What is distributional semantics?

- A corpus-based approach to the representation of meaning based on a very simple intuition: **distributional hypothesis**
  
  - similar context $\iff$ similar meaning

- An empirical method that produces usage-based lexical entries for words, which to the computer look like this:
  
  $\begin{pmatrix}
  10, 0, 0, 0, 100, 40 \\
  -1.3, 1.4, 0.4, -0.2, 1.3, 2.7, -0.001
  \end{pmatrix}$

- Closely related to neuronal word embeddings

- Maths behind it can be complicated . . .
  
  . . . but you can apply DS to many research questions with existing software packages if you understand the basic concepts clearly
  
  - Beware of the black box problem!

Goals of this course

1. Introduce the basic concepts of **distributional semantics** (DS) and – at the same time – teach you to take your own steps into DS with the *wordspace* package for R

2. Show you **what can be done** with DS in two domains of interdisciplinary application, including hands-on exercises
   
   - Linguistic Theory
     
     - Motivation: test theories, enlarge scope of investigation
     
     - Challenge: operationalization
       
       (theoretical concepts $\rightarrow$ empirical properties)
   
   - Cognitive modeling
     
     - Motivation: corpus data are behavioural data after all
     
     - Challenge: continuous variables, large vocabularies

3. Equip you with the “coordinates” to navigate the current DS literature beyond the scope of this course

Today’s plan

**Introduction**

- The distributional hypothesis
- Distributional semantic models
- DSM and semantic similarity

**Course Outline**

**Getting practical**

- Software and further information
- R as a (toy) laboratory
Introduction

The distributional hypothesis

Outline

Introduction

The distributional hypothesis

Distributional semantic models

DSM and semantic similarity

Course Outline

Getting practical

Software and further information

R as a (toy) laboratory

Meaning & distribution

▷ “Die Bedeutung eines Wortes liegt in seinem Gebrauch.”
   — Ludwig Wittgenstein

☞ meaning = use = distribution in language

▷ “You shall know a word by the company it keeps!”
   — J. R. Firth (1957)

☞ distribution = collocations = habitual word combinations

▷ Distributional hypothesis: difference of meaning correlates with difference of distribution (Zellig Harris 1954)

☞ semantic distance

▷ “What people know when they say that they know a word is not how to recite its dictionary definition – they know how to use it [...] in everyday discourse.” (Miller 1986)

What is the meaning of “bardivac”?

Can we infer meaning from usage?

▷ He handed her her glass of bardivac.

▷ Beef dishes are made to complement the bardivacs.

▷ Nigel staggered to his feet, face flushed from too much bardivac.

▷ Malbec, one of the lesser-known bardivac grapes, responds well to Australia’s sunshine.

▷ I dined off bread and cheese and this excellent bardivac.

▷ The drinks were delicious: blood-red bardivac as well as light, sweet Rhenish.

bardivac is a heavy red alcoholic beverage made from grapes

All examples from British National Corpus (handpicked and slightly edited).

Word sketch of “cat”

Can we infer meaning from collocations (as Firth suggests)?

https://the.sketchengine.co.uk/
A thought experiment: deciphering hieroglyphs

<table>
<thead>
<tr>
<th></th>
<th>knife</th>
<th>cat</th>
<th>dog</th>
<th>cup</th>
<th>pig</th>
<th>banana</th>
</tr>
</thead>
<tbody>
<tr>
<td>get</td>
<td>51</td>
<td>52</td>
<td>115</td>
<td>98</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>sij</td>
<td>20</td>
<td>58</td>
<td>83</td>
<td>14</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>ius</td>
<td>84</td>
<td>4</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>hir</td>
<td>0</td>
<td>4</td>
<td>42</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>iit</td>
<td>3</td>
<td>6</td>
<td>33</td>
<td>1</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>kil</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>27</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ \text{sim}(\text{knife}, \text{cat}) = 0.770 \]
\[ \text{sim}(\text{dog}, \text{pig}) = 0.939 \]
\[ \text{sim}(\text{boat}, \text{cup}) = 0.961 \]

