Yehoshua Stern 314963927

Eyal Cohen 207947086

**Ass 3 NLP report**

**1) Summarize statistics:**

1. Thresholds: at the top of the file we have the function words list, we used this list when computing the window words and we ignored them, actually the as much as the code concerns these words does not exist in the sentence.

Another threshold we decided to do is to not collect Data that relate to function words, which decreased the running time by 70% (from 20 minutes total running time to 6 minutes).

1. Number of words which passed the 75-frequency threshold: 10057 words
2. Number of features considered in our computation for each co-occurrence type: 100 (can be seen when we do the efficient algorithm, in the first for loop:

for att in u\_att\_counter\_tuple[u].most\_common(100):

**2)Qualitive conclusion:**

Full list in the end of the report Appendix - Top 20 similars co-occurrence (top20.csv and top20.txt). we have been asked to add the txt file, so we added it even though it's easier to read the csv file.

Comparative characteristics between there types of similarity: a good example to look at is the word bomb, the window type gave us words like bombing, explosion, explode, the sentence based features data gives us words like damage, evacuate, destroy and the dependency based data gives us rocket, weapon, missile.

You can see that the words of the window type are words that are closer in relation to bomb line explode (the bomb exploded etc.…) and when we look at the sentence its words like evacuate and destroy which are related to bomb but in pretty far relation, and when we see the dependency based data we get words that are types of bombs (missile, rocket) or describe bomb (weapon, device)

The main characteristics we found was that for words in the same sentence the relation is clear but not always the first thing we would have thought of when we hear that exact word (another example is for bomb the word Luftwaffe which is a German aircraft from WWII that terrorized soldiers by doing dive bombing, as we said the relation is there but pretty far). As for window-based data, we can see that it is easier to see the connection between the original word and the closest words we get (in table we get periotic and tennis which are kinds of tables), and last for the dependency connections we get words of object that bejhave the same like the first word (like in horse we can see motorcycle and bicycle which are ridable objects like horse).

**3) Top contexts attributes:**

Full list in the end of the report, (Top 20 1st order similar co-occurrence (top20\_1st\_order.csv)).

As we can see in the lists, it's pretty easy to notice the difference between the 1st order similarities and 2nd order similarities, take the word car as an example, in the 1st order similarity we see words like touring, racing, armored which are features of car (touring car, racing car, armored car etc.…) but when we look at the 2nd order similarity we see words like truck, vehicle, motorcycle which are more related to car by being in the same category as car by being a transport machine just like car.

It's even more noticeable when looking at the word hospital, 1st order top word is psychiatric and when we look at 2nd order we see that the first word is clinic which is almost the same as hospital.

**4) manual judgments:**

**MAP:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Window | | Sentence | | Dependency | |
| Semantic | Topical | Semantic | Topical | Semantic | Topical |
| 0.55 | 0.9 | 0.5 | 1 | 0.65 | 0.85 |

Full lists at the end of the report (top20\_car\_semantically/topical, top20\_piano\_semantically/topical) that’s the judgment lists for the similarity output for the co-occurrence mission):

You can see that the results are as expected, the highest score in the semantical (words that behave the same as the original word) type we got in the dependency co-occurrence type, the best score for the topical (which is reasonable because it's words that had the same words around them like the original word) we have at the sentence type and best of both worlds we have when we look at the window type.

The main conclusion we see here is you can use different features to get similar words according to your needs, if you need words that have better semantic relation to your word you might be better using dependency feature collection, but if you need words that have topical connection to your word its better for you to choose sentence based data collection, if you need both you can use window based collection.

**5) Implementation brief description:**

1. PMI: we used the basic PMI estimation technique using the counters to estimate probability and we didn’t save a matrix of the PMI values, we calculated the values in place in the algorithm, for each of the co-occurrence types we had different PMI function (log\_PMI\_u\_att\_<windows/words/tuple>) which used different counters to count the attributes.
2. Implementation of the efficient algorithm: almost exactly like the pseudo-code we saw in class with a minor improvement that we do not save the matrix, we use counters to mimic the rows and columns in the matrix, and every time we calculate the needed PMI value to save space, also we compute the PMi for the word every row only once.

At the end we sort the lists we need to print and return the top 20 words and contexts for each word.

After we ignored completely the function words the running time became reasonable so we didn’t feel there is need for another improvement (about 6 minutes for the whole corpus)

The efficient function names are: most\_similar\_efficient\_att\_<tuple/words/windows>

The function words list is at the top of the code, named function\_words.

When counting words to the unigram counter you can see there is a list, we filter the words that didn’t pass the 75 words threshold, plus we everyu time in the main loop check whether the word is in the words\_set(words seen above 75 times) and if its not in the function words set.

**6) word2vec results:**

Python related files are:

1. vec\_similarity.py – the running file with the algorithm for comparison and computing cosine distance for each vector after reading vectors.
2. vecs\_asarray.py – instead of searching the vectors of each word we already got them from the files and saved them in this file to save time in each run.

**2)Qualitive conclusion(word2vec):**

As in the co-occurrence code we can see almos the same behavior, the words in the dependency were words that serve the same role in the sentence as the original goal ("I drive my car" can be "I drive my lory") and when it comes to bow5 we see more of words that relate to the original word but not necessarily can replace it in the sentence.

Full list an the appendinx at the end under the title Top20\_word2vec(top20\_word2vec.txt top20\_word2vec.csv)

**3) Top contexts attributes(word2vec):**

For word2vec we found that its harder to see the difference between 1st order and 2nd order, as we see in the list in bow5 the word mid-engine appear in both lists, its easier to see why its in the 1st order (mid-engine car).

Like in the co-occurrence explanation we see that we get in the contexts words that relate to the main word by modifying it or addressing it, and when we look at the 2nd similarity type we get words that have the same behavior in the sentence like the original word.

**4) manual judgments:**

When it comes to word2vec we had much better results in the manual judgment then we had in the co-occurrence judgment

**MAP:**

|  |  |  |  |
| --- | --- | --- | --- |
| Window | | Dependency | |
| Semantic | Topical | Semantic | Topical |
| 0.8 | 1 | 0.975 | 1 |

As we see in this word2vec test we get the same results, dependency get better semantic score but this time it doesn’t get worse when it comes to topical score.

One thing to keep in mind about the scores of word2vec and co-occurrence, the word2vec vectors are the product of learning the full Wikipedia corpus, our output is only after learning a sample from the corpus, and that’s why the scores we got in co-occurrence judgement are lower overall then the scores we get in the word2vec judgement.

