Submitters:
Ameen Ali   318429792
Tal Klein   312592520

Writeup 3 - Distributional Semantics

1. **Summarize statistics regarding your experiments:**
   a. **The various thresholds you employed.**

   The words considered for similarity:
   - words which passed the 100 frequency threshold.
   - Words with a POS tag in the following set:
   'NN', 'NNS', 'NNP', 'NNPS', 'VB', 'VBD', 'VBG', 'VBN', 'VBP', 'JJ', 'JJR', 'JJS', 'VBZ', 'WRB'.

   The tuples of (target word, feature word) considered for similarity:

   - Appeared of more than 3 times
   - Having non-negative PMI

   b. **Number of words considered for similarity: the number of words which passed the 100 frequency threshold (columns in the word-feature co-occurrence matrix).**

   8486 words passed the 100 frequency threshold and 7923 words passed all thresholds mentioned above.

   c. **The number of features considered in your computation for each co-occurrence type (rows in the matrix).**

   The number of features depends of the **chosen context**:
   Sentence (1): Max number of features is 8486 and the average number of features is 863
   Window, k=2 (2): Max number of features is 8486 and the average number of features is 315
   Tree (3): Max number of features is 152990 and the average number of features is 16

2. **The 20 most similar words (2nd-order similarity) for each of the target words in the qualitative evaluation, for each co-occurrence type.**
   Write in the report your conclusions from examining these lists, as instructed above, along with examples.

Words that are marked in red are less related to the given word or too general.
We marked the first 5 words manually.

car

Sentence (1)	Window, k=2 (2)	Tree (3)
'drive', 'race', 'men', 'train', 'aircraft', 'driver', 'put', 'choose', 'truck', 'pass', 'long', 'performance', 'total', 'replace', 'track', 'board', 'draw', 'stop', 'hand', 'medium'	'one', 'truck', 'vehicle', 'boat', 'engine', 'train', 'body', 'aircraft', 'drive', 'fact', 'choose', 'race', 'player', 'store', 'hand', 'light', 'special', 'addition', 'story', 'mission'	'vehicle', 'truck', 'boat', 'engine', 'aircraft', 'automobile', 'train', 'locomotive', 'bus', 'ship', 'model', 'product', 'driver', 'motorcycle', 'machine', 'unit', 'weapon', 'track', 'item', 'vessel'

bus

Sentence (1)	Window, k=2 (2)	Tree (3)
'passenger', 'traffic', 'rail', 'transit', 'transportation', 'jack', 'metro', 'arm', 'double', 'vessel', 'web', 'direction', 'citizen', 'airport', 'bit', 'locomotive', 'plot', 'southeast', 'train', 'connection'	'passenger', 'rail', 'boat', 'airport', 'hotel', 'mine', 'truck', 'pacific', 'achieve', 'fly', 'affair', 'success', 'transit', 'metro', 'express', 'engine', 'port', 'golden', 'factory', 'genus'	'rail', 'tram', 'train', 'taxi', 'ferry', 'automobile', 'car', 'vehicle', 'subway', 'transit', 'freight', 'truck', 'railway', 'locomotive', 'boat', 'traffic', 'passenger', 'commuter', 'bike', 'plane'

hospital

Sentence (1)	Window, k=2 (2)	Tree (3)
'medical', 'chinese', 'latter', 'learn', 'step', 'store', 'stand', 'grant', 'transfer', 'separate', 'successful', 'associate', 'entire', 'condition', 'worker', 'big', 'seek', 'double', 'manage', 'intend'	'camp', 'effort', 'conduct', 'attend', 'mission', 'medical', 'status', 'care', 'category', 'institution', 'press', 'green', 'health', 'store', 'honor', 'job', 'merge', 'entire', 'daily', 'picture'	'clinic', 'center', 'prison', 'college', 'museum', 'school', 'library', 'facility', 'campus', 'hotel', 'airport', 'stadium', 'hall', 'centre', 'institution', 'monastery', 'park', 'theatre', 'office', 'camp'

hotel

Sentence (1)	Window, k=2 (2)	Tree (3)
'restaurant', 'wood', 'room', 'abandon', 'fill', 'brand', 'slow', 'memorial', 'dr', 'contact', 'double', 'horse', 'walk', 'palace', 'garden', 'buy', 'rich', 'occupy', 'difficult', 'grant'	'beach', 'worker', 'garden', 'estate', 'i', 'shop', 'reserve', 'bus', 'avenue', 'restaurant', 'sense', 'pakistan', 'temple', 'sing', 'floor', 'administration', 'responsible', 'square', 'britain', 'bill'	'restaurant', 'shop', 'apartment', 'theater', 'house', 'building', 'mill', 'estate', 'inn', 'pub', 'resort', 'store', 'synagogue', 'hall', 'castle', 'library', 'mosque', 'temple', 'tower', 'theatre'

gun

Sentence (1)	Window, k=2 (2)	Tree (3)
'rifle', 'fit', 'machine', 'enemy', 'demand', 'weapon', 'bar', 'intend', 'initial', 'battery', 'entire', 'pick', 'tank', 'wound', 'eye', 'hear', 'crew', 'shot', 'advance', 'piece'	'rifle', 'weapon', 'aircraft', 'tank', 'missile', 'battery', 'box', 'target', 'machine', 'engine', 'artillery', 'rocket', 'bomb', 'mortar', 'whole', 'plane', 'pull', 'bar', 'brigade', 'step'	'cannon', 'pistol', 'weapon', 'rifle', 'tank', 'battery', 'aircraft', 'sword', 'artillery', 'engine', 'camera', 'bullet', 'device', 'machine', 'vessel', 'equipment', 'ship', 'boat', 'rocket', 'shot'

bomb

Sentence (1)	Window, k=2 (2)	Tree (3)
'raid', 'commit', 'enemy', 'fell', 'weapon', 'civilian', 'dozen', 'bomber', 'opponent', 'pilot', 'shot', 'mine', 'abandon', 'injure', 'daniel', 'reading', 'crash', 'sydney', 'defence', 'attract'	'bomber', 'rescue', 'storm', 'bombing', 'tank', 'crash', 'squadron', 'pilot', 'flight', 'rocket', 'gun', 'latter', 'enemy', 'fighter', 'everything', 'aircraft', 'torpedo', 'chain', 'patrol', 'push'	'torpedo', 'bombing', 'missile', 'rocket', 'weapon', 'explosion', 'reactor', 'bomber', 'earthquake', 'shell', 'raid', 'attack', 'grenade', 'eruption', 'fire', 'aircraft', 'ball', 'strike', 'storm', 'bombardment'

