dance
17	sonata	string	orchestra
18	orchestra	op	composition
19	acoustic	trio	organ

First order top 20 results (Window Sentence Dependency)

_____

car

	window	sentence	dependency
0	touring	car	compmod_P_parking
1	accident	parking	compmod_P_wash
2	parking	driver	compmod_P_accident
3	crash	racing	conj_D_truck
4	wash	lap	amod_D_armoured
5	racing	motor	compmod_D_concept
6	driver	accident	compmod_D_f1
7	truck	truck	conj_D_car
8	ferry	formula	dobj_P_race
9	bomb	crash	compmod_P_crash
10	race	race	compmod_P_bomb
11	motor	ford	conj_P_car
12	drive	drive	amod_D_armored
13	cable	wheel	compmod_P_ferry
14	manufacturer	passenger	partmod_D_race
15	formula	vehicle	amod_D_race
16	passenger	stock	dobj_P_drive
17	stock	engine	partmod_D_drive
18	steal	sport	compmod_D_formula
19	wheel	safety	compmod_P_racing

bus

	window	sentence	dependency
0	interchange	bus	conj_P_tram
1	rout	transit	conj_P_subway
2	terminal	commuter	conj_D_tram
3	depot	taxi	compmod_P_stop
4	taxi	interchange	compmod_D_serial
5	tram	tram	nsubj_P_rout
6	route	terminal	conj_D_taxi
7	shuttle	depot	compmod_P_terminus
8	operator	route	nsubj_P_route
9	parking	metro	compmod_P_interchange
10	shelter	transport	conj_D_bus
11	terminus	terminus	conj_P_bus
12	stop	passenger	conj_D_rail
13	bus	truck	compmod_P_shelter
14	service	operator	compmod_P_route
15	transit	stop	compmod_P_terminal
16	rapid	transportation	conj_P_rail
17	station	service	compmod_P_stand
18	lane	connect	compmod_P_passenger

| | | 19 | metro | rail | compmod_P_depot |

hospital

	window	sentence	dependency
0	psychiatric	psychiatric	amod_D_teaching
1	bed	hospital	appos_D_hospital
2	clinic	bombay	compmod_P_admission
3	sick	clinic	compmod_D_mercy
4	hospital	bed	amod_D_psychiatric
5	teaching	medical	compmod_P_bed
6	rush	surgery	conj_P_hospital
7	medical	patient	conj_D_hospital
8	emergency	tamil	compmod_P_stay
9	memorial	teaching	conj_D_clinic
10	recover	care	poss_D_christ
11	patient	health	appos_P_hospital
12	doctor	emergency	compmod_D_hopkins
13	care	cancer	compmod_D_teaching
14	injury	medicine	nsubj_P_treat
15	admit	doctor	compmod_D_memorial
16	treat	facility	poss_D_mary
17	dr	memorial	compmod_D_eye
18	nearby	treatment	compmod_D_general
19	stay	trust	compmod_D_presbyterian

hotel

	window	sentence	dependency
0	casino	hotel	conj_D_casino
1	hilton	luxury	appos_D_hotel
2	hotel	resort	compmod_P_lobby
3	luxury	casino	appos_P_hotel
4	restaurant	plaza	compmod_D_hilton
5	resort	restaurant	compmod_D_plaza
6	lobby	inn	conj_D_resort
7	plaza	room	compmod_D_neutral
8	room	tourist	compmod_D_savoy
9	inn	guest	conj_D_hotel
10	chain	chain	compmod_D_luxury
11	tourism	shop	conj_D_restaurant
12	lodge	palace	compmod_D_resort
13	guest	stay	conj_P_hotel
14	stay	purchase	compmod_P_casino
15	bar	tower	compmod_D_palace
16	downtown	owner	compmod_P_room
17	palace	bar	compmod_D_windsor
18	grand	grand	compmod_P_guest
19	historic	historic	compmod_D_milk

