Lecture 7: Distributional semantics (continued)

Distributions as usage representations

Distributional word clustering

Selectional preferences
Outline.

Distributions as usage representations

Distributional word clustering

Selectional preferences
Distributional methods are a usage representation

- Distributions are a good conceptual representation if you believe that ‘the meaning of a word is given by its usage’.
- Corpus-dependent, culture-dependent, register-dependent.
  Example: similarity between *policeman* and *cop*: 0.23
### Distribution for *policeman*

<table>
<thead>
<tr>
<th>Term</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>ball_n+poss_rel</td>
<td>0.59</td>
</tr>
<tr>
<td>and_c+civilian_n</td>
<td>0.48</td>
</tr>
<tr>
<td>soldier_n+and_c</td>
<td>0.42</td>
</tr>
<tr>
<td>and_c+soldier_n</td>
<td>0.41</td>
</tr>
<tr>
<td>secret_a</td>
<td>0.38</td>
</tr>
<tr>
<td>people_n+include_v</td>
<td>0.37</td>
</tr>
<tr>
<td>corrupt_a</td>
<td>0.36</td>
</tr>
<tr>
<td>uniformed_a</td>
<td>0.35</td>
</tr>
<tr>
<td>uniform_n+poss_rel</td>
<td>0.35</td>
</tr>
<tr>
<td>civilian_n+and_c</td>
<td>0.35</td>
</tr>
<tr>
<td>iraqi_a</td>
<td>0.31</td>
</tr>
<tr>
<td>lot_n+poss_rel</td>
<td>0.31</td>
</tr>
<tr>
<td>chechen_a</td>
<td>0.31</td>
</tr>
<tr>
<td>laugh_v</td>
<td>0.30</td>
</tr>
<tr>
<td>and_c+criminal_n</td>
<td>0.29</td>
</tr>
<tr>
<td>incompetent_a</td>
<td>0.28</td>
</tr>
<tr>
<td>pron_rel+_shoot_v</td>
<td>0.28</td>
</tr>
<tr>
<td>hat_n+poss_rel</td>
<td>0.28</td>
</tr>
<tr>
<td>terrorist_n+and_c</td>
<td>0.28</td>
</tr>
<tr>
<td>and_c+crowd_n</td>
<td>0.27</td>
</tr>
<tr>
<td>military_a</td>
<td>0.27</td>
</tr>
<tr>
<td>helmet_n+poss_rel</td>
<td>0.27</td>
</tr>
<tr>
<td>father_n+be_v</td>
<td>0.27</td>
</tr>
<tr>
<td>on_p()+duty_n</td>
<td>0.26</td>
</tr>
<tr>
<td>salary_n+poss_rel</td>
<td>0.25</td>
</tr>
<tr>
<td>on_p()+horseback_n</td>
<td>0.25</td>
</tr>
<tr>
<td>armed_a</td>
<td>0.25</td>
</tr>
<tr>
<td>and_c+nurse_n</td>
<td>0.24</td>
</tr>
<tr>
<td>job_n+as_p()</td>
<td>0.24</td>
</tr>
<tr>
<td>open_v+fire_n</td>
<td>0.24</td>
</tr>
</tbody>
</table>
Distribution for *cop*