horse

Sentence (1)	Window, k=2 (2)	Tree (3)
'poor', 'farm', 'fast', 'dog', 'existence', 'male', 'hunt', 'eye', 'farmer', 'owner', 'guide', 'combination', 'felt', 'wood', 'thousand', 'bar', 'citizen', 'latter', 'attract', 'zone'	'catch', 'incident', 'following', 'assistant', 'animal', 'poland', 'couple', 'drop', 'factory', 'empire', 'reserve', 'bar', 'hear', 'fish', 'gun', 'administrator', 'back', 'whole', 'earth', 'square'	'dog', 'cattle', 'cat', 'sheep', 'bike', 'bull', 'motorcycle', 'bird', 'bicycle', 'elf', 'car', 'boy', 'animal', 'specimen', 'motor', 'athlete', 'men', 'monster', 'boat', 'bulb'

fox

Sentence (1)	Window, k=2 (2)	Tree (3)
'nbc', 'jackson', 'desire', 'adam', 'andrew', 'cbs', 'struggle', 'bbc', 'chris', 'abc', 'johnson', 'bob', 'latter', 'fame', 'cable', 'comedy', 'hire', 'nomination', 'drop', 'roll'	'nbc', 'abc', 'interested', 'cbs', 'israeli', 'saturday', 'global', 't', 'programming', 'giant', 'gay', 'wave', 'morning', 'expansion', 'struggle', 'j', 'distribution', 'bell', 'documentary', 'colour'	'cbs', 'cbc', 'nbc', 'abc', 'shaw', 'anderson', 'baker', 'patterson', 'itv', 'parker', 'hudson', 'hart', 'pb', 'kennedy', 'johnson', 'smith', 'carter', 'bbc', 'espn', 'mcdonald'

table

Sentence (1)	Window, k=2 (2)	Tree (3)
'step', 'piece', 'couple', 'simple', 'opportunity', 'entry', 'hear', 'handle', 'plot', 'horse', 'internet', 'concept', 'i', 'entire', 'latter', 'basic', 'response', 'lay', 'mass', 'anything'	'box', 'poor', 'easy', 'task', 'double', 'ball', 'web', 'wood', 'exchange', 'map', 'document', 'file', 'minor', 'piece', 'fine', 'background', 'element', 'bill', 'spot', 'code'	'section', 'paragraph', 'text', 'diffs', 'encyclopedia', 'entry', 'diagram', 'page', 'summary', 'column', 'template', 'ranking', 'image', 'category', 'map', 'module', 'box', 'picture', 'diff', 'poll'

bowl

Sentence (1)	Window, k=2 (2)	Tree (3)
'nfl', 'improvement', 'ncaa', 'super', 'pitch', 'quarterback', 'deliver', 'j', 'cricket', 'planet', 'defensive', 'pas', 'writing', 'victory', 'defend', 'limited', 'touchdown', 'trophy', 'chris', 'celebrate'	'exhibition', 'nfl', 'felt', 'gay', 'harbor', 'spot', 'commonwealth', 'qualify', 'thousand', 'playoff', 'shore', 'bell', 'indie', 'fc', 'shopping', 'organ', 'weather', 'golden', 'stock', 'portugal'	'super', 'sec', 'ncaa', 'tournament', 'nba', 'cup', 'mlb', 'boxing', 'alfonso', 'nhl', 'wrestling', 'rugby', 'polo', 'playoff', 'afl', 'gt', 'leo', 'championship', 'tennis', 'league'

guitar

Sentence (1)	Window, k=2 (2)	Tree (3)
'bass', 'drum', 'vocal', 'guitarist', 'piano', 'keyboard', 'bassist', 'solo', 'singer', 'musician', 'steve', 'chris', 'drummer', 'mike', 'tony', 'vocalist', 'percussion', 'brian', 'jim', 'dave'	'vocal', 'bass', 'piano', 'drum', 'guitarist', 'trumpet', 'solo', 'keyboard', 'vocalist', 'voice', 'musician', 'tune', 'bob', 'drummer', 'violin', 'taylor', 'saxophone', 'steve', 'folk', 'flute'	'drum', 'bass', 'keyboard', 'piano', 'vocal', 'violin', 'saxophone', 'cello', 'flute', 'trumpet', 'instrument', 'horn', 'viola', 'percussion', 'solo', 'organ', 'dance', 'string', 'cd', 'music'

piano

Sentence (1)	Window, k=2 (2)	Tree (3)
'trumpet', 'bass', 'string', 'flute', 'composer', 'violin', 'solo', 'guitar', 'composition', 'recording', 'don', 'drum', 'chorus', 'orchestra', 'cello', 'trio', 'ensemble', '1970s', 'saxophone', 'concerto'	'violin', 'flute', 'cello', 'concerto', 'trumpet', 'viola', 'guitar', 'string', 'saxophone', 'soloist', 'bass', 'ensemble', 'sonata', 'solo', 'orchestra', 'op', 'vocal', 'percussion', 'keyboard', 'drum'	'violin', 'viola', 'guitar', 'bass', 'cello', 'flute', 'keyboard', 'horn', 'saxophone', 'trumpet', 'drum', 'instrument', 'percussion', 'organ', 'dance', 'vocal', 'composition', 'music', 'choir', 'mathematics'

Conclusion:

The "sentence" method performed the worst, "window" method was somewhat better and the "tree" method yielded great results.

The words produced by the "tree" method have a syntactic context while the other methods have "bag of words" context.

3. **For each target word, print in a convenient table format the 20 top context attributes for that word, for each of the 3 co-occurrence types (those attributes with highest PMI values in the target word's vector, considered as 1st-order similarities).Write a short qualitative comparison between the lists in item 2 (2nd-order) and the lists in item 3 (1st-order).**

Words that are marked in red are less related to the given word or too general. We marked the first 3 words manually.