gun

	window	sentence	dependency
0	anti-aircraft	anti-aircraft	compmod_D_anti-aircraft
1	turret	turret	amod_D_rifle
2	machine	gun	compmod_D_machine
3	ammunition	armament	compmod_P_barrel
4	mortar	ammunition	dobj_P_aim
5	barrel	machine	compmod_D_deck
6	mm	barrel	amod_D_mounted
7	battery	rifle	conj_D_ammunition
8	gun	battery	conj_P_rifle
9	rifle	artillery	dobj_P_man
10	n	deck	conj_D_mortar
11	mount	fire	compmod_D_mm
12	shield	weapon	conj_D_gun
13	fire	mount	nsubjpass_P_mount
14	artillery	tank	conj_P_gun
15	deck	assault	dobj_P_jump
16	cannon	rose	compmod_P_nest
17	shot	shot	amod_D_anti-aircraft
18	instal	heavy	rcmod_D_fire
19	shoot	shoot	conj_P_sword

bomb

	window	sentence	dependency
0	explode	explode	partmod_P_suicide
1	atomic	luftwaffe	amod_D_pipe
2	disposal	atomic	nsubj_P_explode
3	explosion	bomb	nsubj_D_luftwaffe
4	blast	bombing	compmod_P_explode
5	luftwaffe	ira	compmod_P_disposal
6	raid	explosion	compmod_P_blast
7	hydrogen	explosive	partmod_D_plant
8	explosive	bomber	dobj_D_raid
9	ira	injure	compmod_P_explosion
10	sink	hydrogen	compmod_D_parcel
11	drop	raid	amod_D_atomic
12	shelter	nuclear	nsubj_D_nato
13	bomb	drop	compmod_D_hydrogen
14	terrorist	sink	compmod_D_cluster
15	suicide	target	dobj_P_plant
16	conventional	aircraft	conj_D_sink
17	raf	plane	compmod_D_ira
18	pipe	kill	nsubjpass_P_plant
19	injure	damage	nsubjpass_D_battery

horse

	window	sentence	dependency
0	thoroughbred	thoroughbred	conj_D_pony
1	mule	mounted	conj_D_mule
2	pony	bombay	compmod_D_harness
3	harness	horse	amod_D_thoroughbred
4	rid	jockey	compmod_D_trojan
5	trojan	rid	amod_D_bay
6	drawn	rider	dobj_P_ride
7	riding	trailer	conj_D_wagon
8	cart	riding	compmod_D_crazy
9	racing	breed	compmod_P_breeding
10	jockey	racing	conj_D_rider
11	rider	cattle	compmod_D_thoroughbred
12	crazy	ride	dobj_P_rid
13	cattle	cavalry	conj_P_bull
14	ride	stake	compmod_P_tram
15	artillery	regiment	compmod_P_cart
16	carriage	artillery	compmod_P_riding
17	trailer	brigade	compmod_P_racing
18	race	race	partmod_D_rid
19	brigade	wild	partmod_D_race

fox

	window	sentence	dependency
0	hound	fox	conj_D_hound
1	coyote	hound	conj_P_coyote
2	sac	affiliation	conj_P_hare
3	hunting	cbs	conj_P_sac
4	news	nbc	conj_P_abc
5	affiliate	abc	conj_P_beaver
6	sport	affiliate	compmod_P_hunting
7	rabbit	sport	compmod_P_ohio
8	reservation	news	compmod_D_gardner
9	affiliation	programming	conj_D_cat
10	fox	20th	compmod_P_net
11	20th	channel	compmod_D_twentieth
12	deer	network	compmod_P_squirrel
13	64	kid	compmod_D_kit
14	nbc	hunt	compmod_P_affiliation
15	abc	soccer	amod_D_gray
16	programming	switch	compmod_P_news
17	kid	entertainment	compmod_P_affiliate
18	network	morning	compmod_D_inspector
19	broadcasting	broadcast	compmod_P_sport

table

	window	sentence	dependency
0	tennis	table	compmod_D_periodic
1	picnic	picnic	compmod_D_following
2	periodic	periodic	compmod_P_tennis
3	table	tennis	amod_D_picnic
4	content	gospel	compmod_D_picnic
5	column	row	compmod_P_salt
6	chair	column	compmod_D_medal
7	following	database	amod_D_format
8	salt	following	amod_D_periodic
9	database	content	compmod_D_altar
10	row	element	compmod_D_aggregate
11	update	medal	conj_P_table
12	list	format	compmod_D_summary
13	medal	template	amod_D_rotary
14	round	finish	compmod_D_truth
15	mountain	key	compmod_D_breakfast
16	entry	entry	conj_P_chair
17	indicate	round	compmod_D_html
18	format	contain	compmod_D_column
19	finish	list	conj_D_table