<table>
<thead>
<tr>
<th>Word</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>crooked_a</td>
<td>0.45</td>
</tr>
<tr>
<td>corrupt_a</td>
<td>0.45</td>
</tr>
<tr>
<td>maniac_a</td>
<td>0.44</td>
</tr>
<tr>
<td>dirty_a</td>
<td>0.38</td>
</tr>
<tr>
<td>honest_a</td>
<td>0.37</td>
</tr>
<tr>
<td>uniformed_a</td>
<td>0.36</td>
</tr>
<tr>
<td>tough_a</td>
<td>0.35</td>
</tr>
<tr>
<td>pron_rel+_call_v</td>
<td>0.33</td>
</tr>
<tr>
<td>funky_a</td>
<td>0.32</td>
</tr>
<tr>
<td>bad_a</td>
<td>0.32</td>
</tr>
<tr>
<td>veteran_a</td>
<td>0.29</td>
</tr>
<tr>
<td>and_c+robot_n</td>
<td>0.29</td>
</tr>
<tr>
<td>and_c+criminal_n</td>
<td>0.28</td>
</tr>
<tr>
<td>bogus_a</td>
<td>0.28</td>
</tr>
<tr>
<td>talk_v+to_p()+pron_rel_</td>
<td>0.28</td>
</tr>
<tr>
<td>investigate_v+murder_n</td>
<td>0.27</td>
</tr>
<tr>
<td>on_p()+force_n</td>
<td>0.26</td>
</tr>
<tr>
<td>parody_n+of_p()</td>
<td>0.25</td>
</tr>
<tr>
<td>Mason_n+and_c</td>
<td>0.25</td>
</tr>
<tr>
<td>pron_rel+_kill_v</td>
<td>0.25</td>
</tr>
<tr>
<td>racist_a</td>
<td>0.25</td>
</tr>
<tr>
<td>addicted_a</td>
<td>0.24</td>
</tr>
<tr>
<td>gritty_a</td>
<td>0.23</td>
</tr>
<tr>
<td>and_c+interference_n</td>
<td>0.23</td>
</tr>
<tr>
<td>arrive_v</td>
<td>0.23</td>
</tr>
<tr>
<td>and_c+detective_n</td>
<td>0.23</td>
</tr>
<tr>
<td>look_v+way_n</td>
<td>0.22</td>
</tr>
<tr>
<td>dead_a</td>
<td>0.22</td>
</tr>
<tr>
<td>pron_rel+_stab_v</td>
<td>0.22</td>
</tr>
<tr>
<td>evade_v</td>
<td>0.21</td>
</tr>
</tbody>
</table>
Distributions and knowledge

What kind of information do distributions encode?

- lexical knowledge
- world knowledge
- cultural biases
- boundaries between them are blurry

Distributions are partial lexical semantic representations, but useful and theoretically interesting.
Outline.

Distributions as usage representations

Distributional word clustering

Selectional preferences
Clustering

- clustering techniques group objects into clusters
- similar objects in the same cluster, dissimilar objects in different clusters
- allows us to obtain generalisations over the data
- widely used in various NLP tasks:
  - semantics (e.g. word clustering);
  - summarization (e.g. sentence clustering);
  - text mining (e.g. document clustering).
Distributional word clustering

We will:

- cluster words based on the contexts in which they occur
- assumption: words with similar meanings occur in similar contexts, i.e. are distributionally similar
- we will consider noun clustering as an example
- cluster 2000 nouns – most frequent in the British National Corpus
- into 200 clusters
Clustering nouns

- truck
- lorry
- bike
- car
- bicycle
- taxi
- driver
- mechanic
- engineer
- plumber
- writer
- journalist
- lab
- building
- house
- shack
- dwelling
- highvay
- way
- street
- road
- avenue
- path
- office
- flat
Clustering nouns
Feature vectors

- can use different kinds of context as features for clustering
  - window based context
  - parsed or unparsed
  - syntactic dependencies
- different types of context yield different results
- **Example experiment**: use verbs that take the noun as a direct object or a subject as features for clustering
- **Feature vectors**: verb lemmas, indexed by dependency type, e.g. subject or direct object
- **Feature values**: corpus frequencies
## Extracting feature vectors: Examples