**Appendix:**

**Judgment lists:**

**Top20\_car\_semantically(top20\_car\_semantic.csv):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | window |  | sentence |  | dependency |
| 1 | truck | 1 | toyota | 1 | vehicle |
| 1 | vehicle | 0 | driver | 1 | truck |
| 1 | motorcycle | 0 | chassis | 1 | motorcycle |
| 0 | motor | 0 | racing | 1 | automobile |
| 0 | racing | 1 | chevrolet | 1 | train |
| 1 | automobile | 1 | nascar | 1 | bus |
| 1 | bicycle | 1 | ferrari | 1 | boat |
| 0 | driver | 1 | bmw | 0 | motor |
| 1 | bus | 1 | truck | 1 | aircraft |
| 0 | rail | 0 | axle | 1 | bicycle |
| 1 | boat | 0 | motor | 0 | engine |
| 1 | train | 0 | gt | 0 | racing |
| 0 | race | 0 | v8 | 0 | driver |
| 1 | ford | 0 | wheel | 1 | locomotive |
| 1 | bike | 1 | f1 | 1 | horse |
| 0 | crash | 0 | formula | 1 | ford |
| 1 | auto | 0 | lap | 1 | van |
| 1 | freight | 1 | ford | 0 | formula |
| 0 | passenger | 1 | porsche | 0 | ship |
| 1 | formula | 1 | auto | 1 | auto |

**Top20\_car\_topical(top20\_car\_topically.csv):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | window |  | sentence |  | dependency |
| 1 | truck | 1 | toyota | 1 | vehicle |
| 1 | vehicle | 1 | driver | 1 | truck |
| 1 | motorcycle | 1 | chassis | 1 | motorcycle |
| 1 | motor | 1 | racing | 1 | automobile |
| 1 | racing | 1 | chevrolet | 1 | train |
| 1 | automobile | 1 | nascar | 1 | bus |
| 1 | bicycle | 1 | ferrari | 1 | boat |
| 1 | driver | 1 | bmw | 1 | motor |
| 1 | bus | 1 | truck | 1 | aircraft |
| 0 | rail | 1 | axle | 1 | bicycle |
| 1 | boat | 1 | motor | 1 | engine |
| 1 | train | 1 | gt | 1 | racing |
| 1 | race | 1 | v8 | 1 | driver |
| 1 | ford | 1 | wheel | 0 | locomotive |
| 1 | bike | 1 | f1 | 0 | horse |
| 1 | crash | 1 | formula | 1 | ford |
| 1 | auto | 1 | lap | 1 | van |
| 0 | freight | 1 | ford | 1 | formula |
| 1 | passenger | 1 | porsche | 0 | ship |
| 1 | formula | 1 | auto | 1 | auto |

**top20\_piano\_semantic(top20\_piano\_semantic.csv):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | window |  | sentence |  | dependency |
| 1 | violin | 1 | violin | 1 | violin |
| 1 | cello | 1 | cello | 1 | cello |
| 1 | clarinet | 1 | clarinet | 1 | flute |
| 0 | solo | 1 | flute | 1 | guitar |
| 0 | sonata | 0 | concerto | 1 | clarinet |
| 1 | flute | 0 | sonata | 0 | bass |
| 0 | concerto | 1 | trumpet | 1 | viola |
| 0 | op | 1 | viola | 0 | keyboard |
| 0 | keyboard | 0 | percussion | 0 | horn |
| 0 | string | 1 | saxophone | 1 | drum |
| 0 | bass | 0 | op | 1 | organ |
| 1 | trumpet | 0 | orchestra | 0 | percussion |
| 1 | guitar | 1 | harp | 0 | orchestra |
| 1 | saxophone | 0 | soloist | 1 | instrument |
| 1 | viola | 0 | quartet | 1 | trumpet |
| 1 | organ | 0 | ensemble | 0 | music |
| 0 | horn | 1 | tenor | 1 | saxophone |
| 0 | lesson | 1 | instrument | 1 | harp |
| 0 | classical | 0 | bass | 0 | solo |
| 0 | recital | 0 | soprano | 0 | choir |

**top20\_piano\_topically(top20\_piano\_topically.csv):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | window |  | sentence |  | dependency |
| 1 | violin | 1 | violin | 1 | violin |
| 1 | cello | 1 | cello | 1 | cello |
| 1 | clarinet | 1 | clarinet | 1 | flute |
| 1 | solo | 1 | flute | 1 | guitar |
| 1 | sonata | 1 | concerto | 1 | clarinet |
| 1 | flute | 1 | sonata | 1 | bass |
| 1 | concerto | 1 | trumpet | 1 | viola |
| 1 | op | 1 | viola | 0 | keyboard |
| 0 | keyboard | 1 | percussion | 0 | horn |
| 1 | string | 1 | saxophone | 1 | drum |
| 1 | bass | 1 | op | 1 | organ |
| 1 | trumpet | 1 | orchestra | 1 | percussion |
| 1 | guitar | 1 | harp | 1 | orchestra |
| 1 | saxophone | 1 | soloist | 1 | instrument |
| 1 | viola | 1 | quartet | 1 | trumpet |
| 1 | organ | 1 | ensemble | 1 | music |
| 0 | horn | 1 | tenor | 1 | saxophone |
| 1 | lesson | 1 | instrument | 1 | harp |
| 1 | classical | 1 | bass | 1 | solo |
| 1 | recital | 1 | soprano | 0 | choir |

**word2vec\_top20\_car\_semantic.csv:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | bow5 |  | dependency |
| 1 | cars | 1 | truck |
| 1 | truck | 1 | suv |
| 1 | automobile | 1 | vehicle |
| 1 | vehicle | 1 | minivan |
| 1 | motorbike | 1 | cars |
| 1 | motorcycle | 1 | speedboat |
| 0 | driver | 1 | racecar |
| 1 | minivan | 1 | automobile |
| 1 | suv | 1 | motorcar |
| 1 | lorry | 1 | jeep |
| 1 | motorcar | 1 | limousine |
| 0 | mid-engined | 1 | minibus |
| 1 | limousine | 1 | lorry |
| 0 | front-engined | 1 | limo |
| 1 | moped | 1 | motorcycle |
| 1 | motorhome | 1 | bike |
| 0 | mercedes-benz | 1 | motorhome |
| 1 | bike | 1 | taxicab |
| 0 | rear-engined | 1 | roadster |
| 0 | three-wheeled | 0 | wagon |

**word2vec\_top20\_car\_topical.csv:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | bow5 |  | dependency |
| 1 | cars | 1 | truck |
| 1 | truck | 1 | suv |
| 1 | automobile | 1 | vehicle |
| 1 | vehicle | 1 | minivan |
| 1 | motorbike | 1 | cars |
| 1 | motorcycle | 1 | speedboat |
| 1 | driver | 1 | racecar |
| 1 | minivan | 1 | automobile |
| 1 | suv | 1 | motorcar |
| 1 | lorry | 1 | jeep |
| 1 | motorcar | 1 | limousine |
| 1 | mid-engined | 1 | minibus |
| 1 | limousine | 1 | lorry |
| 1 | front-engined | 1 | limo |
| 1 | moped | 1 | motorcycle |
| 1 | motorhome | 1 | bike |
| 1 | mercedes-benz | 1 | motorhome |
| 1 | bike | 1 | taxicab |
| 1 | rear-engined | 1 | roadster |
| 1 | three-wheeled | 1 | wagon |

**word2vec\_top20\_piano\_semantic.csv:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | bow5 |  | dependency |
| 1 | violin | 1 | violin |
| 1 | cello | 1 | cello |
| 1 | harpsichord | 1 | harpsichord |
| 1 | clarinet | 1 | saxophone |
| 1 | viola | 1 | clarinet |
| 1 | flute | 1 | guitar |
| 1 | bassoon | 1 | trombone |
| 1 | violoncello | 1 | mandolin |
| 1 | oboe | 1 | vibraphone |
| 1 | concerto | 1 | marimba |
| 1 | saxophone | 1 | accordion |
| 1 | accordion | 1 | pianoforte |
| 1 | harp | 1 | bassoon |
| 1 | trombone | 1 | fortepiano |
| 0 | sonatas | 1 | violoncello |
| 1 | trumpet | 1 | trumpet |
| 1 | mandolin | 1 | harmonica |
| 1 | pianoforte | 1 | clavinet |
| 1 | vibraphone | 1 | clavichord |
| 0 | concertos | 1 | euphonium |