car

Sentence (1)	Window, k=2 (2)	Tree (3)
'touring', 'chevrolet', 'bmw', 'gt', 'ferrari', 'gm', 'chassis', 'jaguar', 'honda', 'toyota', 'holden', 'lotus', 'parking', 'trailer', 'lap', 'racing', 'auto', 'driver', 'luxury', 'compartment'	'touring', 'chevrolet', 'rental', 'vintage', 'armored', 'accident', 'luxury', 'jaguar', 'parking', 'toyota', 'honda', 'lotus', 'collide', 'gt', 'wash', 'ferrari', 'bmw', 'compartment', 'toss', 'dining'	'adpmod to design adpobj up', 'trend poss down', 'adpmod in engine adpobj up', 'adpmod of driver adpobj up', 'plymouth poss down', 'wreck dobj up', 'price amod down', 'race nsubjpass up', 'locomotive conj down', 'dome amod down', 'exit nsubj up', 'ford poss down', 'adpmod per driver adpobj up', 'adpmod of plate adpobj up', 'chase amod down', 'parking compmod up', 'adpmod by minute adpobj up', 'wash compmod up', 'hire compmod down', 'adpmod with race adpobj up'

bus

Sentence (1)	Window, k=2 (2)	Tree (3)
'transit', 'livery', 'rout', 'commuter', 'interchange', 'tram', 'taxi', 'subway', 'terminal', 'fare', 'depot', 'metro', 'route', 'hyderabad', 'shuttle', 'terminus', 'rack', 'chassis', 'parking', 'underneath'	'rout', 'taxi', 'interchange', 'transit', 'subway', 'tram', 'shuttle', 'depot', 'intel', 'terminal', 'commuter', 'barton', 'trevor', 'livery', 'parking', 'fare', 'toss', 'route', 'tramway', 'gasoline'	'taxi conj up', 'route nsubj up', 'tram conj up', 'depot poss up', 'tram conj down', 'subway conj up', 'adpmod from competition adpobj up', 'rout nsubj up', 'terminus compmod up', 'taxi conj down', 'automobile conj up', 'stop compmod up', 'serial compmod down', 'interchange compmod up', 'bus conj up', 'adpmod by provide compmod up', 'adpmod by accessible adpobj up', 'paint nsubjpass up', 'rail conj down', 'bus conj down'

hospital

Sentence (1)	Window, k=2 (2)	Tree (3)

'here', 'psychiatric', 'nh', 'clinic', 'acute', 'mir', 'surgical', 'healthcare', 'insane', 'mercy', 'bed', 'nursing', 'ambulance', 'nurse', 'surgery', 'addiction', 'middlesex', 'surgeon', 'andrea', 'sick'	'psychiatric', 'clinic', 'acute', 'nh', 'insane', 'bed', 'mercy', 'middlesex', 'ruby', 'hopkins', 'alexandra', 'sick', 'mayo', 'davidson', 'overnight', 'nursing', 'andrea', 'rush', 'teaching', 'stabilize'	'bed amod down', 'sailor poss down', 'adpmod at physician adpobj up', 'vincent poss down', 'adpmod a facility adpobj up', 'hospital dep up', 'priory conj up', 'undergo rcmod down', 'breach compmod down', 'adpmod at pronounce adpobj up', 'recover partmod down', 'guy poss down', 'adpmod at nurse adpobj up', 'adpmod at doctor adpobj up', 'adpmod to rush adpobj up', 'accommodate nsubj up', 'adpmod with bed adpobj down', 'teaching amod down', 'adpmod at resident adpobj up', 'anthony poss down'

hotel

Sentence (1)	Window, k=2 (2)	Tree (3)
'casino', 'luxury', 'resort', 'sol', 'dining', 'plaza', 'havana', 'inn', 'restaurant', 'nightclub', 'accommodation', 'traveller', 'bedroom', 'windsor', 'lagoon', 'traveler', 'monica', 'cater', 'canary', 'carlton'	'casino', 'luxury', 'sol', 'plaza', 'lobby', 'nightclub', 'resort', 'restaurant', 'dining', 'havana', 'inn', 'accommodation', 'cater', 'famed', 'paramount', 'riverside', 'suite', 'laptop', 'northumberland', 'windsor'	'inn appos up', 'inn conj down', 'adjacent appos down', 'cottage conj down', 'adpmod a purchase adpobj up', 'casino conj up', 'existence compmod up', 'adpmod at restaurant adpobj up', 'hotel appos down', 'casino conj down', 'adpmod of lobby adpobj up', 'hotel appos up', 'lobby compmod up', 'adpmod into check adpobj up', 'adpmod of restaurant adpobj up', 'adpmod at dinner adpobj up', 'rename rcmod down', 'bar poss up', 'plaza compmod down', 'resort conj down'

gun

Sentence (1)	Window, k=2 (2)	Tree (3)
'turret', 'armament', 'mm', 'ammunition', 'mk', 'mortar', 'machine', 'barrel', 'grenade', 'firing', 'pistol', 'he', 'rifle', 'battleship', 'bolt', 'firearm', 'pod', 'battery', 'bullet', 'aa'	'turret', 'mm', 'machine', 'ammunition', 'mortar', 'firing', 'armament', 'barrel', 'mk', 'bolt', 'augment', 'sniper', 'battery', 'rifle', 'frigate', 'deck', 'bunker', 'rose', 'annie', 'silence'	'adpmod with fire adpobj up', 'rifle amod down', 'adpmod from shot adpobj up', 'prisoner conj up', 'rifle appos down', 'cm amod down', 'barrel dobj up', 'jam nsubj up', 'adpmod with merit adpobj up', 'adpmod with damage adpobj up', 'knife conj down', 'velocity poss up', 'aim dobj up', 'barrel compmod up', 'machine compmod down', 'armament compmod up', 'store rcmod down', 'bullet conj down', 'ready nsubj up', 'adpmod by engage adpobj up'

bomb

Sentence (1)	Window, k=2 (2)	Tree (3)
'explode', 'luftwaffe', 'atomic', 'ira', 'disposal', 'bombing', 'explosive', 'explosion', 'baghdad', 'injure', 'fuse', 'hydrogen', 'neutron', 'blast', 'grenade', 'bomber', 'rack', 'libyan', 'invader', 'jin'	'explode', 'atomic', 'disposal', 'ira', 'luftwaffe', 'hydrogen', 'explosive', 'blast', 'fuse', 'baghdad', 'explosion', 'bounce', 'rig', 'libyan', 'raid', 'mortar', 'tanker', 'rip', 'crude', 'dismantle'	'adpmod by luftwaffe adpobj down', 'explode compmod up', 'pipe amod down', 'torpedo conj down', 'suicide partmod up', 'explode rcmod down', 'explosive conj up', 'explode partmod down', 'explode nsubj up', 'adpmod by destroy adpcomp up', 'luftwaffe nsubj down', 'adpmod of explosion adpobj up', 'explosive conj down', 'adpmod by ally adpobj down', 'explode dobj up', 'plant partmod down', 'blast compmod up', 'explosion compmod up', 'hide partmod down', 'aim nsubjpass up'