<table>
<thead>
<tr>
<th>tree (Dobj)</th>
<th>crop (Dobj)</th>
<th>tree (Subj)</th>
<th>crop (Subj)</th>
</tr>
</thead>
<tbody>
<tr>
<td>85 plant_v</td>
<td>76 grow_v</td>
<td>131 grow_v</td>
<td>78 grow_v</td>
</tr>
<tr>
<td>82 climb_v</td>
<td>44 produce_v</td>
<td>49 plant_v</td>
<td>23 yield_v</td>
</tr>
<tr>
<td>48 see_v</td>
<td>16 harvest_v</td>
<td>40 stand_v</td>
<td>10 sow_v</td>
</tr>
<tr>
<td>46 cut_v</td>
<td>12 plant_v</td>
<td>26 fell_v</td>
<td>9 fail_v</td>
</tr>
<tr>
<td>27 fall_v</td>
<td>10 ensure_v</td>
<td>25 look_v</td>
<td>8 plant_v</td>
</tr>
<tr>
<td>26 like_v</td>
<td>10 cut_v</td>
<td>23 make_v</td>
<td>7 spray_v</td>
</tr>
<tr>
<td>23 make_v</td>
<td>9 yield_v</td>
<td>22 surround_v</td>
<td>7 come_v</td>
</tr>
<tr>
<td>23 grow_v</td>
<td>9 protect_v</td>
<td>21 show_v</td>
<td>6 produce_v</td>
</tr>
<tr>
<td>22 use_v</td>
<td>9 destroy_v</td>
<td>20 seem_v</td>
<td>6 feed_v</td>
</tr>
<tr>
<td>22 round_v</td>
<td>7 spray_v</td>
<td>20 overhang_v</td>
<td>6 cut_v</td>
</tr>
<tr>
<td>20 get_v</td>
<td>7 lose_v</td>
<td>20 fall_v</td>
<td>5 sell_v</td>
</tr>
<tr>
<td>18 hit_v</td>
<td>6 sell_v</td>
<td>19 cut_v</td>
<td>5 make_v</td>
</tr>
<tr>
<td>18 fell_v</td>
<td>6 get_v</td>
<td>18 take_v</td>
<td>5 include_v</td>
</tr>
<tr>
<td>18 bark_v</td>
<td>5 support_v</td>
<td>18 go_v</td>
<td>5 harvest_v</td>
</tr>
<tr>
<td>17 want_v</td>
<td>5 see_v</td>
<td>18 become_v</td>
<td>4 follow_v</td>
</tr>
<tr>
<td>16 leave_v</td>
<td>5 raise_v</td>
<td>17 line_v</td>
<td>3 ripen_v</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Feature vectors: Examples

**tree**
- 131 grow_v_Subj
- 85 plant_v_Dobj
- 82 climb_v_Dobj
- 49 plant_v_Subj
- 48 see_v_Dobj
- 46 cut_v_Dobj
- 40 stand_v_Subj
- 27 fall_v_Dobj
- 26 like_v_Dobj
- 26 fell_v_Subj
- 25 look_v_Subj
- 23 make_v_Subj
- 23 make_v_Dobj
- 23 grow_v_Dobj
- 22 use_v_Dobj
- 22 surround_v_Subj
- 22 round_v_Dobj
- 20 overhang_v_Subj
- ...

**crop**
- 78 grow_v_Subj
- 76 grow_v_Dobj
- 44 produce_v_Dobj
- 23 yield_v_Subj
- 16 harvest_v_Dobj
- 12 plant_v_Dobj
- 10 sow_v_Subj
- 10 ensure_v_Dobj
- 10 cut_v_Dobj
- 9 yield_v_Dobj
- 9 protect_v_Dobj
- 9 fail_v_Subj
- 9 destroy_v_Dobj
- 8 plant_v_Subj
- 7 spray_v_Subj
- 7 spray_v_Dobj
- 7 lose_v_Dobj
- 6 feed_v_Subj
- ...

...
many clustering algorithms are available
example algorithm: K-means clustering
given a set of \( N \) data points \( \{x_1, x_2, \ldots, x_N\} \)
partition the data points into \( K \) clusters \( C = \{C_1, C_2, \ldots, C_K\} \)
minimize the sum of the squares of the distances of each data point to the cluster mean vector \( \mu_i \):

\[
\arg\min_C \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{1}
\]
K-means clustering