**word2vec\_top20\_piano\_topical.csv:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | bow5 |  | dependency |
| 1 | violin | 1 | violin |
| 1 | cello | 1 | cello |
| 1 | harpsichord | 1 | harpsichord |
| 1 | clarinet | 1 | saxophone |
| 1 | viola | 1 | clarinet |
| 1 | flute | 1 | guitar |
| 1 | bassoon | 1 | trombone |
| 1 | violoncello | 1 | mandolin |
| 1 | oboe | 1 | vibraphone |
| 1 | concerto | 1 | marimba |
| 1 | saxophone | 1 | accordion |
| 1 | accordion | 1 | pianoforte |
| 1 | harp | 1 | bassoon |
| 1 | trombone | 1 | fortepiano |
| 1 | sonatas | 1 | violoncello |
| 1 | trumpet | 1 | trumpet |
| 1 | mandolin | 1 | harmonica |
| 1 | pianoforte | 1 | clavinet |
| 1 | vibraphone | 1 | clavichord |
| 1 | concertos | 1 | euphonium |

**Output lists 1st order and 2nd order:**

**Top 20 similars co-occurrence (top20.csv and top20.txt):**

|  |  |  |  |
| --- | --- | --- | --- |
| **word** | **window** | **sentence** | **dependency** |
| car |  |  |  |
|  | truck | toyota | vehicle |
|  | vehicle | driver | truck |
|  | motorcycle | chassis | motorcycle |
|  | motor | racing | automobile |
|  | racing | chevrolet | train |
|  | automobile | nascar | bus |
|  | bicycle | ferrari | boat |
|  | driver | bmw | motor |
|  | bus | truck | aircraft |
|  | rail | axle | bicycle |
|  | boat | motor | engine |
|  | train | gt | racing |
|  | race | v8 | driver |
|  | ford | wheel | locomotive |
|  | bike | f1 | horse |
|  | crash | formula | ford |
|  | auto | lap | van |
|  | freight | ford | formula |
|  | passenger | porsche | ship |
|  | formula | auto | auto |
| bus |  |  |  |
|  | commuter | commuter | rail |
|  | transit | transit | train |
|  | rail | rail | tram |
|  | tram | tram | ferry |
|  | passenger | passenger | car |
|  | operator | metro | operator |
|  | ferry | interchange | taxi |
|  | transport | terminal | road |
|  | metro | freight | transit |
|  | interchange | amtrak | railway |
|  | terminal | subway | vehicle |
|  | freight | terminus | passenger |
|  | train | transport | line |
|  | taxi | tramway | bicycle |
|  | shuttle | route | metro |
|  | connect | taxi | route |
|  | route | transportation | highway |
|  | truck | ferry | railroad |
|  | express | depot | freight |
|  | rout | hub | subway |
| hospital |  |  |  |
|  | clinic | clinic | clinic |
|  | psychiatric | medical | home |
|  | medical | nursing | school |
|  | nh | psychiatric | facility |
|  | care | surgical | hotel |
|  | nursing | healthcare | library |
|  | facility | rehabilitation | center |
|  | patient | surgery | camp |
|  | hopkins | psychiatry | museum |
|  | charity | care | ward |
|  | bed | health | prison |
|  | health | patient | college |
|  | memorial | psychiatrist | institution |
|  | mercy | nh | airport |
|  | college | acute | office |
|  | trust | medicine | store |
|  | presbyterian | clinical | centre |
|  | institution | dental | theatre |
|  | mental | diagnosis | house |
|  | teaching | pharmacy | laboratory |
| hotel |  |  |  |
|  | luxury | restaurant | restaurant |
|  | plaza | inn | store |
|  | casino | casino | resort |
|  | resort | resort | building |
|  | near | plaza | shop |
|  | restaurant | tourist | inn |
|  | dining | luxury | theater |
|  | shopping | shop | house |
|  | windsor | dining | palace |
|  | chain | renovate | apartment |
|  | room | attraction | park |
|  | nearby | accommodation | factory |
|  | inn | retail | theatre |
|  | riverside | tavern | station |
|  | hilton | apartment | casino |
|  | apartment | mall | home |
|  | savoy | shopping | tower |
|  | palace | cater | facility |
|  | grand | pub | residence |
|  | theatre | demolish | hall |
| gun |  |  |  |
|  | cannon | launcher | cannon |
|  | anti-aircraft | anti-aircraft | artillery |
|  | artillery | turret | pistol |
|  | fire | ammunition | rifle |
|  | machine | cannon | weapon |
|  | tank | armament | mortar |
|  | rifle | rifle | shotgun |
|  | heavy | mortar | fire |
|  | mortar | grenade | battery |
|  | battery | artillery | camera |
|  | armament | projectile | engine |
|  | mm | battery | turret |
|  | enemy | tank | tube |
|  | turret | weapon | rocket |
|  | weapon | fire | unit |
|  | ammunition | mm | ammunition |
|  | gunner | armored | tank |
|  | radar | firing | vehicle |
|  | laser | battleship | ship |
|  | aa | gunner | machine |
| bomb |  |  |  |
|  | bombing | explode | bombing |
|  | explosion | bombing | torpedo |
|  | explode | luftwaffe | rocket |
|  | nuclear | bomber | attack |
|  | bomber | explosive | weapon |
|  | blast | explosion | shell |
|  | raid | raid | fire |
|  | attack | injure | device |
|  | disposal | blast | missile |
|  | tank | damage | explosion |
|  | atomic | tanker | car |
|  | destroy | evacuate | strike |
|  | aircraft | destroy | tank |
|  | plane | airfield | explosive |
|  | fighter | ira | grenade |
|  | fly | grenade | earthquake |
|  | luftwaffe | plane | bomber |
|  | during | aircraft | storm |
|  | allied | sink | aircraft |
|  | explosive | allied | crash |
| horse |  |  |  |
|  | thoroughbred | thoroughbred | dog |
|  | racing | jockey | motorcycle |
|  | brigade | pony | bicycle |
|  | regiment | racecourse | animal |
|  | cavalry | mounted | car |
|  | pony | breeder | bike |
|  | artillery | ride | pony |
|  | breed | rider | racing |
|  | motorcycle | rid | bull |
|  | harness | sire | volunteer |
|  | race | racing | thoroughbred |
|  | stable | stake | team |
|  | ride | handicap | truck |
|  | infantry | breed | goat |
|  | rider | cavalry | woman |
|  | camel | stud | cattle |
|  | guard | camel | man |
|  | bike | saddle | race |
|  | cattle | riding | athlete |
|  | jockey | mare | dragon |
| fox |  |  |  |
|  | nbc | nbc | abc |
|  | abc | cbs | nbc |
|  | cbs | abc | cbs |
|  | news | espn | television |
|  | affiliate | affiliation | bbc |
|  | network | programming | network |
|  | tv | cbc | espn |
|  | television | newscast | tv |
|  | cbc | affiliate | sport |
|  | programming | broadcasting | channel |
|  | broadcasting | channel | radio |
|  | channel | broadcast | cbc |
|  | radio | cnn | cable |
|  | sport | television | news |
|  | espn | pb | itv |
|  | pb | daytime | smith |
|  | cnn | network | shaw |
|  | matthew | hbo | cnn |
|  | anchor | news | cox |
|  | drama | weekday | entertainment |
| table |  |  |  |
|  | tennis | heading | list |
|  | periodic | html | column |
|  | picnic | picnic | page |
|  | below | volleyball | section |
|  | file | skiing | template |
|  | etc | header | category |
|  | badminton | row | screen |
|  | html | badminton | bottom |
|  | summarize | excel | discussion |
|  | database | infobox | room |
|  | pdf | softball | diagram |
|  | update | handball | infobox |
|  | rotary | gymnastics | article |
|  | swim | xml | bench |
|  | ten | subscription | row |
|  | altar | bottom | entry |
|  | dinner | playground | image |
|  | water | database | window |
|  | following | mediawiki | box |
|  | chess | bracket | magazine |
| bowl |  |  |  |
|  | super | batsman | super |
|  | bowler | bowler | cup |
|  | mvp | steelers | championship |
|  | quiz | right-handed | league |
|  | pro | mvp | tournament |
|  | rose | wicket | game |
|  | tournament | playoff | playoff |
|  | all-american | inning | football |
|  | lawn | consecutive | club |
|  | consecutive | nfl | team |
|  | golden | 49ers | trophy |
|  | compass | redskin | season |
|  | playoff | packer | final |
|  | orange | afc | conference |
|  | victory | super | match |
|  | ball | berth | division |
|  | bermuda | quarterback | nfl |
|  | classic | bowling | wrestling |
|  | cotton | all-time | baseball |
|  | championship | cowboy | square |
| guitar |  |  |  |
|  | bass | bass | drum |
|  | piano | percussion | bass |
|  | drum | acoustic | piano |
|  | keyboard | drum | keyboard |
|  | vocal | vocal | vocal |
|  | acoustic | keyboard | flute |
|  | solo | saxophone | instrument |
|  | violin | flute | violin |
|  | instrument | piano | trumpet |
|  | trumpet | guitarist | cello |
|  | pedal | rhythm | saxophone |
|  | playing | violin | organ |
|  | saxophone | instrument | clarinet |
|  | intro | bassist | string |
|  | flute | backing | horn |
|  | jazz | trumpet | solo |
|  | rhythm | cello | music |
|  | tenor | instrumentation | tenor |
|  | electric | synthesizer | percussion |
|  | string | clarinet | orchestra |
| piano |  |  |  |
|  | violin | violin | violin |
|  | cello | cello | cello |
|  | clarinet | clarinet | flute |
|  | solo | flute | guitar |
|  | sonata | concerto | clarinet |
|  | flute | sonata | bass |
|  | concerto | trumpet | viola |
|  | op | viola | keyboard |
|  | keyboard | percussion | horn |
|  | string | saxophone | drum |
|  | bass | op | organ |
|  | trumpet | orchestra | percussion |
|  | guitar | harp | orchestra |
|  | saxophone | soloist | instrument |
|  | viola | quartet | trumpet |
|  | organ | ensemble | music |
|  | horn | tenor | saxophone |
|  | lesson | instrument | harp |
|  | classical | bass | solo |
|  | recital | soprano | choir |