A thought experiment: deciphering hieroglyphs

<table>
<thead>
<tr>
<th></th>
<th>knife</th>
<th>cat</th>
<th>boat</th>
<th>cup</th>
<th>pig</th>
<th>banana</th>
</tr>
</thead>
<tbody>
<tr>
<td>get</td>
<td>51</td>
<td>52</td>
<td>59</td>
<td>98</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>sij</td>
<td>20</td>
<td>58</td>
<td>39</td>
<td>14</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>ius</td>
<td>84</td>
<td>4</td>
<td>23</td>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>hir</td>
<td>0</td>
<td>4</td>
<td>42</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>iit</td>
<td>3</td>
<td>6</td>
<td>33</td>
<td>1</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>kil</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>27</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ \text{sim}(\text{cat}, \text{boat}) = 0.961 \]
Introduction

The distributional hypothesis

English as seen by the computer...

<table>
<thead>
<tr>
<th></th>
<th>get</th>
<th>see</th>
<th>use</th>
<th>hear</th>
<th>eat</th>
<th>kill</th>
</tr>
</thead>
<tbody>
<tr>
<td>knife</td>
<td>51</td>
<td>20</td>
<td>84</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>cat</td>
<td>52</td>
<td>58</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td>dog</td>
<td>115</td>
<td>83</td>
<td>10</td>
<td>42</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td>boat</td>
<td>59</td>
<td>39</td>
<td>23</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cup</td>
<td>98</td>
<td>14</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>pig</td>
<td>12</td>
<td>17</td>
<td>3</td>
<td>2</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>banana</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>18</td>
<td>0</td>
</tr>
</tbody>
</table>

verb-object counts from British National Corpus

Geometric interpretation

- row vector $\mathbf{x}_{\text{dog}}$ describes usage of word dog in the corpus
- can be seen as coordinates of point in $n$-dimensional Euclidean space
- illustrated for two dimensions: get and use
- $\mathbf{x}_{\text{dog}} = (115,10)$

co-occurrence matrix $\mathbf{M}$

- similarity = spatial proximity (Euclidean dist.)
- location depends on frequency of noun ($f_{\text{dog}} \approx 2.7 \cdot f_{\text{cat}}$)

Two dimensions of English V–Obj DSM

- $d = 63.3$ for dog–cat
- $d = 57.5$ for boat–cat

© Evert/Lapesa/Lenci/Baroni (CC-by-sa)

DSM Tutorial – Part 1
wordspace.collocations.de 11/60
Geometric interpretation

- Vector can also be understood as an arrow from origin.
- Direction is more important than location.
- Use angle $\alpha$ as a distance measure.
- Or normalise length $\|x_{dog}\|$ of arrow.

Two dimensions of English V–Obj DSM

$\alpha = 54.3^\circ$

General definition of DSMs

A distributional semantic model (DSM) is a scaled and/or transformed co-occurrence matrix $M$, such that each row $x$ represents the distribution of a target term across contexts.

<table>
<thead>
<tr>
<th>Term</th>
<th>get</th>
<th>see</th>
<th>use</th>
<th>hear</th>
<th>eat</th>
<th>kill</th>
</tr>
</thead>
<tbody>
<tr>
<td>knife</td>
<td>0.027</td>
<td>-0.024</td>
<td>0.206</td>
<td>-0.022</td>
<td>-0.044</td>
<td>-0.042</td>
</tr>
<tr>
<td>cat</td>
<td>0.031</td>
<td>0.143</td>
<td>-0.243</td>
<td>-0.015</td>
<td>-0.009</td>
<td>0.131</td>
</tr>
<tr>
<td>dog</td>
<td>-0.026</td>
<td>0.021</td>
<td>-0.212</td>
<td>0.064</td>
<td>0.013</td>
<td>0.014</td>
</tr>
<tr>
<td>boat</td>
<td>-0.022</td>
<td>0.009</td>
<td>-0.044</td>
<td>-0.040</td>
<td>-0.074</td>
<td>-0.042</td>
</tr>
<tr>
<td>cup</td>
<td>-0.014</td>
<td>-0.173</td>
<td>-0.249</td>
<td>-0.099</td>
<td>-0.119</td>
<td>-0.042</td>
</tr>
<tr>
<td>pig</td>
<td>-0.069</td>
<td>0.094</td>
<td>-0.158</td>
<td>0.000</td>
<td>0.094</td>
<td>0.265</td>
</tr>
<tr>
<td>banana</td>
<td>0.047</td>
<td>-0.139</td>
<td>-0.104</td>
<td>-0.022</td>
<td>0.267</td>
<td>-0.042</td>
</tr>
</tbody>
</table>