horse

Sentence (1)	Window, k=2 (2)	Tree (3)
'thoroughbred', 'likewise', 'breeder', 'jockey', 'cart', 'stud', 'rid', 'arabian', 'handicap', 'trailer', 'drawn', 'mare', 'riding', 'rider', 'cattle', 'breed', 'bombay', 'racing', 'breeding', 'trainer'	'thoroughbred', 'drawn', 'rid', 'cart', 'crazy', 'jockey', 'cattle', 'riding', 'racing', 'breeder', 'arabian', 'cow', 'rider', 'goat', '"the"', 'carriage', 'ride', 'sheep', 'stud', 'hobby'	'adpmod from fell adpobj up', 'fool conj up', 'hobby compmod down', 'jockey conj down', 'shoe dobj up', 'bull conj up', 'bay amod down', 'cart conj down', 'adpmod for line adpobj up', 'crazy compmod down', 'rid partmod down', 'thoroughbred amod down', 'adpmod a breed adpobj up', 'ride dobj up', 'wagon conj down', 'breeding compmod up', 'adpmod by earn adpobj up', 'rid dobj up', 'adpmod by pull adpobj up', 'rider conj down'

fox

Sentence (1)	Window, k=2 (2)	Tree (3)
'affiliation', 'hare', 'rabbit', 'newscast', 'subspecies', 'cnn', 'ufc', 'deer', 'cbs', 'nbc', 'abc', 'abundance', 'affiliate', 'holly', 'hunting', 'espn', 'beaver', 'sport', 'dana', 'statewide'	'hare', 'cnn', 'dana', 'pb', 'affiliation', 'rabbit', 'holly', 'hunting', 'affiliate', 'twentieth', '20th', 'gardner', 'abundance', 'newscast', 'reservation', 'ruby', 'sioux', 'news', 'abc', 'beaver'	'hare conj up', 'beaver conj up', 'adpmod on morning adpobj up', 'adpmod of subspecies adpobj up', 'adpmod on televise adpobj up', 'gardner compmod down', 'abc conj up', 'hunting compmod up', 'ohio compmod up', 'kit compmod down', 'cat conj down', 'affiliation compmod up', 'twentieth compmod down', 'gray amod down', 'net compmod up', 'mask compmod up', 'kid compmod up', 'sport compmod up', 'affiliate compmod up', 'news compmod up'

table

Sentence (1)	Window, k=2 (2)	Tree (3)
'for', 'with', 'periodic', 'picnic', 'tennis', 'below', 'html', 'gospel', 'marking', 'heading', 'besides', 'slice', 'subscription', 'row', 'summarize', 'column', 'pdf', 'width', 'header', 'excel'	'periodic', 'picnic', 'tennis', 'below', 'summarize', 'rotary', 'html', 'excel', 'canon', 'header', 'heading', 'coffee', 'pdf', 'following', 'dining', 'gaming', 'rout', 'aggregate', 'overview', 'astronomer'	'adpmod at eat adpobj up', 'contain appos up', 'adpmod below show adpobj down', 'adpmod below list adpobj down', 'periodic compmod down', 'following compmod down', 'align dobj up', 'adpmod in row adpobj up', 'tennis compmod up', 'table conj up', 'picnic amod down', 'adpmod of column adpobj up', 'format amod down', 'column compmod down', 'below compmod up', 'picnic compmod down', 'salt compmod up', 'adpmod in column adpobj up', 'aggregate compmod down', 'periodic amod down'

bowl

Sentence (1)	Window, k=2 (2)	Tree (3)
'quiz', 'steelers', 'batsman', 'super', 'over', 'bowler', 'rose', 'wicket', 'pro', 'alley', 'sugar', 'pace', 'starter', 'bowling', 'streak', 'stir', 'bronco', 'interception', 'nfl', 'cotton'	'batsman', 'super', 'quiz', 'over', 'alley', 'pro', 'rose', 'pace', 'bermuda', 'lawn', 'bowler', 'sugar', 'mvp', 'cotton', 'orange', 'subdivision', 'dust', 'snap', 'soup', 'xiii'	'over dobj down', 'hand nsubjpass up', 'over rcmod up', 'concede partmod down', 'fast dobj down', 'batsman rcmod up', 'alley amod up', 'adpmod of coverage compmod up', 'adpmod in giant compmod up', 'alley dobj down', 'insight compmod down', 'adpmod in victory compmod up', 'bowl conj down', 'bowl conj up', 'adpmod without take adpcomp down', 'adpmod to select adpobj up', 'bermuda compmod down', 'adpmod in rush adpobj up', 'ball rcmod up', 'pace dobj down'

guitar

Sentence (1)	Window, k=2 (2)	Tree (3)
'percussion', 'bass', 'acoustic', 'keyboard', 'vocal', 'drum', 'saxophone', 'line-up', 'backing', 'rhythm', 'vintage', 'gibson', 'electric', 'violin', 'chord', 'bo', 'tenor', 'piano', 'trumpet', 'cello'	'acoustic', 'bass', 'keyboard', 'percussion', 'vocal', 'rhythm', 'saxophone', 'drum', 'gibson', 'vintage', 'electric', 'intro', 'backing', 'tenor', 'barrett', 'flute', 'mick', 'trumpet', 'chord', 'violin'	'vocal appos down', 'adpmod on line-up adpobj up', 'adpmod on lineup adpobj up', 'backing conj down', 'bass dep down', 'intro compmod up', 'flourish compmod up', 'adpmod from black compmod down', 'vocal appos up', 'adpmod on ; adpobj up', 'keyboard conj down', 'saxophone conj up', 'adpmod of basic adpobj up', 'vocal conj up', 'vocal conj down', 'tune dobj up', 'bass conj down', 'rag

		compmod up', 'bass appos down', 'rhythm compmod down'

piano

Sentence (1)	Window, k=2 (2)	Tree (3)
'cello', 'viola', 'op', 'sonata', 'concerto', 'violin', 'saxophone', 'trumpet', 'flute', 'soprano', 'pianist', 'percussion', 'lesson', 'trio', 'tenor', 'keyboard', 'franz', 'organist', 'conservatory', 'emerson'	'sonata', 'viola', 'concerto', 'cello', 'op', 'violin', 'saxophone', 'lesson', 'trumpet', 'soprano', 'hammond', 'flute', 'trio', 'salon', 'organ', 'quartet', 'percussion', 'duet', 'solo', 'suite'	'adpmod for piece conj up', 'bass appos up', 'viola conj up', 'op conj down', 'organ conj up', 'cello conj up', 'flute conj down', 'adpmod for trio adpobj up', 'saxophone conj up', 'adpmod for concerto adpobj up', 'voice conj up', 'adpmod for composition adpobj up', 'concerto compmod up', 'violin conj up', 'sonata compmod up', 'ph compmod down', 'horn conj up', 'violin conj down', 'string conj down', 'trumpet conj up'

Conclusion:

Unlike the second order similarity, the "sentence" method performed like the "window" method the "tree" method. It can be explained by the transitivity of the second order similarity model. That means that the wider the context window is, the higher chance to produce more unrelated results in the second order.

