Purpose of the demo task
= to show several things related to gold standard data for a supervised machine learning task, especially

- Manual annotation and basic data analysis
- Gold Standard data distribution
- Inter-annotator agreement
- Confusion matrices
- Error analysis
Verb Patterns Classification is a kind of *lexical disambiguation* of verbs. The task is similar to the traditional *word sense disambiguation* (WSD). The two tasks differ in how the semantic categories are defined (word senses vs. patterns of typical verb usage).

Let’s focus on two English verbs, namely *cry* and *enlarge*. 
CRY -- dictionary definitions

cry ※ *****

1 cry; cries; crying; cried
When you cry, tears come from your eyes, usually because you are unhappy or hurt.
   I hung up the phone and started to cry.
   Please don't cry.
   He cried with anger and frustration.
   ...a crying baby.

   VB

2 cry; cries; crying; cried
If you cry something, you shout it or say it loudly.
   'Nancy Drew,' she cried, 'you're under arrest!'.
   I cried: 'It's wonderful news!'

   VB

5 cry; cries
You can refer to a public protest about something or appeal for something as a cry of some kind. (JOURNALISM)
   There have been cries of outrage about this expenditure.
   Many other countries have turned a deaf ear to their cries for help.

N-COUNT: usu N of/for n
**ENLARGE -- dictionary definitions**

*enlarge* ♞  ••••• enlarge; enlarges; enlarging; enlarged

1. **When you enlarge something or when it enlarges, it becomes bigger.**
   ...
   the plan to enlarge Ewood Park into a 30,000 all-seater stadium...
   The glands in the neck may enlarge.

V-ERG

@ **enlarged**

   The UN secretary-general yesterday recommended an enlarged peacekeeping force.

ADJ-GRADED

2. **To enlarge a photograph means to develop a bigger print of it.**
   ...
   newly-weds wishing to enlarge snaps of their big day.

VB

3. **If you enlarge on something that has been mentioned, you give more details about it.** (FORMAL)
   He didn't enlarge on the form that the interim government and assembly would take.
   I wish to enlarge upon a statement made by Gary Docking.

VB

= expand
# CRY -- Pattern definitions

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Pattern Definition</th>
<th>Explanation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 1</td>
<td>[Human] cry [no object]</td>
<td>[[Human]] weeps usually because [[Human]] is unhappy or in pain</td>
<td><em>His advice to stressful women was: ‘If you cry, do n’t cry alone.</em></td>
</tr>
<tr>
<td>Pattern 4</td>
<td>[Human] cry [THAT-CL</td>
<td>WH-CL</td>
<td>QUOTE] ({out})</td>
</tr>
<tr>
<td>Pattern 7</td>
<td>[Entity</td>
<td>State] cry [{out}] [{for} Action] [no object]</td>
<td>[[Entity</td>
</tr>
</tbody>
</table>
## ENLARGE -- Pattern definitions

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Pattern Definition</th>
<th>Explanation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 1</td>
<td>$[[\text{Human}]^[\text{Eventuality}]] \text{ enlarge } [\text{Entity}]$</td>
<td>$[[\text{Human}</td>
<td>\text{Eventuality}]] \text{ causes } [[\text{Entity}]] \text{ to grow or become larger}$</td>
</tr>
<tr>
<td>Pattern 2</td>
<td>$[\text{Entity}] \text{ enlarge } [\text{no object}]$</td>
<td>$[[\text{Entity}]] \text{ grows or becomes larger}$</td>
<td><em>As infants grow, their bodies not only enlarge but change both in shape and colour.</em></td>
</tr>
<tr>
<td>Pattern 3</td>
<td>$[[\text{Human}]^[\text{Document}]] \text{ enlarge } [{\text{on}</td>
<td>\text{upon}}] \text{ Anything } = \text{ Topic} [\text{no object}]$</td>
<td>$[[\text{Human}]] \text{ speaks or writes at length on } [[\text{Anything } = \text{ Topic}]] \text{ or } [[\text{Document}]] \text{ contains long-winded comments on } [[\text{Topic}]]$</td>
</tr>
<tr>
<td>Pattern 4</td>
<td>enlarged</td>
<td>now larger than before, without any deliberate causer or causer irrelevant</td>
<td><em>The fluid filled spaces or ventricles appear to be enlarged, and the blood flow to the front of the brain is reduced.</em></td>
</tr>
</tbody>
</table>
You will classify *cry* and *enlarge* manually.