Term = word, lemma, phrase, morpheme, word pair, …
Nearest neighbours

DSM based on verb-object relations from BNC, reduced to 100 dim. with SVD

Neighbours of **trousers** (cosine angle):
- shirt (18.5), blouse (21.9), scarf (23.4), jeans (24.7), skirt (25.9), sock (26.2), shorts (26.3), jacket (27.8), glove (28.1), coat (28.8), cloak (28.9), hat (29.1), tunic (29.3), overcoat (29.4), pants (29.8), helmet (30.4), apron (30.5), robe (30.6), mask (30.8), tracksuit (31.0), jersey (31.6), shawl (31.6), ...

Neighbours of **rage** (cosine angle):
- anger (28.5), fury (32.5), sadness (37.0), disgust (37.4), emotion (39.0), jealousy (40.0), grief (40.4), irritation (40.7), revulsion (40.7), scorn (40.7), panic (40.8), bitterness (41.6), resentment (41.8), indignation (41.9), excitement (42.0), hatred (42.5), envy (42.8), disappointment (42.9), ...

Semantic maps

**Semantic map (V−Obj from BNC)**

Clustering

**Clustering of concrete nouns (V−Obj from BNC)**
DSM vectors as word embeddings

DSM vector as sub-symbolic meaning representation
- feature vector for machine learning algorithm
- input for neural network
- such distributed representations are known as **embeddings**

Computation in semantic space
- find meaningful subdimensions in DSM space (⇒ correlation)
- linear operations on vectors

Outline

Introduction
- The distributional hypothesis
- Distributional semantic models

DSM and semantic similarity

Getting practical
- Software and further information
- R as a (toy) laboratory

Distributional similarity as semantic similarity

- DSM similarity as a **quantitative notion**
  - if \(a\) is closer to \(b\) than to \(c\) in the distributional vector space, then \(a\) is more semantically similar to \(b\) than to \(c\)

- DSM similarity as a **graded notion**, differently from categorical nature of most theoretical accounts

- DSM similarity as the empirical correlate of a heterogeneous set of phenomena
  - ... which we may want to tease apart!

- DSM similarity is **symmetric** – cognition is not
  - ... can we fix this?
Characterizing DSM similarity

- DSMs are thought to represent taxonomic similarity
  - words that tend to occur in the same contexts
- Words that share many contexts share many properties (attributes) and are thus taxonomically/ontologically similar
  - synonyms (rhino/rhinoceros)
  - antonyms and values on a scale (good/bad)
  - co-hyponyms (rock/jazz)
  - hyper- and hyponyms (rock/basalt)
- Taxonomic similarity is seen as the fundamental semantic relation organising the vocabulary of a language, allowing categorization, generalization and inheritance...

Is distributional similarity just taxonomic?

Manual annotation: what are the properties of car? Humans vs DSM

Attributional similarity vs. Semantic relatedness

- Attributional similarity (← taxonomical) – two words sharing a large number of salient features (attributes)
  - synonymy (car/automobile)
  - co-hyponymy (car/van/truck)
  - hyperonymy (car/vehicle)
  - antonymy (hot/cold)
  - Problem: they are the opposite of similar, and yet...
- Semantic relatedness (Budanitsky & Hirst 2006) – two words semantically associated without necessarily being similar
  - function (car/driver)
  - meronymy (car/tyre)
  - location (car/road)
  - attribute (car/fast)
  - Problem: they are the opposite of similar, and yet...