**Top 20 1st order similar co-occurrence (top20 1st\_order.csv):**

|  |  |  |  |
| --- | --- | --- | --- |
| **word** | **window** | **sentence** | **dependency** |
| car |  |  |  |
|  | touring | racing | ('dealership'; 'compmod'; 'TO-ME') |
|  | armoured | driver | ('parking'; 'compmod'; 'TO-ME') |
|  | accident | lap | ('wash'; 'compmod'; 'TO-ME') |
|  | armored | motor | ('accident'; 'compmod'; 'TO-ME') |
|  | rental | accident | ('truck'; 'conj'; 'FROM-ME') |
|  | luxury | truck | ('armoured'; 'amod'; 'FROM-ME') |
|  | crash | formula | ('concept'; 'compmod'; 'FROM-ME') |
|  | parking | crash | ('race'; 'dobj'; 'TO-ME') |
|  | wash | ford | ('crash'; 'compmod'; 'TO-ME') |
|  | racing | race | ('bomb'; 'compmod'; 'TO-ME') |
|  | funny | drive | ('hit'; 'adpobj'; 'TO-ME'; 'adpmod'; 'by') |
|  | cable | wheel | ('armored'; 'amod'; 'FROM-ME') |
|  | sprint | passenger | ('ferry'; 'compmod'; 'TO-ME') |
|  | freight | vehicle | ('race'; 'amod'; 'FROM-ME') |
|  | dealer | stock | ('drive'; 'dobj'; 'TO-ME') |
|  | stock | engine | ('formula'; 'compmod'; 'FROM-ME') |
|  | motor | sport | ('drive'; 'partmod'; 'FROM-ME') |
|  | ferry | speed | ('racing'; 'compmod'; 'TO-ME') |
|  | race | championship | ('luxury'; 'compmod'; 'FROM-ME') |
|  | driver | front | ('cable'; 'compmod'; 'FROM-ME') |
| bus |  |  |  |
|  | interchange | transit | ('tram'; 'conj'; 'TO-ME') |
|  | rout | interchange | ('ֳ‰ireann'; 'compmod'; 'TO-ME') |
|  | shuttle | terminal | ('subway'; 'conj'; 'TO-ME') |
|  | terminal | taxi | ('stop'; 'compmod'; 'TO-ME') |
|  | depot | depot | ('rout'; 'nsubj'; 'TO-ME') |
|  | route | route | ('serial'; 'compmod'; 'FROM-ME') |
|  | commuter | metro | ('terminus'; 'compmod'; 'TO-ME') |
|  | operator | transport | ('route'; 'nsubj'; 'TO-ME') |
|  | transit | terminus | ('interchange'; 'compmod'; 'TO-ME') |
|  | shelter | passenger | ('accessible'; 'adpobj'; 'TO-ME'; 'adpmod'; 'by') |
|  | stop | truck | ('rail'; 'conj'; 'FROM-ME') |
|  | service | operator | ('shelter'; 'compmod'; 'TO-ME') |
|  | serial | stop | ('route'; 'compmod'; 'TO-ME') |
|  | hybrid | transportation | ('terminal'; 'compmod'; 'TO-ME') |
|  | parking | service | ('rail'; 'conj'; 'TO-ME') |
|  | rapid | operate | ('stand'; 'compmod'; 'TO-ME') |
|  | terminus | connect | ('passenger'; 'compmod'; 'TO-ME') |
|  | station | rail | ('depot'; 'compmod'; 'TO-ME') |
|  | motor | station | ('transit'; 'compmod'; 'TO-ME') |
|  | driver | airport | ('board'; 'dobj'; 'TO-ME') |
| hospital |  |  |  |
|  | psychiatric | psychiatric | ('doctor'; 'adpobj'; 'TO-ME'; 'adpmod'; 'at') |
|  | bed | clinic | ('rush'; 'adpobj'; 'TO-ME'; 'adpmod'; 'to') |
|  | hopkins | bed | ('teaching'; 'amod'; 'FROM-ME') |
|  | nh | medical | ('psychiatric'; 'amod'; 'FROM-ME') |
|  | mercy | surgery | ('bed'; 'compmod'; 'TO-ME') |
|  | teaching | patient | ('stay'; 'compmod'; 'TO-ME') |
|  | acute | bombay | ('hopkins'; 'compmod'; 'FROM-ME') |
|  | mental | care | ('teaching'; 'compmod'; 'FROM-ME') |
|  | memorial | teaching | ('rush'; 'adpobj'; 'TO-ME') |
|  | presbyterian | cancer | ('transport'; 'adpobj'; 'TO-ME'; 'adpmod'; 'to') |
|  | admission | health | ('take'; 'adpobj'; 'TO-ME'; 'adpmod'; 'to') |
|  | rush | medicine | ('patient'; 'adpobj'; 'TO-ME') |
|  | general | emergency | ('memorial'; 'compmod'; 'FROM-ME') |
|  | christ | doctor | ('admit'; 'adpobj'; 'TO-ME'; 'adpmod'; 'to') |
|  | recover | dr | ('die'; 'adpobj'; 'TO-ME'; 'adpmod'; 'at') |
|  | charity | memorial | ('eye'; 'compmod'; 'FROM-ME') |
|  | massachusetts | facility | ('general'; 'compmod'; 'FROM-ME') |
|  | veteran | trust | ('visit'; 'adpobj'; 'TO-ME'; 'adpmod'; 'in') |
|  | naval | treatment | ('doctor'; 'adpobj'; 'TO-ME') |
|  | nearby | tamil | ('massachusetts'; 'compmod'; 'FROM-ME') |
| hotel |  |  |  |
|  | hilton | resort | ('casino'; 'conj'; 'FROM-ME') |
|  | luxury | luxury | ('lobby'; 'compmod'; 'TO-ME') |
|  | savoy | casino | ('hilton'; 'compmod'; 'FROM-ME') |
|  | plaza | plaza | ('plaza'; 'compmod'; 'FROM-ME') |
|  | dining | restaurant | ('resort'; 'conj'; 'FROM-ME') |
|  | lobby | inn | ('savoy'; 'compmod'; 'FROM-ME') |
|  | paramount | room | ('neutral'; 'compmod'; 'FROM-ME') |
|  | casino | tourist | ('luxury'; 'compmod'; 'FROM-ME') |
|  | room | chain | ('restaurant'; 'conj'; 'FROM-ME') |
|  | riverside | guest | ('resort'; 'compmod'; 'FROM-ME') |
|  | chain | shop | ('stay'; 'adpobj'; 'TO-ME'; 'adpmod'; 'at') |
|  | windsor | palace | ('palace'; 'compmod'; 'FROM-ME') |
|  | resort | purchase | ('room'; 'compmod'; 'TO-ME') |
|  | grand | stay | ('convert'; 'adpobj'; 'TO-ME'; 'adpmod'; 'into') |
|  | palace | tower | ('milk'; 'compmod'; 'FROM-ME') |
|  | expensive | owner | ('chain'; 'compmod'; 'TO-ME') |
|  | midland | bar | ('restaurant'; 'adpobj'; 'TO-ME') |