In addition, the words produced are odd in some cases, for example, the word's "hospital" features include (window): middlesex, andrea, hopkins, alexandra and (tree): sailor, vincent, anthony.
But the "good" words make up for the noise.

4. **Report your manual judgments for the two words (possibly in appendix), the MAP results for the 3 co-occurrence types (AP for each of the two words and their average as MAP), and provide some insight on these results.**

Manually check of those words in question 2.
Result(car,sentence)= 1 1 0 1 1 1 0 0 1 0 0 0 0 0 1 0 0 0 1 0
AP(car,sentence)= (1/1 + 2/2 + 0 + 3/4 + 4/5 + 5/6 + 0 + 0 + 6/9 + 0 + 0 + 0 + 0 + 0 + 7/15 + 0 + 0 + 0 + 8/19 + 0) / 8 = 0.7422

Result(car,window)= 0 1 1 1 1 1 0 1 1 0 0 1 0 1 1 0 0 0 0 0
AP(car,window)=(0 + 1/2 + 2/3 + 3/4 + 4/5 + 5/6 + 0 + 6/8 + 7/9 + 0 + 0 + 8/12 + 0 + 9/14 + 10/15 + 0 + 0 + 0 + 0 + 0)/10 = 0.7054

Result(car,tree)= 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1
AP(car,tree)=(1/1 + 2/2 + 3/3 + 4/4 + 5/5 + 6/6 + 7/7 + 8/8 + 9/9 + 10/10 + 11/11 + 12/12 + 13/13 + 14/14 + 15/15 + 16/16 + 17/17 + 18/18 + 0 + 19/20)/19 = 0.9973

Result(piano,sentence)= 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1
AP(piano,sentence)= (1/1 + 2/2 + 3/3 + 4/4 + 5/5 + 6/6 + 7/7 + 8/8 + 9/9 + 10/10 + 11/11 + 12/12 + 13/13 + 14/14 + 15/15 + 16/16 + 17/17 + 0 + 18/19 + 19/20)/19=0.9946

Result(piano,window)= 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
AP(piano,window)= 1

Result(piano,tree)= 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
AP(piano,tree)= 1

MAP(sentence)=avg(AP(car,sentence),AP(piano,sentence)) = (0.7422+0.9946)/2= 0.8684
MAP(window)=(0.7054+1)/2=0.8527
MAP(tree)=(0.9973+1)/2=0.9987

The sentence and window method yielded similar result, even though the number of correct words produced by the window method was higher.
As expected, the tree method gave the highest result.

5. Write a brief description of your implementation of:
a. The estimation of the PMI values

Input: a sparse matrix of target words (rows) and feature words (columns) that is implemented as dictionary of dictionaries. To create symmetry, we turned the columns as rows as well.

Algorithm:
1. For each row (= word) in the matrix:
   1.1.   Calculate the sum of the elements (= features) in the row
   1.2.   Calculate the sum of all the elements

2. For each column (= feature) in the matrix:
   2.1.   Calculate the sum of the elements (= words) in the column

3. For each word w (row):
   3.1.   Calculate

sum of matrix row $\longrightarrow$ $P(w) = \dfrac{\#(w,\square)}{\sum_{w,c \in D} \#(w,c)}$

4. For each feature c (column):

4.1.   Calculate   sum of matrix column $\longrightarrow$ $P(c) = \dfrac{\#(\square, c)}{\sum_{w,c \in D} \#(w,c)}$

5. For each (w,c):
    5.1.   Calculate

matrix entry

sum of matrix entries $\longrightarrow$ $P(w,c) = \dfrac{\#(w,c)}{\sum_{w,c \in D} \#(w,c)}$

And calculate the log function of $PMI(w,c) = \dfrac{P(w,c)}{P(w)P(c)}$
As long as the PMI is positive
Otherwise, filter out those results (PPMI + not allowing PMI=0
because log(0) is -inifinty)

Complexity:
n = number of target words, k=number of features - O(2nk)

b. The efficient algorithm for computing all similarities for a target word.

Calculating cosine distance:

Algorithm:
1. For each row (= word) in the matrix, calculate the length of the vector of the word
2. For each word w (row) in the target words given (bounded number)
    2.1.   For each feature f (column):
        2.1.1. For each word v with feature f (v in the column of f)
               Calculate:
               dot(w,v) = dot(w,v) + log(freq_w_feature) * log(freq_v_feature)
3. For each word w (row) in the target words given
    3.1.   For each feature f (column):
        3.1.1. For each word v with feature f (v in the column of f)
               Calculate:
               sim(w,v) = dot(w,v) / length(w) * length(v)

Complexity:
n = number of target words, k=number of features - O(2nk)

6. **Report the results of the word2vec experiment, with the same types of content as in items 2, 3 and 4 above, for the two vector types of word2vec. For item 3, report the 10 context words for which the context vector has highest dot product similarity with the target word. Write a short comparison between the results of items**

2nd order similarity:

Top 20 similar words, ordered by similarity in descending order:

car

window, k=5 (2)	tree (3)
'cars', 'truck', 'automobile', 'vehicle', 'motorbike', 'motorcycle', 'driver', 'minivan', 'suv', 'lorry', 'motorcar', 'mid-engined', 'limousine', 'front-engined', 'moped', 'motorhome', 'mercedes-benz', 'bike', 'rear-engined', 'three-wheeled'	'truck', 'suv', 'vehicle', 'minivan', 'cars', 'speedboat', 'racecar', 'automobile', 'motorcar', 'jeep', 'limousine', 'minibus', 'lorry', 'limo', 'motorcycle', 'bike', 'motorhome', 'taxicab', 'roadster', 'wagon'

bus

window, k=5 (2)	tree (3)
'buses', 'tram', 'metrobus', 'intercity', 'busses', 'fixed-route', 'minibus', 'inter-city', 'ksrtc', 'commuter', 'apsrtc', 'msrtc', 'inter-urban', 'dial-a-ride', 'mini-bus', 'light-rail', 'rail', 'transit', 'trolleybus', 'limited-stop'	'minibus', 'tram', 'buses', 'jeepney', 'taxicab', 'motorcoach', 'taxi', 'trolleybus', 'lorry', 'truck', 'metrobus', 'streetcar', 'busses', 'ferryboat', 'trolley', 'tramcar', 'railcar', 'railmotor', 'intercityexpress', 'train'