Why similar in DSMs? They co-occur → share contexts
DSM similarities: terminological coordinates

Attributional vs. Relational Similarity

- **Attributional similarity** (← taxonomical) – two words sharing a large number of salient features (attributes)
  - synonymy (car/automobile)
  - co-hyponymy (car/van/truck)
  - hyperonymy (car/vehicle)

- **Relational similarity** (Turney 2006) – similar relation between pairs of words (analogy)
  - policeman : gun :: teacher : book
  - mason : stone :: carpenter : wood
  - traffic : street :: water : riverbed

  ... textbook example of neural embeddings application

Problem: symmetry in DSM similarity

The symmetry assumption does not fit all phenomena

**Solution: neighbor rank can capture (potential) asymmetries**

- Motivation: cognitive processes are notoriously asymmetric
- Advantage: rank makes similarity predictions comparable across models and is applicable to different distance measures
- Interpretation: rank controls for differences in density in the semantic space

---

**Introduction**

Course Outline

- Introduction
  - The distributional hypothesis
  - Distributional semantic models
  - DSM and semantic similarity

- Getting practical
  - Software and further information
  - R as a (toy) laboratory

Day 1: Introduction

Summing up what we learnt

- A DSM is a **matrix**, which contains
  - ... targets: **rows**
  - ... contexts: **columns**
  - ... co-occurrence scores (or fancier versions of co-occurrence) for target/context pairs: **matrix cells**

- The row corresponding to a target **vector** is the best **approximation** we have for it its meaning
  - **Goal**: make **comparisons** (recall the hieroglyphs)
    - **Similarity** as context overlap

- Geometric interpretation: vectors as coordinates in space
  - **Similarity** as distance
  - Neighbors reveal the semantic nuances a DSM is capturing
  - Visualization: neighbor maps
  - Neighbor rank as a way to get asymmetric similarity predictions

---

© Evert/Lapesa/Lenci/Baroni (CC-by-sa) DSM Tutorial – Part 1 wordspace.collocations.de 32/60

© Evert/Lapesa/Lenci/Baroni (CC-by-sa) DSM Tutorial – Part 1 wordspace.collocations.de 33/60

© Evert/Lapesa/Lenci/Baroni (CC-by-sa) DSM Tutorial – Part 1 wordspace.collocations.de 34/60

© Evert/Lapesa/Lenci/Baroni (CC-by-sa) DSM Tutorial – Part 1 wordspace.collocations.de 35/60
Roadmap: First steps in distributional semantics

▶ Day 2: Building a DSM, step by step
  ▶ DSM parameters: formal definition & taxonomy
  ▶ Collecting co-occurrence data: what counts as a context?
  ▶ Mathematical operations on the DSM vectors
  ▶ Computing distances/similarities
  ▶ Practice: building DSMs and exploring parameters

▶ Day 3: Which meaning is a DSM capturing (if any?)
  ▶ Evaluation: conceptual coordinates
  ▶ Standard evaluation tasks: multiple choice, prediction of similarity ratings, clustering
  ▶ Narrowing down similarity: classifying semantic relations
  ▶ Practice: evaluation of selected tasks

Roadmap: Interdisciplinary applications

▶ Day 4: DS beyond NLP – Linguistic theory
  ▶ Linguistic exploitation of distributional representations
  ▶ A textbook challenge for DSMs: polysemy
  ▶ Success stories: semantic compositionality (belown and above word level), morphological transparency, argument structure
  ▶ Issues: not all words have a (straightforward) DS meaning
  ▶ Practice: word sense disambiguation & modeling of morphological derivation

▶ Day 5: DS beyond NLP – Cognitive modelling
  ▶ DSMs for cognitive modeling: general issues
  ▶ Free association norms as a window into the organization of the mental lexicon
  ▶ Predicting free associations with DSMs
  ▶ Practice: combine DSMs with first-order co-occurrence in the FAST free association task
Recent workshops and tutorials