bowl

	window	sentence	dependency
0	right-handed	bowl	rcmod_P_batsman
1	super	super	amod_P_alley
2	batsman	right-handed	dobj_D_pace
3	alley	quiz	conj_D_bowl
4	quiz	steelers	compmod_D_quiz
5	pro	batsman	conj_P_bowl
6	subdivision	bowler	compmod_D_insight
7	rose	wicket	compmod_D_rose
8	mvp	rose	compmod_D_bermuda
9	bowl	pro	compmod_D_pro
10	bowler	sugar	compmod_D_sugar
11	lawn	nfl	compmod_P_subdivision
12	sugar	pace	rcmod_P_ball
13	orange	orange	compmod_D_super
14	wicket	cotton	compmod_D_compass
15	cotton	inning	compmod_D_toilet
16	insight	consecutive	compmod_P_barrow
17	dust	touchdown	compmod_D_orange
18	inning	ball	compmod_D_lawn
19	consecutive	yard	dobj_D_green

guitar

	window	sentence	dependency

0	bass	bass	appos_P_vocal
1	acoustic	guitar	appos_D_vocal
2	amplifier	percussion	conj_D_amplifier
3	vocal	acoustic	compmod_P_flourish
4	keyboard	vocal	conj_D_vocal
5	rhythm	drum	conj_P_vocal
6	synthesizer	keyboard	conj_D_keyboard
7	percussion	amplifier	conj_D_bass
8	electric	saxophone	appos_D_bass
9	saxophone	rhythm	compmod_D_bass
10	drum	line-up	conj_D_percussion
11	pedal	backing	compmod_D_rhythm
12	solo	electric	compmod_P_solo
13	backing	gibson	conj_D_synthesizer
14	tenor	piano	amod_D_acoustic
15	chord	violin	conj_D_drum
16	gibson	string	compmod_P_intro
17	piano	guitarist	conj_D_guitar
18	guitar	solo	conj_D_saxophone
19	hero	melody	conj_D_trumpet

piano

	window	sentence	dependency
0	concerto	cello	conj_P_viola
1	sonata	sonata	compmod_P_opus
2	op	concerto	conj_P_cello
3	violin	viola	amod_D_aid
4	viola	clarinet	conj_P_violin
5	cello	violin	compmod_P_concerto
6	recital	opus	conj_D_op
7	clarinet	op	compmod_P_sonata
8	harp	piano	compmod_P_recital
9	lesson	saxophone	conj_P_harp
10	synthesizer	trumpet	compmod_P_cum
11	saxophone	flute	conj_D_violin
12	trumpet	trio	conj_P_clarinet
13	trio	pianist	conj_P_appreciation
14	flute	lesson	compmod_P_lesson
15	solo	keyboard	conj_P_saxophone
16	percussion	guitar	conj_P_trumpet
17	duet	solo	conj_P_voice
18	organ	orchestra	conj_D_flute
19	piano	drum	compmod_D_ph

Appendix 2 Word2Vec Comparison Tables

Word2Vec Comparison Tables

The format is the same as described for items 2 and 3 with the following column order:

Bow5 Deps

Second order top 20 results (Bow5 Deps)

_____

car

	bow5	deps
0	cars	truck
1	truck	suv
2	automobile	vehicle
3	vehicle	minivan
4	motorbike	cars
5	motorcycle	speedboat
6	driver	racecar
7	minivan	automobile
8	suv	motorcar
9	lorry	jeep
10	motorcar	limousine
11	mid-engined	minibus
12	limousine	lorry
13	front-engined	limo
14	moped	motorcycle
15	motorhome	bike
16	mercedes-benz	motorhome
17	bike	taxicab
18	rear-engined	roadster
19	three-wheeled	wagon