hospital

window, k=5 (2)	tree (3)
'clinic', 'hospitals', 'infirmary', 'hospice', 'lying-in', 'dispensary', 'polyclinic', 'sanatorium', 'convalescent', 'mulago', 'addenbrooke', 'bethlem', 'psychiatric', 'maudsley', 'siriraj', 'sanitarium', 'in-patient', 'incurables', 'orthopaedic', 'westmead'	'sanatorium', 'hospice', 'sanitorium', 'hospitals', 'sanitarium', 'clinic', 'infirmary', 'polyclinic', 'dispensary', 'orphanage', 'poorhouse', 'almshouse', 'workhouse', 'institutet', 'leprosarium', 'rikshospitalet', 'heliport', 'gaol', 'guesthouse', 'motherhouse'

hotel

window, k=5 (2)	tree (3)
'motel', 'restaurant', 'doubletree', 'sheraton', 'hotels', 'ritz-carlton', 'sofitel', 'westin', 'ramada', 'casino', 'kempinski', 'mansion', 'inn', 'cafe', 'tavern', 'apartments',	'motel', 'hotels', 'casino', 'restaurant', 'inn', 'guesthouse', 'tavern', 'cafe', 'ritz-carlton', 'nightclub', 'travelodge', 'pizzeria', 'roadhouse',

'boutique', 'nightclub', 'marriott', 'travelodge'	'boardinghouse', 'caf\xc3\xa9', 'condo', 'brewpub', 'sheraton', 'steakhouse', 'brasserie'

gun

window, k=5 (2)	tree (3)
'guns', 'cannon', 'howitzer', 'sub-machine', 'flamethrower', 'belt-fed', '37mm', 'smoothbore', 'pistol', 'shkas', '105mm', '40mm', 'gatling', 'recoilless', '76mm', '3-inch', 'rifle', '88mm', 'large-caliber', 'autocannons'	'guns', 'handgun', 'machinegun', 'howitzer', 'pistol', 'rifle', 'shotgun', 'firearm', 'cannon', 'musket', 'crossbow', 'autocannon', 'phaser', 'flamethrower', 'revolver', 'carbine', 'machine-gun', 'weapon', 'carronade', 'pounder'

bomb

window, k=5 (2)	tree (3)
'bombs', 'detonated', 'detonates', 'detonate', 'booby-trap', 'detonating', 'firebomb', 'car-bomb', 'exploded', 'detonation', 'warhead', '500-pound', 'b61', 'laser-guided', 'blast', 'explosives', 'detonations', 'landmine', 'tallboy', 'thermonuclear'	'bombs', 'firebomb', 'landmine', 'car-bomb', 'grenade', 'torpedo', 'ied', 'warhead', 'bomber', 'bomblets', 'missile', 'nuke', 'detonator', 'booby-trap', 'kamikaze', 'munition', 'explosives', 'machinegun', 'a-bomb', 'firebombs'

horse

window, k=5 (2)	tree (3)
'horses', 'standardbred', 'saddlebred', 'gelding', 'thoroughbred', 'stallion', 'racehorses', 'dog', 'riderless', 'gaited', 'bronc', 'percheron', 'pony', 'trotting', 'harness', 'chariot', 'appaloosa', 'sulky', 'racehorse', 'greyhound'	'horses', 'goat', 'dog', 'stallion', 'mule', 'bronc', 'cow', 'unicycle', 'greyhound', 'bareback', 'camel', 'appaloosa', 'saddlebred', 'colt', 'zebu', 'donkey', 'sidesaddle', 'racehorse', 'elephant', 'pony'

fox

window, k=5 (2)	tree (3)
'abc', 'cbs', 'nbc', 'wxyz-tv', 'msnbc', 'ctv', 'wsvn', 'familynet', 'wttg', 'wjbk', 'wfxt', 'espn', 'wofl', 'cnn', 'oln', 'blitzer', 'nesn', 'wesh', 'wb', 'wgn-tv' table 'tables', 'sortable', 'wikitable', 'look-up', 'foosball', 'toc', 'bulleted', 'ping-pong', 'billiard', 'table-tennis', 'textbox', 'tray',	'daystar', 'nbc', 'byutv', 'kron', 'wolf', 'cbs-tv', 'familynet', 'abc', 'wccb', 'oln', 'wjar', 'hdnet', 'telefutura', 'woodchuck', 'nbc-tv', 'soapnet', 'cinemax', 'wdiv', 'mundofox', 'coyote' table 'tables', 'leaderboard', 'sideboard', 'chessboard', 'textbox', 'taskbar', 'gameboard', 'worksheet', 'tray', 'viewport', 'dais', 'flowchart', 'playfield', 'mantelpiece', 'stepladder', 'cladogram',

| 'lookup', 'wikitables', 'brackets', 'header', 'footer', 'tabular', 'menu', 'carom' | |

table

window, k=5 (2)	tree (3)
'tables', 'sortable', 'wikitable', 'look-up', 'foosball', 'toc', 'bulleted', 'ping-pong', 'billiard', 'table-tennis', 'textbox', 'tray', 'lookup', 'wikitables', 'brackets', 'header', 'footer', 'tabular', 'menu', 'carom'	'tables', 'leaderboard', 'sideboard', 'chessboard', 'textbox', 'taskbar', 'gameboard', 'worksheet', 'tray', 'viewport', 'dais', 'flowchart', 'playfield', 'mantelpiece', 'stepladder', 'cladogram', 'letterbox', 'windowsill', 'bookcase', 'wikitable'

bowl

window, k=5 (2)	tree (3)
'xliii', 'xlii', 'xliv', 'bowls', 'xlvi', 'tostitos', 'xli', 'xxxviii', 'xlv', 'xxxv', 'xxxix', 'xxxvii', 'xlvii', 'xxxvi', 'xxxiv', 'bluebonnet', 'gator', 'xxviii', 'xxxii', 'xxxi'	'bowls', 'superbowl', 'arenabowl', 'wcws', 'wnit', 'nlcs', 'arenacup', 'postseason', 'nit', 'xliii', 'llws', 'beanpot', 'xlv', 'tripleman\xc3\xada', 'alcs', 'nlds', 'cup', 'kvalserien', 'tourney', 'cws'

guitar

window, k=5 (2)	tree (3)
'harmonica', 'mandolin', 'bass', 'drums', 'guitars', 'keyboards', 'accordion', 'banjo', 'saxophone', '12-string', 'ukulele', 'trombone', 'fiddle', 'autoharp', 'melodica', 'percussion', 'vibraphone', 'tambourine', 'vocals', 'fretless'	'saxophone', 'bass', 'mandolin', 'harmonica', 'accordion', 'trombone', 'violin', 'banjo', 'guitars', 'cello', 'piano', 'vibraphone', 'sax', 'trumpet', 'autoharp', 'clarinet', 'sitar', 'fiddle', 'drums', 'marimba'

piano

window, k=5 (2)	tree (3)
'violin', 'cello', 'harpsichord', 'clarinet', 'viola', 'flute', 'bassoon', 'violoncello', 'oboe', 'concerto', 'saxophone', 'accordion', 'harp', 'trombone', 'sonatas', 'trumpet', 'mandolin', 'pianoforte', 'vibraphone', 'concertos'	'violin', 'cello', 'harpsichord', 'saxophone', 'clarinet', 'guitar', 'trombone', 'mandolin', 'vibraphone', 'marimba', 'accordion', 'pianoforte', 'bassoon', 'fortepiano', 'violoncello', 'trumpet', 'harmonica', 'clavinet', 'clavichord', 'euphonium'