Iteration 1

Iteration 2

Iteration 3

Iteration 4

Iteration 5

Iteration 6
## Noun clusters

<table>
<thead>
<tr>
<th>Tree</th>
<th>Crop</th>
<th>Flower</th>
<th>Plant</th>
<th>Root</th>
<th>Leaf</th>
<th>Seed</th>
<th>Rose</th>
<th>Wood</th>
<th>Grain</th>
<th>Stem</th>
<th>Forest</th>
<th>Garden</th>
</tr>
</thead>
<tbody>
<tr>
<td>consent</td>
<td>permission</td>
<td>concession</td>
<td>injunction</td>
<td>licence</td>
<td>approval</td>
<td>lifetime</td>
<td>quarter</td>
<td>period</td>
<td>century</td>
<td>succession</td>
<td>stage</td>
<td>generation</td>
</tr>
<tr>
<td>phase</td>
<td>interval</td>
<td>future</td>
<td>subsidy</td>
<td>compensation</td>
<td>damages</td>
<td>allowance</td>
<td>payment</td>
<td>pension</td>
<td>grant</td>
<td>carriage</td>
<td>bike</td>
<td>vehicle</td>
</tr>
<tr>
<td>train</td>
<td>truck</td>
<td>lorry</td>
<td>coach</td>
<td>taxi</td>
<td>official</td>
<td>officer</td>
<td>inspector</td>
<td>journalist</td>
<td>detective</td>
<td>constable</td>
<td>police</td>
<td>policeman</td>
</tr>
<tr>
<td>reporter</td>
<td>girl</td>
<td>other</td>
<td>woman</td>
<td>child</td>
<td>person</td>
<td>people</td>
<td>length</td>
<td>past</td>
<td>mile</td>
<td>metre</td>
<td>distance</td>
<td>inch</td>
</tr>
<tr>
<td>yard</td>
<td>tide</td>
<td>breeze</td>
<td>flood</td>
<td>wind</td>
<td>rain</td>
<td>storm</td>
<td>weather</td>
<td>wave</td>
<td>current</td>
<td>heat</td>
<td>sister</td>
<td>daughter</td>
</tr>
<tr>
<td>parent</td>
<td>relative</td>
<td>lover</td>
<td>cousin</td>
<td>friend</td>
<td>wife</td>
<td>mother</td>
<td>husband</td>
<td>brother</td>
<td>father</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Clustering nouns

car  bicycle  bike  taxi  lorry  driver  mechanic  plumber  engineer  writer  scientist  journalist  truck  proceedings  journal  book  newspaper  lab  office  building  shack  house  flat  dwelling  highway  road  avenue  street  path  way  dwelling

distributional word clustering
Clustering nouns

car — bicycle — bike — taxi — lorry — driver — mechanic — plumber — engineer — writer — scientist — journalist

highway — way — street — house — path — shack — flat — dwelling — road — avenue — street — way — path
We can also cluster verbs...

<table>
<thead>
<tr>
<th>sparkle</th>
<th>glow</th>
<th>widen</th>
<th>flash</th>
<th>flare</th>
<th>gleam</th>
<th>darken</th>
<th>narrow</th>
<th>flicker</th>
<th>shine</th>
<th>blaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>bulge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gulp</td>
<td>drain</td>
<td>stir</td>
<td>empty</td>
<td>pour</td>
<td>sip</td>
<td>spill</td>
<td>swallow</td>
<td>drink</td>
<td>pollute</td>
<td>seep</td>
</tr>
<tr>
<td>purify</td>
<td>ooze</td>
<td>pump</td>
<td>bubble</td>
<td>splash</td>
<td>ripple</td>
<td>simmer</td>
<td>boil</td>
<td>tread</td>
<td></td>
<td></td>
</tr>
<tr>
<td>polish</td>
<td>clean</td>
<td>scrape</td>
<td>scrub</td>
<td>soak</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kick</td>
<td>hurl</td>
<td>push</td>
<td>fling</td>
<td>throw</td>
<td>pull</td>
<td>drag</td>
<td>haul</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rise</td>
<td>fall</td>
<td>shrink</td>
<td>drop</td>
<td>double</td>
<td>fluctuate</td>
<td>dwindle</td>
<td>decline</td>
<td>plunge</td>
<td>decrease</td>
<td>soar</td>
</tr>
<tr>
<td>initiate</td>
<td>inhibit</td>
<td>aid</td>
<td>halt</td>
<td>trace</td>
<td>track</td>
<td>speed</td>
<td>obstruct</td>
<td>impede</td>
<td>accelerate</td>
<td>slow</td>
</tr>
<tr>
<td>work</td>
<td>escape</td>
<td>fight</td>
<td>head</td>
<td>ride</td>
<td>fly</td>
<td>arrive</td>
<td>travel</td>
<td>come</td>
<td>run</td>
<td>go</td>
</tr>
</tbody>
</table>
Hard and soft clustering