|  | milk | grand | ('chain'; 'adpobj'; 'TO-ME'; 'adpmod'; 'of') |
|  | ace | historic | ('shop'; 'conj'; 'TO-ME') |
|  | nearby | spring | ('purchase'; 'dobj'; 'TO-ME') |
| gun |  |  |  |
|  | anti-aircraft | anti-aircraft | ('navarone'; 'adpobj'; 'FROM-ME') |
|  | machine | turret | ('anti-aircraft'; 'compmod'; 'FROM-ME') |
|  | mm | armament | ('emplacement'; 'compmod'; 'TO-ME') |
|  | mounted | machine | ('rifle'; 'amod'; 'FROM-ME') |
|  | barrel | ammunition | ('machine'; 'compmod'; 'FROM-ME') |
|  | mk | barrel | ('battery'; 'adpobj'; 'TO-ME'; 'adpmod'; 'of') |
|  | deck | rifle | ('barrel'; 'compmod'; 'TO-ME') |
|  | turret | battery | ('arm'; 'adpobj'; 'TO-ME'; 'adpmod'; 'with') |
|  | bunker | artillery | ('aim'; 'dobj'; 'TO-ME') |
|  | n | deck | ('deck'; 'compmod'; 'FROM-ME') |
|  | armstrong | fire | ('mounted'; 'amod'; 'FROM-ME') |
|  | heavy | mount | ('ammunition'; 'conj'; 'FROM-ME') |
|  | artillery | weapon | ('man'; 'dobj'; 'TO-ME') |
|  | rifle | tank | ('mm'; 'compmod'; 'FROM-ME') |
|  | shield | shot | ('mortar'; 'conj'; 'FROM-ME') |
|  | jam | assault | ('rifle'; 'conj'; 'TO-ME') |
|  | lewis | rose | ('shoot'; 'adpobj'; 'TO-ME'; 'adpmod'; 'with') |
|  | nest | heavy | ('mount'; 'nsubjpass'; 'TO-ME') |
|  | battery | shoot | ('jump'; 'dobj'; 'TO-ME') |
|  | jump | arm | ('battery'; 'adpobj'; 'TO-ME') |
| bomb |  |  |  |
|  | atomic | explode | ('aimer'; 'compmod'; 'TO-ME') |
|  | explode | luftwaffe | ('detonate'; 'nsubjpass'; 'TO-ME') |
|  | disposal | atomic | ('detonate'; 'dobj'; 'TO-ME') |
|  | petrol | ira | ('pipe'; 'amod'; 'FROM-ME') |
|  | hydrogen | bombing | ('explode'; 'nsubj'; 'TO-ME') |
|  | parcel | explosion | ('luftwaffe'; 'nsubj'; 'FROM-ME') |
|  | luftwaffe | explosive | ('explodes'; 'compmod'; 'TO-ME') |
|  | ira | injure | ('defuse'; 'dobj'; 'TO-ME') |
|  | blast | bomber | ('explode'; 'compmod'; 'TO-ME') |
|  | explosion | hydrogen | ('aircraft'; 'adpobj'; 'FROM-ME'; 'adpmod'; 'by') |
|  | raid | raid | ('disposal'; 'compmod'; 'TO-ME') |
|  | premature | nuclear | ('blast'; 'compmod'; 'TO-ME') |
|  | neutron | drop | ('plant'; 'partmod'; 'FROM-ME') |
|  | pipe | sink | ('explosion'; 'compmod'; 'TO-ME') |
|  | bounce | target | ('raid'; 'dobj'; 'FROM-ME') |
|  | suicide | aircraft | ('atomic'; 'amod'; 'FROM-ME') |
|  | explosive | kill | ('hydrogen'; 'compmod'; 'FROM-ME') |
|  | nuclear | damage | ('cluster'; 'compmod'; 'FROM-ME') |
|  | conventional | plane | ('wave'; 'adpobj'; 'FROM-ME'; 'adpmod'; 'in') |
|  | nato | destroy | ('plant'; 'dobj'; 'TO-ME') |
| horse |  |  |  |
|  | thoroughbred | thoroughbred | ('pony'; 'conj'; 'FROM-ME') |
|  | trojan | mounted | ('mule'; 'conj'; 'FROM-ME') |
|  | harness | jockey | ('harness'; 'compmod'; 'FROM-ME') |
|  | rid | rid | ('trojan'; 'compmod'; 'FROM-ME') |
|  | crazy | trailer | ('ride'; 'dobj'; 'TO-ME') |
|  | racing | rider | ('wagon'; 'conj'; 'FROM-ME') |
|  | riding | riding | ('crazy'; 'compmod'; 'FROM-ME') |
|  | drawn | racing | ('rider'; 'conj'; 'FROM-ME') |
|  | arabian | breed | ('thoroughbred'; 'compmod'; 'FROM-ME') |
|  | cattle | cattle | ('rid'; 'dobj'; 'TO-ME') |
|  | ride | bombay | ('bull'; 'conj'; 'TO-ME') |
|  | artillery | ride | ('riding'; 'compmod'; 'TO-ME') |
|  | wild | stake | ('racing'; 'compmod'; 'TO-ME') |
|  | breed | cavalry | ('rid'; 'partmod'; 'FROM-ME') |
|  | race | regiment | ('race'; 'partmod'; 'FROM-ME') |
|  | rider | artillery | ('troop'; 'adpobj'; 'TO-ME'; 'adpmod'; 'of') |
|  | brigade | brigade | ('man'; 'adpobj'; 'TO-ME'; 'adpmod'; 'on') |
|  | dark | race | ('fall'; 'adpobj'; 'TO-ME'; 'adpmod'; 'from') |
|  | trainer | wild | ('age'; 'partmod'; 'FROM-ME') |
|  | shoe | trail | ('race'; 'adpobj'; 'TO-ME'; 'adpmod'; 'for') |
| fox |  |  |  |
|  | sac | hound | ('sauk'; 'conj'; 'TO-ME') |
|  | 20th | affiliation | ('hound'; 'conj'; 'FROM-ME') |
|  | hunting | cbs | ('searchlight'; 'compmod'; 'TO-ME') |
|  | news | nbc | ('footy'; 'compmod'; 'TO-ME') |
|  | twentieth | abc | ('mulder'; 'compmod'; 'TO-ME') |
|  | affiliate | affiliate | ('hare'; 'conj'; 'TO-ME') |
|  | sport | sport | ('abc'; 'conj'; 'TO-ME') |
|  | reservation | news | ('terrier'; 'compmod'; 'TO-ME') |
|  | matthew | programming | ('hunting'; 'compmod'; 'TO-ME') |
|  | 64 | channel | ('televise'; 'adpobj'; 'TO-ME'; 'adpmod'; 'on') |
|  | kid | 20th | ('ohio'; 'compmod'; 'TO-ME') |
|  | gray | network | ('cat'; 'conj'; 'FROM-ME') |
|  | broadcasting | kid | ('net'; 'compmod'; 'TO-ME') |
|  | programming | soccer | ('twentieth'; 'compmod'; 'FROM-ME') |
|  | abc | switch | ('kit'; 'compmod'; 'FROM-ME') |
|  | network | entertainment | ('televise'; 'adpobj'; 'TO-ME') |
|  | kit | morning | ('affiliation'; 'compmod'; 'TO-ME') |
|  | nbc | broadcast | ('gray'; 'amod'; 'FROM-ME') |
|  | soccer | television | ('news'; 'compmod'; 'TO-ME') |