• You will be given 10+10 sentences with the given verbs

• For each sentence you will assign a pattern that fits best the given sentence
  • there are 3 predefined patterns for the verb *cry*
  • there are 4 predefined patterns for the verb *enlarge*
  • if you think that no pattern matches the sentence, choose "u"
  • if you think that the given word is not a verb, choose "x"

• Use the forms posted at https://ufal.mff.cuni.cz/courses/npfl054/demo
Gold standard data sets are posted on the course web page (DEMO).

CRY – 250 instances in the GS set

<table>
<thead>
<tr>
<th>class</th>
<th>1</th>
<th>4</th>
<th>7</th>
<th>u</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>131</td>
<td>59</td>
<td>13</td>
<td>33</td>
<td>14</td>
</tr>
</tbody>
</table>

ENLARGE – 300 instances in the GS set

<table>
<thead>
<tr>
<th>class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>230</td>
<td>21</td>
<td>20</td>
<td>26</td>
<td>3</td>
</tr>
</tbody>
</table>
Automatic classifier is a function that assigns certain output class to each input instance.

Output class is a discrete (possibly categorical) value.

In the demo task: Pattern tags are categorical output values, sentences containing the verbs in question are input instances.

Classifier accuracy is often estimated using a test data sample as a percentage of correctly classified instances in the sample. This estimate is called sample accuracy.

Automatic predictions made by automatic classifier (our best model F1) are posted on the course web page (DEMO).

– NOTE that it is the same GS set, and it was also used as training data (!).
– Thus, you can compute only the training error, not the test error.
Manual annotation

Annotated data – a subset of the GS
– the same data set annotated by each group

2014 – 2 groups

- A (5 Czech)
- B (2 Czech, 3 foreign)

2015 – 4 groups

- A (6 Czech)
- B (6 Czech)
- C (6 Czech)
- D (6 Czech)

Now we can analyse/compare

- which group is closer to the Gold Standard
- inter-annotator agreement between groups
- error types
  - made by people
  - made by automatic classifier
A, B and GS distributions - CRY (2014)

Cry A

Cry B

Cry C

Cry D
A vs GS - confusion matrix - CRY (2014)

Number of agreements: 35 (70%)
Number of disagreements: 15 (30%)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>4</th>
<th>7</th>
<th>u</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS</td>
<td>27</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>u</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>x</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>GS</td>
</tr>
<tr>
<td>----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GS</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GS</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GS</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GS</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Agreement:**
- A vs GS: 41 (68%)
- B vs GS: 40 (67%)
- C vs GS: 45 (75%)
- D vs GS: 40 (67%)

**Disagreement:**
- A vs GS: 19 (32%)
- B vs GS: 20 (33%)
- C vs GS: 15 (25%)
- D vs GS: 20 (33%)
A, B, C, D vs GS - confusion m. - ENLARGE (2015)

Agreement: 38 (63%)
Disagreement: 22 (37%)

Agreement: 36 (60%)
Disagreement: 24 (40%)

Agreement: 28 (47%)
Disagreement: 32 (53%)

Agreement: 28 (47%)
Disagreement: 32 (53%)
Practical exercises in R

1. Download two files with annotated data cry-A.csv and cry-C.csv.
   - [https://ufal.mff.cuni.cz/courses/npfl054/demo](https://ufal.mff.cuni.cz/courses/npfl054/demo)

2. Run R and read the data using `read.csv()`.
   - Hint: see the posted Tutorial, Part I.
   - ... and create objects `cry.A` and `cry.C`.

3. Make the confusion matrix between groups A and C.
   - Hint: use `table(cry.A$class, cry.C$class)`

4. Compute simple agreement (in percentage) between A and C.
   - Hint: use `diag()` and `sum()`
Summary of Lab #1

Homework

• Go through the posted Tutorial
  ... and get familiar with all details in Parts I and II

• Get familiar with the data.table package
  – just to understand Part II

• Do exercise B) in Part III