- **Hard clustering**: each data point assigned to one cluster only (as in our k-means experiment)
- **Soft clustering**: each data point is associated with multiple clusters with a membership probability
Uses of word clustering in NLP

- Noun and verb clustering are typically used in NLP for lexical acquisition tasks
  - to automatically create large-scale lexical resources to support other NLP tasks
  - e.g. selectional preferences
- Dimensionality reduction for other statistical models
  - to reduce negative effects of sparse data
Outline.

Distributions as usage representations

Distributional word clustering

Selectional preferences
Subcategorization and selection

Subcategorization:
- different predicates (e.g. verbs) can take a different number and types of arguments (syntactic)

Selection:
- different predicates place different semantic constraints on the types of arguments that they can take
Selectional preferences

Subcategorization

- verbs (and other types of words) have different numbers and types of syntactic arguments:

  * Kim adored
  * Kim gave Sandy
  * Kim adored to sleep
  * Kim liked to sleep
  * Kim devoured
  * Kim ate

- verbs can be intransitive (1 argument: subj); transitive (2 arguments: subj and obj); ditransitive (3 arguments: subj and 2 obj).
Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate’s argument slot than others.

The authors wrote a new paper.

Watch out, the cat is eating your sausage!

*The carrot ate the keys.

*The law sang a driveway.
Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate’s argument slot than others.

The authors wrote a new paper.

Watch out, the cat is eating your sausage!

*The carrot ate the keys.

*The law sang a driveway.
Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate’s argument slot than others.

The authors wrote a new paper.

Watch out, the cat is eating your sausage!

*The carrot ate the keys.

*The law sang a driveway.
Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate’s argument slot than others.

The authors wrote a new paper.
Watch out, the cat is eating your sausage!
*The carrot ate the keys.
*The law sang a driveway.
Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate’s argument slot than others.

The authors wrote a new paper.

Watch out, the cat is eating your sausage!

*The carrot ate the keys.

*The law sang a driveway.
Learning selectional preferences from distributions

[animate] eat [food]
[person] sing [song]
[person] read [book]

1. Need to define a set of argument classes that can fill the argument slot of the predicate
   e.g. use noun clusters for this purpose
2. Need to quantify the level of association of a particular verb with a particular noun class
Selectional preference model


Selectional preference strength

\[ S_R(v) = D_{KL}(P(c|v) || P(c)) = \sum_c P(c|v) \log \frac{P(c|v)}{P(c)} \]

\( D_{KL} \) is Kullback–Leibler divergence

Selectional association

\[ A_R(v, c) = \frac{1}{S_R(v)} P(c|v) \log \frac{P(c|v)}{P(c)} \]

\( P(c) \) is the prior probability of the noun class;
\( P(c|v) \) its posterior probability given the verb; \( R \) is the grammatical relation
Calculating probabilities

\[ P(c) = \frac{f(c)}{\sum_k f(c_k)}, \]

\[ P(c|v) = \frac{f(v, c)}{f(v)}, \]

\[ f(c) = \sum_{n_i \in c} f(n_i) \]

- \( f(v, c) \): frequency of verb \( v \) co-occurring with the noun class \( c \)
- \( f(v) \): total frequency of verb \( v \) with all noun classes
- \( f(c) \): total frequency of the noun class \( c \)
Selectional preferences of kill (Dobj)