- 2007: CoSMo Workshop (at Context ’07)
- 2008: ESSLLI Wshp & Shared Task, Italian J of Linguistics
- 2009: GeMS Wshp (EACL), DiSCo Wshp (CogSci), ESSLLI
- 2010: 2nd GeMS (ACL), ESSLLI Wshp, Tutorial (NAACL), J Natural Language Engineering
- 2011: 2nd DiSCo (ACL), 3rd GeMS (EMNLP)
- 2012: DiDaS Wshp (ICSC), ESSLLI Course
- 2013: CVSC Wshp (ACL), TFDS Wshp (IWCS), Dagstuhl
- 2014: 2nd CVSC (EACL), DSM Wshp (Insight)
- 2015: VSM4NLP (NAACL), ESSLLI Course, TAL Journal
- 2016: DSALT Wshp (ESSLLI), Tutorial (COLING), Tutorial (Konvens), ESSLLI Course, Computational Linguistics
- 2017: ESSLLI Course
- 2018: Tutorial (LREC), ESSLLI Course1 & Course2

click on Workshop name to open Web page

Software packages

<table>
<thead>
<tr>
<th>Package</th>
<th>Language</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infomap NLP</td>
<td>C</td>
<td>classical LSA-style DSM</td>
</tr>
<tr>
<td>HiDEx</td>
<td>C++</td>
<td>re-implementation of the HAL model (Lund &amp; Burgess 1996)</td>
</tr>
<tr>
<td>SemanticVectors</td>
<td>Java</td>
<td>scalable architecture based on random indexing representation</td>
</tr>
<tr>
<td>S-Space</td>
<td>Java</td>
<td>complex object-oriented framework</td>
</tr>
<tr>
<td>JobimText</td>
<td>Java</td>
<td>UIMA / Hadoop framework</td>
</tr>
<tr>
<td>Gensim</td>
<td>Python</td>
<td>complex framework, focus on parallelization and out-of-core algorithms</td>
</tr>
<tr>
<td>Vecto</td>
<td>Python</td>
<td>user-friendly, designed for research on compositional semantics</td>
</tr>
<tr>
<td>DISSECT</td>
<td>Python</td>
<td></td>
</tr>
<tr>
<td>wordspace</td>
<td>R</td>
<td>interactive research laboratory, but scales to real-life data sets</td>
</tr>
<tr>
<td>text2vec</td>
<td>R</td>
<td>GloVe embeddings &amp; topic models</td>
</tr>
</tbody>
</table>

click on package name to open Web page

Further information

- Handouts & other materials available from wordspace wiki at http://wordspace.collocations.de/
- based on joint work with Marco Baroni and Alessandro Lenci
- Tutorial is open source (CC), and can be downloaded from http://r-forge.r-project.org/projects/wordspace/
- We should be working on a textbook Distributional Semantics for Synthesis Lectures on HLT (Morgan & Claypool)

Outline

- Introduction
- The distributional hypothesis
- Distributional semantic models
- DSM and semantic similarity
- Course Outline

Getting practical

- Software and further information
- R as a (toy) laboratory
Prepare to get your hands dirty...

We will use the statistical programming environment R as a toy laboratory in this tutorial but one that scales to real-life applications.

Software installation

- R version 4.0 or newer from http://www.r-project.org/
- RStudio from http://www.rstudio.com/
- R packages from CRAN (through RStudio menu)
  - recommended: e1071, text2vec, Rtsne, uwot
  - optional: tm, quanteda, data.table, wordcloud, shiny, spacyr, udpipe, coreNLP

- Get data sets, precompiled DSMs and wordspaceEval package (with some non-public data sets) from http://wordspace.collocations.de/doku.php/course:material

First steps in R

Start each session by loading the wordspace package.

```r
> library(wordspace)
```

The package includes various example data sets, some of which should look familiar to you.

```r
> DSM_HieroglyphsMatrix
get see use hear eat kill
knife 51 20 84 0 3 0
cat 52 58 4 4 6 26
dog 115 83 10 42 33 17
boat 59 39 23 4 0 0
cup 98 14 6 2 1 0
pig 12 17 3 2 9 27
banana 11 2 2 0 18 0
```
**Term-term matrix** records co-occurrence frequencies with feature terms for each target term.

```
> DSM_TermTermMatrix
```