bus

	bow5	deps
0	buses	minibus
1	tram	tram
2	metrobus	buses
3	intercity	jeepney
4	busses	taxicab
5	fixed-route	motorcoach
6	minibus	taxi
7	inter-city	trolleybus
8	ksrtc	lorry
9	commuter	truck
10	apsrtc	metrobus
11	msrtc	streetcar
12	inter-urban	busses
13	dial-a-ride	ferryboat
14	mini-bus	trolley
15	light-rail	tramcar
16	rail	railcar

```
17      transit          railmotor
18    trolleybus   intercityexpress
19   limited-stop              train
```

```
                    hospital
                bow5              deps
0              clinic        sanatorium
1           hospitals           hospice
2            infirmary        sanitorium
3             hospice         hospitals
4             lying-in        sanitarium
5           dispensary            clinic
6           polyclinic          infirmary
7           sanatorium         polyclinic
8         convalescent         dispensary
9               mulago          orphanage
10          addenbrooke         poorhouse
11             bethlem          almshouse
12          psychiatric         workhouse
13             maudsley         institutet
14              siriraj        leprosarium
15           sanitarium    rikshospitalet
16           in-patient          heliport
17            incurables             gaol
18           orthopaedic       guesthouse
19             westmead        motherhouse
```

```
                    hotel
                bow5              deps
0                motel             motel
1           restaurant            hotels
2           doubletree            casino
3             sheraton        restaurant
4               hotels               inn
5         ritz-carlton        guesthouse
6              sofitel            tavern
7               westin              cafe
8               ramada      ritz-carlton
9               casino          nightclub
10           kempinski         travelodge
11             mansion           pizzeria
12                 inn          roadhouse
13                cafe      boardinghouse
14              tavern              café
15          apartments             condo
16            boutique           brewpub
17           nightclub          sheraton
18             marriott        steakhouse
19           travelodge         brasserie
```

```
                     gun
                bow5              deps
0                 guns              guns
1               cannon           handgun
```

	bow5	deps
2	howitzer	machinegun
3	sub-machine	howitzer
4	flamethrower	pistol
5	belt-fed	rifle
6	37mm	shotgun
7	smoothbore	firearm
8	pistol	cannon
9	shkas	musket
10	105mm	crossbow
11	40mm	autocannon
12	gatling	phaser
13	recoilless	flamethrower
14	76mm	revolver
15	3-inch	carbine
16	rifle	machine-gun
17	88mm	weapon
18	large-caliber	carronade
19	autocannons	pounder

bomb

	bow5	deps
0	bombs	bombs
1	detonated	firebomb
2	detonates	landmine
3	detonate	car-bomb
4	booby-trap	grenade
5	detonating	torpedo
6	firebomb	ied
7	car-bomb	warhead
8	exploded	bomber
9	detonation	bomblets
10	warhead	missile
11	500-pound	nuke
12	b61	detonator
13	laser-guided	booby-trap
14	blast	kamikaze
15	explosives	munition
16	detonations	explosives
17	landmine	machinegun
18	tallboy	a-bomb
19	thermonuclear	firebombs

horse

	bow5	deps
0	horses	horses
1	standardbred	goat
2	saddlebred	dog
3	gelding	stallion
4	thoroughbred	mule
5	stallion	bronc
6	racehorses	cow
7	dog	unicycle
8	riderless	greyhound
9	gaited	bareback

	bronc	camel
10	bronc	camel
11	percheron	appaloosa
12	pony	saddlebred
13	trotting	colt
14	harness	zebu
15	chariot	donkey
16	appaloosa	sidesaddle
17	sulky	racehorse
18	racehorse	elephant
19	greyhound	pony

fox

	bow5	deps
0	abc	daystar
1	cbs	nbc
2	nbc	byutv
3	wxyz-tv	kron
4	msnbc	wolf
5	ctv	cbs-tv
6	wsvn	familynet
7	familynet	abc
8	wttg	wccb
9	wjbk	oln
10	wfxt	wjar
11	espn	hdnet
12	wofl	telefutura
13	cnn	woodchuck
14	oln	nbc-tv
15	blitzer	soapnet
16	nesn	cinemax
17	wesh	wdiv
18	wb	mundofox
19	wgn-tv	coyote

table

	bow5	deps
0	tables	tables
1	sortable	leaderboard
2	wikitable	sideboard
3	look-up	chessboard
4	foosball	textbox
5	toc	taskbar
6	bulleted	gameboard
7	ping-pong	worksheet
8	billiard	tray
9	table-tennis	viewport
10	textbox	dais
11	tray	flowchart
12	lookup	playfield
13	wikitables	mantelpiece
14	brackets	stepladder
15	header	cladogram
16	footer	letterbox
17	tabular	windowsill