|  | century | tv | ('affiliate'; 'compmod'; 'TO-ME') |
| table |  |  |  |
|  | picnic | picnic | ('list'; 'adpobj'; 'FROM-ME'; 'adpmod'; 'below') |
|  | periodic | tennis | ('carpianum'; 'conj'; 'TO-ME') |
|  | tennis | gospel | ('periodic'; 'compmod'; 'FROM-ME') |
|  | badminton | row | ('following'; 'compmod'; 'FROM-ME') |
|  | rotary | column | ('tennis'; 'compmod'; 'TO-ME') |
|  | html | database | ('picnic'; 'amod'; 'FROM-ME') |
|  | following | following | ('picnic'; 'compmod'; 'FROM-ME') |
|  | coffee | content | ('content'; 'adpobj'; 'FROM-ME'; 'adpmod'; 'of') |
|  | canon | template | ('salt'; 'compmod'; 'TO-ME') |
|  | summarize | medal | ('medal'; 'compmod'; 'FROM-ME') |
|  | breakfast | format | ('column'; 'adpobj'; 'TO-ME'; 'adpmod'; 'of') |
|  | altar | finish | ('periodic'; 'amod'; 'FROM-ME') |
|  | operating | element | ('truth'; 'compmod'; 'FROM-ME') |
|  | salt | list | ('content'; 'adpobj'; 'FROM-ME') |
|  | database | contain | ('sit'; 'adpobj'; 'TO-ME'; 'adpmod'; 'at') |
|  | dinner | entry | ('canon'; 'compmod'; 'FROM-ME') |
|  | mathematical | key | ('row'; 'adpobj'; 'TO-ME'; 'adpmod'; 'of') |
|  | pool | code | ('row'; 'adpobj'; 'TO-ME') |
|  | medal | water | ('knight'; 'adpobj'; 'TO-ME'; 'adpmod'; 'of') |
|  | setting | text | ('column'; 'adpobj'; 'TO-ME') |
| bowl |  |  |  |
|  | super | right-handed | ('batsman'; 'rcmod'; 'TO-ME') |
|  | batsman | super | ('xliv'; 'compmod'; 'TO-ME') |
|  | quiz | quiz | ('xlii'; 'compmod'; 'TO-ME') |
|  | pro | steelers | ('xxxiv'; 'compmod'; 'TO-ME') |
|  | rose | batsman | ('xliii'; 'compmod'; 'TO-ME') |
|  | alley | bowler | ('xxxvi'; 'compmod'; 'TO-ME') |
|  | sugar | rose | ('alley'; 'amod'; 'TO-ME') |
|  | lawn | wicket | ('xl'; 'compmod'; 'TO-ME') |
|  | bermuda | pro | ('select'; 'adpobj'; 'TO-ME'; 'adpmod'; 'to') |
|  | subdivision | sugar | ('pace'; 'dobj'; 'FROM-ME') |
|  | cotton | pace | ('rose'; 'compmod'; 'FROM-ME') |
|  | orange | nfl | ('bermuda'; 'compmod'; 'FROM-ME') |
|  | mvp | orange | ('quiz'; 'compmod'; 'FROM-ME') |
|  | dust | cotton | ('pro'; 'compmod'; 'FROM-ME') |
|  | compass | inning | ('sugar'; 'compmod'; 'FROM-ME') |
|  | barrow | touchdown | ('subdivision'; 'compmod'; 'TO-ME') |
|  | bowler | consecutive | ('invitation'; 'adpobj'; 'TO-ME'; 'adpmod'; 'to') |
|  | xiii | ball | ('super'; 'compmod'; 'FROM-ME') |
|  | all-american | yard | ('ball'; 'rcmod'; 'TO-ME') |
|  | toilet | stadium | ('orange'; 'compmod'; 'FROM-ME') |
| guitar |  |  |  |
|  | acoustic | bass | ('vocal'; 'appos'; 'TO-ME') |
|  | bass | percussion | ('vocal'; 'appos'; 'FROM-ME') |
|  | rhythm | acoustic | ('banjo'; 'conj'; 'TO-ME') |
|  | electric | vocal | ('amplifier'; 'conj'; 'FROM-ME') |
|  | amplifier | keyboard | ('riff'; 'compmod'; 'TO-ME') |
|  | vintage | drum | ('vocal'; 'conj'; 'FROM-ME') |
|  | pedal | amplifier | ('vocal'; 'conj'; 'TO-ME') |
|  | gibson | saxophone | ('keyboard'; 'conj'; 'FROM-ME') |
|  | solo | line-up | ('bass'; 'conj'; 'FROM-ME') |
|  | tenor | rhythm | ('bass'; 'appos'; 'FROM-ME') |
|  | playing | backing | ('bass'; 'compmod'; 'FROM-ME') |
|  | distortion | electric | ('percussion'; 'conj'; 'FROM-ME') |
|  | intro | gibson | ('rhythm'; 'compmod'; 'FROM-ME') |
|  | vocal | piano | ('solo'; 'compmod'; 'TO-ME') |
|  | chord | violin | ('acoustic'; 'amod'; 'FROM-ME') |
|  | hero | guitarist | ('drum'; 'conj'; 'FROM-ME') |
|  | rag | string | ('intro'; 'compmod'; 'TO-ME') |
|  | flute | solo | ('consist'; 'adpobj'; 'TO-ME'; 'adpmod'; 'on') |
|  | concerto | instrument | ('saxophone'; 'conj'; 'FROM-ME') |
|  | signature | melody | ('trumpet'; 'conj'; 'FROM-ME') |
| piano |  |  |  |
|  | concerto | cello | ('accordion'; 'compmod'; 'TO-ME') |
|  | viola | sonata | ('accompaniment'; 'compmod'; 'TO-ME') |
|  | cello | op | ('viola'; 'conj'; 'TO-ME') |
|  | sonata | viola | ('concertino'; 'adpobj'; 'TO-ME'; 'adpmod'; 'for') |
|  | violin | concerto | ('concertino'; 'adpobj'; 'TO-ME') |
|  | recital | violin | ('trio'; 'adpobj'; 'TO-ME'; 'adpmod'; 'for') |
|  | clarinet | clarinet | ('harpsichord'; 'conj'; 'TO-ME') |
|  | lesson | opus | ('cello'; 'conj'; 'TO-ME') |
|  | solo | saxophone | ('concerto'; 'compmod'; 'TO-ME') |
|  | opus | trumpet | ('violin'; 'conj'; 'TO-ME') |
|  | trio | pianist | ('virtuoso'; 'compmod'; 'TO-ME') |
|  | prepared | flute | ('op'; 'conj'; 'FROM-ME') |
|  | beethoven | lesson | ('sonata'; 'compmod'; 'TO-ME') |
|  | trumpet | trio | ('recital'; 'compmod'; 'TO-ME') |
|  | quartet | keyboard | ('harp'; 'conj'; 'TO-ME') |
|  | acoustic | guitar | ('sonata'; 'adpobj'; 'TO-ME'; 'adpmod'; 'for') |
|  | atop | solo | ('concerto'; 'adpobj'; 'TO-ME'; 'adpmod'; 'for') |
|  | saxophone | orchestra | ('clarinet'; 'conj'; 'TO-ME') |
|  | duet | organ | ('violin'; 'conj'; 'FROM-ME') |
|  | suite | drum | ('quintet'; 'compmod'; 'TO-ME') |