Conclusions:

The results of the word2vec were much better than the results of our code which is based by the methods taught in class, as in most of the words were related to the target words. However, most of them are synonyms.
One reason is probably because we did not use smoothed PMIs in our code.
Second reason is the use of a neutral network vs. mathematical formulas.

Bowl produced many unrelated words: roman digits (?)

1st order similarity:

Top 20 similar words, ordered by dot product (target word*context word) in descending order:

car

window, k=5 (2)	tree (3)
'racing', 'mygale', 'bmw', 'driver', 'motor', 'cars', 'dealership', 'parked', 'rear-drive', 'accident', 'mid-engined', 'dealerships', 'touring', 'champ', 'race', 'toyota', 'rental', 'suv', 'truck', 'auto'	'amod_street-legal', 'conj_hovercraft', 'amod_newly-designed', 'amod_liter', 'amod_late-model', 'poss_brink', 'compmod_m1918', 'adpmod:fromI_tossed', 'compmod_high-wing', 'poss_waltrip', 'adpmod:byI_escapes', 'amod_requisitioned', 'compmodI_rallying', 'nsubjI_drivers', 'compmod_rigid', 'adpmod:insideI_placing', 'compmod_hilux', 'apposI_bike', 'poss_matty', 'dobjI_onboard'

bus

window, k=5 (2)	tree (3)
'buses', 'ksrtc', 'samtrans', 'seabus', 'smrt', 'connexxion', 'inter-city', 'intercity', 'msrtc', 'transit', 'double-decker', 'tram', 'commuter', 'apsrtc', 'services', 'routes', 'hiroden', 'metrobus', 'kmb', 'fixed-route'	'amod_east-bound', 'conj_hovercraft', 'adpmod:onI_ridership', 'compmod_airlink', 'compmod_operates', 'conj_busses', 'compmod_xpt', 'compmod_yrt', 'dobjI_onboard', 'amod_20-bit', 'compmod_citybus', 'amod_requisitioned', 'adpmod:between_belfast', 'amod_street-legal', 'compmod_n11', 'compmodI_hookups', 'amod_newly-designed', 'adpmod:like_express', 'compmod_weka', 'compmod_hs2'

hospital

window, k=5 (2)	tree (3)
'bethlem', 'moorfields', 'psychiatric', 'hospitals', 'infirmary', 'foundling', 'siriraj', 'maudsley', 'westmead', 'bellevue', 'convalescent', 'polyclinic', 'maternity', 'sunnybrook', 'soroka', 'clinic', 'mulago', 'in-patient', 'dispensary', 'deaconess'	'compmod_safdarjung', 'adpmod:of_nuova', 'compmod_strangeways', 'amod_maximum-security', 'compmod_armley', 'compmod_combermere', 'compmod_fresnes', 'compmod_eastview', 'compmod_desloge', 'compmod_monklands', 'compmod_nudgee', 'compmod_longmeadow', 'compmod_edgemont', 'compmod_southridge', 'compmod_archbold', 'compmod_shinagawa', 'compmod_eppley', 'compmod_nantong', 'compmod_ravenscourt', 'compmod_manresa'

hotel

window, k=5 (2)	tree (3)
'hotels', 'westin', 'radisson', 'kempinski', 'sofitel', 'ramada', 'biltmore', 'ritz-carlton', 'sheraton', 'marriott', 'waldorf-astoria', 'five-star', '5-star', 'crowne', 'doubletree', 'restaurant', 'hyatt', 'hilton', 'luxury', 'luxurious'	'compmod_dolmabah\xc3\xa7e', 'compmod_whitwell', 'conj_guesthouses', 'compmod_ravenscourt', 'compmod_hanworth', 'compmod_siriraj', 'compmod_strangeways', 'compmod_safdarjung', 'compmod_armley', 'adpmod:onl_renovations', 'compmod_manresa', 'rcmod_renting', 'compmod_portchester', 'compmod_eastview', 'compmod_combermere', 'adpmod:nearl_parked', 'compmod_kirkdale', 'compmod_edgemont', 'compmod_blackness', 'appos_acre'

gun

window, k=5 (2)	tree (3)
'guns', 'submachine', 'machine', 'gatling', 'sub-machine', 'howitzer', 'rifle', '40mm', '11-inch', 'emplacements', 'muzzle-loading', '20mm', '75mm', 'smoothbore', 'turrets', 'recoilless', '5-inch', 'turret', 'breech-loading', 'anti-aircraft'	'adpmod:withl_rearmed', 'compmod_m1918', 'compmodl_fellatio', 'num_60mm', 'adpmod:ofl_hilt', 'compmod_karabiner', 'compmod_t-55', 'adpmod:forl_cartridges', 'compmodl_cwt', 'compmod_varmint', 'amod_newly-designed', 'conjl_turban', 'rcmod_crushes', 'adpmod:froml_bled', 'amod_point-and-shoot', 'compmod_mlrs', 'compmod_bvr', 'compmodl_set-ups', 'adpmod:withl_pistol', 'conjl_hellcat'

bomb

window, k=5 (2)	tree (3)
'bombs', 'detonated', 'detonates', 'detonate', 'atomic', 'exploded', 'detonating', 'bomber', 'bombing', 'blast', 'bombings', 'booby-trap', 'landmine', 'v-1', 'hiroshima', 'nuclear', 'laser-guided', 'explodes', 'blasts', 'detonation'	'adpmod:on_pentagon', 'compmodI_splashes', 'adpmod:byI_unharmed', 'adpmod:withI_rearmed', 'compmod_single-car', 'compmod_m1918', 'compmod_phishing', 'compmodI_fellatio', 'rcmod_bursts', 'compmod_decapitation', 'adpmod:over_berlin', 'num_60mm', 'compmod_nega', 'rcmod_crushes', 'adpmod:at_wonsan', 'conjI_hellcat', 'compmod_bvr', 'compmod_econoline', 'compmod_t-55', 'adpmod:against_fortress'