	menu	bookcase
18	menu	bookcase
19	carom	wikitable

	bowl	
	bow5	deps
0	xliii	bowls
1	xlii	superbowl
2	xliv	arenabowl
3	bowls	wcws
4	xlvi	wnit
5	tostitos	nlcs
6	xli	arenacup
7	xxxviii	postseason
8	xlv	nit
9	xxxv	xliii
10	xxxix	llws
11	xxxvii	beanpot
12	xlvii	xlv
13	xxxvi	triplemanía
14	xxxiv	alcs
15	bluebonnet	nlds
16	gator	cup
17	xxviii	kvalserien
18	xxxii	tourney
19	xxxi	cws

	guitar	
	bow5	deps
0	harmonica	saxophone
1	mandolin	bass
2	bass	mandolin
3	drums	harmonica
4	guitars	accordion
5	keyboards	trombone
6	accordion	violin
7	banjo	banjo
8	saxophone	guitars
9	12-string	cello
10	ukulele	piano
11	trombone	vibraphone
12	fiddle	sax
13	autoharp	trumpet
14	melodica	autoharp
15	percussion	clarinet
16	vibraphone	sitar
17	tambourine	fiddle
18	vocals	drums
19	fretless	marimba

	piano	
	bow5	deps
0	violin	violin
1	cello	cello
2	harpsichord	harpsichord

```
3     clarinet    saxophone
4        viola     clarinet
5        flute       guitar
6      bassoon     trombone
7  violoncello     mandolin
8         oboe    vibraphone
9     concerto      marimba
10   saxophone    accordion
11   accordion   pianoforte
12        harp      bassoon
13    trombone   fortepiano
14     sonatas  violoncello
15     trumpet      trumpet
16    mandolin    harmonica
17  pianoforte     clavinet
18  vibraphone   clavichord
19    concertos    euphonium
```

First order top 10 results (Bow5 Deps)

_____

car
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
accident poss_waltrip
****
bus
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 dobjI_onboard
transit amod_20-bit
****

hospital
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
bellevue compmod_monklands
****

hotel
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
marriott adpmod:onI_renovations
****

gun
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
emplacements compmod_varmint
****

bomb
bombs adpmod:on_pentagon
detonated compmodI_splashes
detonates adpmod:byI_unharmed
detonate adpmod:withI_rearmed
atomic compmod_single-car

exploded compmod_m1918
detonating compmod_phishing
bomber compmodl_fellatio
bombing rcmod_bursts
blast compmod_decapitation
****
horse
horses nsubjpassl_spooked
standardbred compmodl_drovers
thoroughbred dobjl_shoe
trotting adpmod:byl_commute
riding rcmod_roams
racing poss_quixote
ridden compmod_ch-53e
galloping appos_pony
bred compmod_poitevin
gelding adpmod:ofl_hilt
****
fox
vulpes appos_canadensis
news compmodl_subchannels
movietone conj_mattel
owned-and-operated compmod_jacki
nbc conjl_raimi
cbs conjl_corgan
terrier conjl_leda
affiliate conj_bluebird
cnn conj_nolte
msnbc conjl_browder
****
table
sortable adpmod:offl_knocks
tables adpmod:onl_seventh
tennis amod_ten-foot
lookup compmod_abydos
billiards amod_25-metre
foosball adpmod:intol_face
periodic adpmod:belowl_added
billiard adpmod:throughl_throw
toc compmod_three-judge
look-up adpmod:ontol_projects
****

bowl
xli conjI_preseason
bowls adpmod:against_oklahoma
xliv adpmod:inI_2-1
xxxviii conj_play-offs
xliii compmod_pan-pacific
xlii adpmod:forI_stadiums
xlvi adpmod:atI_clinched
xlv adpmod:intoI_win
super amod_25-metre
xxxix conjI_regionals
****

guitar
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:ofI_virtuoso
banjo adpmod:onI_foster
vocals amod_necked
****

piano
violin amod_end-blown
sonata conjI_bongos
cello adpmod:onI_harris
concerto adpmod:ofI_virtuoso
op adpmod:onI_thompson
harpsichord conj_back-up
concertos compmod_kreutzer
viola compmod_vomeronasal
violoncello adpmod:onI_supplemented
sonatas appos_mandolin
****