**Top20 word2vec(top20\_word2vec.txt top20\_word2vec.csv):**

|  |  |  |
| --- | --- | --- |
| **word** | **bow5** | **dependency** |
| car |  |  |
|  | 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 |
| bus |  |  |
|  | 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 |
| hospital |  |  |
|  | clinic | sanatorium |
|  | hospitals | hospice |
|  | infirmary | sanitorium |
|  | hospice | hospitals |
|  | lying-in | sanitarium |
|  | dispensary | clinic |
|  | polyclinic | infirmary |
|  | 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 |
| hotel |  |  |
|  | 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 | cafֳ© |
|  | apartments | condo |
|  | boutique | brewpub |
|  | nightclub | sheraton |
|  | marriott | steakhouse |
|  | travelodge | brasserie |
| gun |  |  |
|  | 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 |
| bomb |  |  |
|  | 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 |
| horse |  |  |
|  | 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 |
| fox |  |  |
|  | 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 |  |  |
|  | 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 |
| bowl |  |  |
|  | xliii | bowls |
|  | xlii | superbowl |
|  | xliv | arenabowl |
|  | bowls | wcws |
|  | xlvi | wnit |
|  | tostitos | nlcs |
|  | xli | arenacup |
|  | xxxviii | postseason |
|  | xlv | nit |
|  | xxxv | xliii |
|  | xxxix | llws |
|  | xxxvii | beanpot |
|  | xlvii | xlv |
|  | xxxvi | triplemanֳ­a |
|  | xxxiv | alcs |
|  | bluebonnet | nlds |
|  | gator | cup |
|  | xxviii | kvalserien |
|  | xxxii | tourney |
|  | xxxi | cws |
| guitar |  |  |
|  | 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 |
| piano |  |  |
|  | 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 |