horse

window, k=5 (2)	tree (3)
'horses', 'standardbred', 'thoroughbred', 'trotting', 'riding', 'racing', 'ridden', 'galloping', 'bred', 'gelding', 'stud', 'appaloosa', 'pommel', 'bareback', 'bronc', 'sired', 'thoroughbreds', 'friesian', 'stables', 'aqha'	'nsubjpassI_spooked', 'compmodI_drovers', 'dobjI_shoe', 'adpmod:byI_commute', 'rcmod_roams', 'poss_quixote', 'compmod_ch-53e', 'appos_pony', 'compmod_poitevin', 'adpmod:ofI_hilt', 'adpmod:byI_escapes', 'compmodI_fillet', 'compmodI_cogburn', 'compmod_beardless', 'apposI_bike', 'compmod_m1918', 'poss_quagmire', 'conjI_turban', 'adpmod:byI_foot', 'appos_canadensis'

fox

window, k=5 (2)	tree (3)
'vulpes', 'news', 'movietone', 'owned-and-operated', 'nbc', 'cbs', 'terrier', 'affiliate', 'cnn', 'msnbc', 'abc', 'espn', 'mynetworktv', 'nine-tailed', 'wb', 'wish-tv', 'wtae-tv', 'blitzer', 'marten', 'cw'	'appos_canadensis', 'compmodI_subchannels', 'conj_mattel', 'compmod_jacki', 'conjI_raimi', 'conjI_corgan', 'conjI_leda', 'conj_bluebird', 'conj_nolte', 'conjI_browder', 'conj_keeler', 'conjI_steinem', 'conjI_thurmond', 'conjI_melly', 'conjI_seaver', 'compmodI_e-type', 'apposI_msnbc', 'conjI_devitt', "conj_o'hare", 'compmod_rayford'

table

window, k=5 (2)	tree (3)
'sortable', 'tables', 'tennis', 'lookup', 'billiards', 'foosball', 'periodic', 'billiard', 'toc', 'look-up', 'hash', 'collapsible', 'snooker',	'adpmod:offI_knocks', 'adpmod:onI_seventh', 'amod_ten-foot', 'compmod_abydos', 'amod_25-metre', 'adpmod:intoI_face',

'below', 'round', 'carom', 'contents', 'tabular', 'bottom', 'headers'	'adpmod:belowI_added', 'adpmod:throughI_throw', 'compmod_three-judge', 'adpmod:ontoI_projects', 'adpmod:aroundI_looked', 'adpmod:onI_substance', 'adpmod:onI_slides', 'amod_rooftop', 'adpmod:insideI_placing', 'adpmod:withinI_display', 'adpmod:onI_entities', 'dobjI_clogs', 'adpmod:aboveI_from', 'adpmod:onI_measurement'

bowl

window, k=5 (2)	tree (3)
'xli', 'bowls', 'xliv', 'xxxviii', 'xliii', 'xlii', 'xlvi', 'xlv', 'super', 'xxxix', 'xxxv', 'xxxi', 'xxxvii', 'xxxvi', 'xlvii', 'xxxiii', 'xxv', 'xxvii', 'xxviii', 'xxxiv'	'conjI_preseason', 'adpmod:against_oklahoma', 'adpmod:inI_2-1', 'conj_play-offs', 'compmod_pan-pacific', 'adpmod:forI_stadiums', 'adpmod:atI_clinched', 'adpmod:intoI_win', 'amod_25-metre', 'conjI_regionals', 'amod_sm-liiga', 'adpmod:against_brentford', 'adpmod:ofI_tackle', 'poss_postman', 'adpmod:inI_semifinal', 'adpmod:against_walsall', 'adpmod:atI_pitcher', 'compmod_doomben', 'compmod_deuces', 'compmod_sundae'

guitar

window, k=5 (2)	tree (3)
'bass', 'guitars', 'drums', 'keyboards', 'nyckelharpa', 'glockenspiel', '6-string', 'tambourine', 'banjo', 'vocals', 'harmonica', 'fretless', 'ukulele', '12-string', 'benmont', 'six-string', 'riffs', 'piano', 'accordion', 'melodica'	'amod_end-blown', 'conj_back-up', 'adpmod:onI_harris', 'adpmod:forI_adagio', 'adpmod:onI_thompson', 'appos_mandolin', 'dep_keyboard', 'adpmod:ofI_virtuoso', 'adpmod:onI_foster', 'amod_necked', 'conjI_flinders', 'adpmod:onI_supplemented', 'possI_pickups', 'amod_banjo', 'compmod_sandberg', 'adpmod:ofI_sweethearts', 'poss_diddley', 'depI_trumpet', 'adpmod:onI_pianist', 'dobjI_syncing'

piano

window, k=5 (2)	tree (3)
'violin', 'sonata', 'cello', 'concerto', 'op', 'harpsichord', 'concertos', 'viola',	'amod_end-blown', 'conjI_bongos', 'adpmod:onI_harris', 'adpmod:ofI_virtuoso',

'violoncello', 'sonatas', 'divertimento', 'pianoforte', 'e-flat', 'concerti', 'concertante', 'clarinet', 'sonatina', 'bassoon', 'nyckelharpa', 'guitar'	'adpmod:onI_thompson', 'conj_back-up', 'compmod_kreutzer', 'compmod_vomeronasal', 'adpmod:onI_supplemented', 'appos_mandolin', 'apposI_prelude', 'adpmod:onI_foster', 'compmod_akiyoshi', 'dep_keyboard', 'compmod_sandberg', 'conjI_kabuki', 'amod_necked', 'dobjI_syncing', 'depI_trumpet', 'amod_12-tone'

Conclusions:

The results of the word2vec – window were really good. However, in the tree results, some of the features did not make sense.

For example, for hospital we got: 'compmod_**safdarjung**' (ruler of an indian state in the 18th century), 'adpmod:of_**nuova**', 'compmod_**strangeways**' (a prison).

The 1st order produced words that are related to the target word and not synonyms like the 2nd order.

MAP results for car and piano:

Both words had a perfect precision in both methods.

AP(car)=AP(piano) = (1+ 1 + 1 + … + 1) / 20 = 1

MAP(window) = MAP(tree) = avg(AP(car), AP(piano))=1

7. How to run our code:

The files:
part1.py – questions 1-3
word2vec.py – question 6

Modules: docopt, numpy

Arguments: both files require arguments to run.
Run "filename.py -h" for details.