<table>
<thead>
<tr>
<th></th>
<th>cat</th>
<th>dog</th>
<th>animal</th>
<th>time</th>
<th>reason</th>
<th>cause</th>
<th>effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>breed</td>
<td>83</td>
<td>13</td>
<td>10</td>
<td>19</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>tail</td>
<td>37</td>
<td>13</td>
<td>29</td>
<td>81</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>feed</td>
<td>1</td>
<td>10</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>-2</td>
<td>1</td>
</tr>
<tr>
<td>kill</td>
<td>-1</td>
<td>4</td>
<td>19</td>
<td>1</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>important</td>
<td>1</td>
<td>1</td>
<td>14</td>
<td>14</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>explain</td>
<td>-1</td>
<td>34</td>
<td>40</td>
<td>34</td>
<td>-1</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>likely</td>
<td>-1</td>
<td>1</td>
<td>34</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>17</td>
</tr>
</tbody>
</table>

**Term-context matrix** records frequency of term in each individual context (e.g. sentence, document, Web page, encyclopaedia article).

```
> DSM_TermContextMatrix
```

<table>
<thead>
<tr>
<th></th>
<th>cat</th>
<th>dog</th>
<th>animal</th>
<th>time</th>
<th>reason</th>
<th>cause</th>
<th>effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>dog</td>
<td>-10</td>
<td>4</td>
<td>11</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>animal</td>
<td>2</td>
<td>15</td>
<td>10</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>time</td>
<td>-1</td>
<td>-2</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>reason</td>
<td>1</td>
<td>-2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>cause</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>effect</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

**Playing with a larger model**

**Term-term matrix, dimensionality-reduced**, built from Web texts for target words in the format *lemma_POS* (e.g. *banana_N*).

```
> DSM_Vectors
> View(DSM_Vectors)
```

Let’s inspect some nearest neighbors:

```
> nearest.neighbours(DSM_Vectors, "banana_N", n=4)
coconut_N pineapple_N watermelon_N bean_N
10.86118 12.60826 13.35160 13.79671
```

```
> nearest.neighbours(DSM_Vectors, "freedom_N", n=4)
peace_N morality_N equality_N conscience_N
30.13420 34.18397 34.23418 34.23894
```

Or create a semantic map for a word we are interested in:

```
> plot(nearest.neighbours(DSM_Vectors, "freedom_N", n=20, dist.matrix=TRUE))
```
Getting practical: R as a (toy) laboratory

... and with an even larger model

You can download several large pre-compiled DSMs from the course wiki, which represent different parameters of the co-occurrence matrix (⇒ part 2).

- e.g. WP500_DepFilter_Lemma.rda
- download this file to subdirectory models

```r
> load("models/WP500_DepFilter_Lemma.rda", verbose=TRUE)
Loading objects:
  WP500_DepFilter_Lemma
> model <- WP500_DepFilter_Lemma # assign to a shorter name

Now try the semantic map again:

```r
> plot(nearest.neighbours(model, "freedom_N", n=20, dist.matrix=TRUE))
```

Freedom in a neural embedding model: word2vec

```r
> load("GoogleNews300_wf200k.rda", verbose=TRUE)
> embeddings <- GoogleNews300_wf200k.rda
> plot(nearest.neighbours(embeddings, "freedom_N", n=20, dist.matrix=TRUE))
```

Bonus: Recreate the hieroglyphs example

```r
# apply log-transformation to de-skew co-occurrence frequencies
M <- log2(DSM_HieroglyphsMatrix + 1) # see part 2
> round(M, 3)

# compute semantic distance (cosine similarity)
> pair.distances("dog", "cat", M, convert=FALSE)
dog/cat
 0.9610952

# find nearest neighbours
> nearest.neighbours(M, "dog", n=3)
cat  pig  cup
 16.03458 20.08826 31.77784

> plot(nearest.neighbours(M, "dog", n=5, dist.matrix=TRUE))
```

Explorations

While you wait for part 2, you can explore some DSM similarity networks online:

- https://corpora.linguistik.uni-erlangen.de/shiny/wordspace/
- built in R with wordspace and shiny

Getting practical: R as a (toy) laboratory
References I


References II


References III


References IV


Getting practical R as a (toy) laboratory

References V


Turney, Peter D. and Pantel, Patrick (2010). From frequency to meaning: Vector space models of semantics. Journal of Artificial Intelligence Research, 37, 141–188.