**Top20 contexts word2vec(top20\_word2vec\_contexts.txt top20\_word2vec\_1st\_order.csv):**

|  |  |  |
| --- | --- | --- |
| **words** | **bow5** | **dependency** |
| car |  |  |
|  | racing | adpmod:byI\_commute |
|  | mygale | amod\_street-legal |
|  | bmw | conj\_hovercraft |
|  | driver | amod\_newly-designed |
|  | motor | amod\_liter |
|  | cars | amod\_late-model |
|  | dealership | poss\_brink |
|  | parked | compmod\_m1918 |
|  | rear-drive | adpmod:fromI\_tossed |
|  | accident | compmod\_high-wing |
|  | mid-engined | poss\_waltrip |
|  | dealerships | adpmod:byI\_escapes |
|  | touring | amod\_requisitioned |
|  | champ | compmodI\_rallying |
|  | race | nsubjI\_drivers |
|  | toyota | compmod\_rigid |
|  | rental | adpmod:insideI\_placing |
|  | suv | compmod\_hilux |
|  | truck | apposI\_bike |
|  | auto | poss\_matty |
| bus |  |  |
|  | buses | adpmod:byI\_commute |
|  | ksrtc | amod\_east-bound |
|  | samtrans | conj\_hovercraft |
|  | seabus | adpmod:onI\_ridership |
|  | smrt | compmod\_airlink |
|  | connexxion | compmod\_operates |
|  | inter-city | conj\_busses |
|  | intercity | compmod\_xpt |
|  | msrtc | compmod\_yrt |
|  | transit | dobjI\_onboard |
|  | double-decker | amod\_20-bit |
|  | tram | compmod\_citybus |
|  | commuter | amod\_requisitioned |
|  | apsrtc | adpmod:between\_belfast |
|  | services | amod\_street-legal |
|  | routes | compmod\_n11 |
|  | hiroden | compmodI\_hookups |
|  | metrobus | amod\_newly-designed |
|  | kmb | adpmod:like\_express |
|  | fixed-route | compmod\_weka |
| hospital |  |  |
|  | bethlem | compmod\_siriraj |
|  | moorfields | compmod\_safdarjung |
|  | psychiatric | adpmod:of\_nuova |
|  | hospitals | compmod\_strangeways |
|  | infirmary | amod\_maximum-security |
|  | foundling | compmod\_armley |
|  | siriraj | compmod\_combermere |
|  | maudsley | compmod\_fresnes |
|  | westmead | compmod\_eastview |
|  | bellevue | compmod\_desloge |
|  | convalescent | compmod\_monklands |
|  | polyclinic | compmod\_nudgee |
|  | maternity | compmod\_longmeadow |
|  | sunnybrook | compmod\_edgemont |
|  | soroka | compmod\_southridge |
|  | clinic | compmod\_archbold |
|  | mulago | compmod\_shinagawa |
|  | in-patient | compmod\_eppley |
|  | dispensary | compmod\_nantong |
|  | deaconess | compmod\_ravenscourt |
|  | hotel |  |
|  | hotels | compmod\_nymphenburg |
|  | westin | compmod\_dolmabahֳ§e |
|  | radisson | compmod\_whitwell |
|  | kempinski | conj\_guesthouses |
|  | sofitel | compmod\_ravenscourt |
|  | ramada | compmod\_hanworth |
|  | biltmore | compmod\_siriraj |
|  | ritz-carlton | compmod\_strangeways |
|  | sheraton | compmod\_safdarjung |
|  | marriott | compmod\_armley |
|  | waldorf-astoria | adpmod:onI\_renovations |
|  | five-star | compmod\_manresa |
|  | 5-star | rcmod\_renting |
|  | crowne | compmod\_portchester |
|  | doubletree | compmod\_eastview |
|  | restaurant | compmod\_combermere |
|  | hyatt | adpmod:nearI\_parked |
|  | hilton | compmod\_kirkdale |
|  | luxury | compmod\_edgemont |
|  | luxurious | compmod\_blackness |
| gun |  |  |
|  | guns | num\_88mm |
|  | submachine | adpmod:withI\_rearmed |
|  | machine | compmod\_m1918 |
|  | gatling | compmodI\_fellatio |
|  | sub-machine | num\_60mm |
|  | howitzer | adpmod:ofI\_hilt |
|  | rifle | compmod\_karabiner |
|  | 40mm | compmod\_t-55 |
|  | 11-inch | adpmod:forI\_cartridges |
|  | emplacements | compmodI\_cwt |
|  | muzzle-loading | compmod\_varmint |
|  | 20mm | amod\_newly-designed |
|  | 75mm | conjI\_turban |
|  | smoothbore | rcmod\_crushes |
|  | turrets | adpmod:fromI\_bled |
|  | recoilless | amod\_point-and-shoot |
|  | 5-inch | compmod\_mlrs |
|  | turret | compmod\_bvr |
|  | breech-loading | compmodI\_set-ups |
|  | anti-aircraft | adpmod:withI\_pistol |
| bomb |  |  |
|  | bombs | num\_88mm |
|  | detonated | adpmod:on\_pentagon |
|  | detonates | compmodI\_splashes |
|  | detonate | adpmod:byI\_unharmed |
|  | atomic | adpmod:withI\_rearmed |
|  | exploded | compmod\_single-car |
|  | detonating | compmod\_m1918 |
|  | bomber | compmod\_phishing |
|  | bombing | compmodI\_fellatio |
|  | blast | rcmod\_bursts |
|  | bombings | compmod\_decapitation |
|  | booby-trap | adpmod:over\_berlin |
|  | landmine | num\_60mm |
|  | v-1 | compmod\_nega |
|  | hiroshima | rcmod\_crushes |
|  | nuclear | adpmod:at\_wonsan |
|  | laser-guided | conjI\_hellcat |
|  | explodes | compmod\_bvr |
|  | blasts | compmod\_econoline |
|  | detonation | compmod\_t-55 |
| horse |  |  |
|  | horses | amod\_gaited |
|  | standardbred | nsubjpassI\_spooked |
|  | thoroughbred | compmodI\_drovers |
|  | trotting | dobjI\_shoe |
|  | riding | adpmod:byI\_commute |
|  | racing | rcmod\_roams |
|  | ridden | poss\_quixote |
|  | galloping | compmod\_ch-53e |
|  | bred | appos\_pony |
|  | gelding | compmod\_poitevin |
|  | stud | adpmod:ofI\_hilt |
|  | appaloosa | adpmod:byI\_escapes |
|  | pommel | compmodI\_fillet |
|  | bareback | compmodI\_cogburn |
|  | bronc | compmod\_beardless |
|  | sired | apposI\_bike |
|  | thoroughbreds | compmod\_m1918 |
|  | friesian | poss\_quagmire |
|  | stables | conjI\_turban |
|  | aqha | adpmod:byI\_foot |
| fox |  |  |
|  | vulpes | amod\_crab-eating |
|  | news | appos\_canadensis |
|  | movietone | compmodI\_subchannels |
|  | owned-and-operated | conj\_mattel |
|  | nbc | compmod\_jacki |
|  | cbs | conjI\_raimi |
|  | terrier | conjI\_corgan |
|  | affiliate | conjI\_leda |
|  | cnn | conj\_bluebird |
|  | msnbc | conj\_nolte |
|  | abc | conjI\_browder |
|  | espn | conj\_keeler |
|  | mynetworktv | conjI\_steinem |
|  | nine-tailed | conjI\_thurmond |
|  | wb | conjI\_melly |
|  | wish-tv | conjI\_seaver |
|  | wtae-tv | compmodI\_e-type |
|  | blitzer | apposI\_msnbc |
|  | marten | conjI\_devitt |
|  | cw | conj\_o'hare |
| table |  |  |
|  | sortable | adpmod:inI\_rows |
|  | tables | adpmod:offI\_knocks |
|  | tennis | adpmod:onI\_seventh |
|  | lookup | amod\_ten-foot |
|  | billiards | compmod\_abydos |
|  | foosball | amod\_25-metre |
|  | periodic | adpmod:intoI\_face |
|  | billiard | adpmod:belowI\_added |
|  | toc | adpmod:throughI\_throw |
|  | look-up | compmod\_three-judge |
|  | hash | adpmod:ontoI\_projects |
|  | collapsible | adpmod:aroundI\_looked |
|  | snooker | adpmod:onI\_substance |
|  | below | adpmod:onI\_slides |
|  | round | amod\_rooftop |
|  | carom | adpmod:insideI\_placing |
|  | contents | adpmod:withinI\_display |
|  | tabular | adpmod:onI\_entities |
|  | bottom | dobjI\_clogs |
|  | headers | adpmod:aboveI\_from |
| bowl |  |  |
|  | xli | adpmod:against\_auburn |
|  | bowls | conjI\_preseason |
|  | xliv | adpmod:against\_oklahoma |
|  | xxxviii | adpmod:inI\_2-1 |
|  | xliii | conj\_play-offs |
|  | xlii | compmod\_pan-pacific |
|  | xlvi | adpmod:forI\_stadiums |
|  | xlv | adpmod:atI\_clinched |
|  | super | adpmod:intoI\_win |
|  | xxxix | amod\_25-metre |
|  | xxxv | conjI\_regionals |
|  | xxxi | amod\_sm-liiga |
|  | xxxvii | adpmod:against\_brentford |
|  | xxxvi | adpmod:ofI\_tackle |
|  | xlvii | poss\_postman |
|  | xxxiii | adpmod:inI\_semifinal |
|  | xxv | adpmod:against\_walsall |
|  | xxvii | adpmod:atI\_pitcher |
|  | xxviii | compmod\_doomben |
|  | xxxiv | compmod\_deuces |
| guitar |  |  |
|  | bass | conjI\_bongos |
|  | guitars | amod\_end-blown |
|  | drums | conj\_back-up |
|  | keyboards | adpmod:onI\_harris |
|  | nyckelharpa | adpmod:forI\_adagio |
|  | glockenspiel | adpmod:onI\_thompson |
|  | 6-string | appos\_mandolin |
|  | tambourine | dep\_keyboard |
|  | banjo | adpmod:ofI\_virtuoso |
|  | vocals | adpmod:onI\_foster |
|  | harmonica | amod\_necked |
|  | fretless | conjI\_flinders |
|  | ukulele | adpmod:onI\_supplemented |
|  | 12-string | possI\_pickups |
|  | benmont | amod\_banjo |
|  | six-string | compmod\_sandberg |
|  | riffs | adpmod:ofI\_sweethearts |
|  | piano | poss\_diddley |
|  | accordion | depI\_trumpet |
|  | melodica | adpmod:onI\_pianist |
| piano |  |  |
|  | violin | adpmod:forI\_adagio |
|  | sonata | amod\_end-blown |
|  | cello | conjI\_bongos |
|  | concerto | adpmod:onI\_harris |
|  | op | adpmod:ofI\_virtuoso |
|  | harpsichord | adpmod:onI\_thompson |
|  | concertos | conj\_back-up |
|  | viola | compmod\_kreutzer |
|  | violoncello | compmod\_vomeronasal |
|  | sonatas | adpmod:onI\_supplemented |
|  | divertimento | appos\_mandolin |
|  | pianoforte | apposI\_prelude |
|  | e-flat | adpmod:onI\_foster |
|  | concerti | compmod\_akiyoshi |
|  | concertante | dep\_keyboard |
|  | clarinet | compmod\_sandberg |
|  | sonatina | conjI\_kabuki |
|  | bassoon | amod\_necked |
|  | nyckelharpa | dobjI\_syncing |
|  | b-flat | depI\_trumpet |