0.38 girl other woman child person people
0.20 being species sheep animal creature horse baby human fish male lamb
bird rabbit female insect cattle mouse monster
0.19 sister daughter parent relative lover cousin friend wife mother husband
brother father
0.04 thousand citizen inhabitant resident minority youngster refugee peasant
miner hundred
0.0378 gene tissue cell particle fragment bacterium protein acid complex
compound molecule organism
0.0336 fleet soldier knight force rebel guard troops crew army pilot
0.0335 official officer inspector journalist detective constable police
policeman reporter
0.0322 victim bull teenager prisoner hero gang enemy rider offender youth
killer thief driver defender hell
0.0136 week month year
...
Selectional preferences of *drink* (Dobj)

0.5831 drink coffee champagne pint wine beer
0.2778 drop tear sweat paint blood water juice
0.1084 mixture salt dose ingredient sugar substance drug milk cream alcohol fibre chemical
0.0515 brush bowl bucket receiver barrel dish glass container plate basket bottle tray
0.0069 couple minute night morning hour time evening afternoon
0.0041 stability efficiency security prospects health welfare survival safety
0.0025 recording music tape song tune radio guitar trick album football organ stuff
0.0005 rage excitement panic anger terror flame laughter
0.0004 ball shot kick arrow stroke bullet punch bomb shell blow missile
0.0003 lunch dinner breakfast meal
...

Selectional preferences of *drink* (Dobj)
Different senses of *run*

The children *ran* to the store
If you see this man, *run*!
Service *runs* all the way to Cranbury
She is *running* a relief operation in Sudan
the story or argument *runs* as follows
Does this old car still *run* well?
Interest rates *run* from 5 to 10 percent
Who’s *running* for treasurer this year?
They *ran* the tapes over and over again
These dresses *run* small
Selectional preferences of *run* (Subj)

0.2125 drop tear sweat paint blood water juice
0.1665 technology architecture program system product version interface software tool computer network processor chip package
0.1657 tunnel road path trail lane route track street bridge
0.1166 carriage bike vehicle train truck lorry coach taxi
0.0919 tide breeze flood wind rain storm weather wave current heat
0.0865 tube lock tank circuit joint filter battery engine device disk furniture machine mine seal equipment machinery wheel motor slide disc instrument
0.0792 ocean canal stream bath river waters pond pool lake
0.0497 rope hook cable wire thread ring knot belt chain string
0.0469 arrangement policy measure reform proposal project programme scheme plan course
0.0352 week month year
0.0351 couple minute night morning hour time evening afternoon
Selectional preferences of *run* (continued)

0.0341 criticism appeal charge application allegation claim objection suggestion case complaint
0.0253 championship open tournament league final round race match competition game contest
0.0218 desire hostility anxiety passion doubt fear curiosity enthusiasm impulse instinct emotion feeling suspicion
0.0183 expenditure cost risk expense emission budget spending
0.0136 competitor rival team club champion star winner squad county player liverpool partner leeds
0.0102 being species sheep animal creature horse baby human fish male lamb bird rabbit female insect cattle mouse monster
...
Selectional preferences of *cut* (Dobj)

0.2845 expenditure cost risk expense emission budget spending
0.1527 dividend price rate premium rent rating salary wages
0.0832 employment investment growth supplies sale import export
production consumption traffic input spread supply flow
0.0738 potato apple slice food cake meat bread fruit
0.0407 stitch brick metal bone strip cluster coffin stone piece tile fabric rock
layer remains block
0.0379 excess deficit inflation unemployment pollution inequality poverty
delay discrimination symptom shortage
0.0366 tree crop flower plant root leaf seed rose wood grain stem forest
garden
0.0330 tail collar strand skirt trousers hair curtain sleeve
0.0244 rope hook cable wire thread ring knot belt chain string
...

Selectional preferences

**Uses of selectional preferences**

Widely used in NLP as a source of lexical information:

- Word sense induction and disambiguation
- Parsing (resolving ambiguous attachments)
- Identifying figurative language and idioms
- Paraphrasing and paraphrase detection
- Natural language inference (e.g. in the entailment identification task)
- etc.

We looked at verbs, other parts of speech (e.g. adjectives) can have preferences too.