

CLARA TreeCourse, December 15, 2010  
Prague, Czech Republic



# Play the Language: an alternative way of annotation

**Barbora Hladká, Jiří Mírovský**

in cooperation with

Jirka Hana, Jan Kohout, Vladimír Rovenský



# Human-machine communication

- applications comfortable and friendly to users
- communication in the user's native language
- knowledge of the language
  - From where? → From (annotated) corpora.

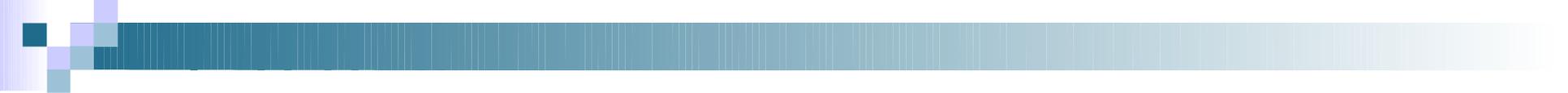
# Annotated corpora

- source of information on natural language structure (e.g. **Prague dependency treebanks**)
  - theoretical linguistics
  - natural language processing
    - supervised learning methods
    - ...
      - **the more data, the better**



# Data, more data, even more data

- 1969
  - Brown corpus of **1M** tokens
- 1994
  - British National Corpus of **100M** tokens
- 2006
  - Web 1T 5gram of **1,024,908,267,229** tokens



# An alternative way of annotation

# Human computation

- The idea: there is a lot of things
  - that **humans** have no problem with, but
  - **computers** cannot yet solve.
  - I.e. tasks trivial for humans, but very challenging for computers.
- How to address these things? Human computation.
  - Use massive groups of networked human minds to solve the problem that computers cannot.
- Hot issue over the last decade.

# In other word, crowd-sourcing

**NOISE TO SIGNAL**  
Rob Cottingham



# Human computation - examples

- Voluntary
  - Wikipedia, Open source software
- Motivation by money
  - Amazon Mechanical Turk
- Motivation by fun
  - Games with a purpose
- No other choice
  - CAPTCHA/reCAPTCHA

# Wikipedia

 [Log in / create account](#)



**WIKIPEDIA**  
The Free Encyclopedia

- [Main page](#)
- [Contents](#)
- [Featured content](#)
- [Current events](#)
- [Random article](#)
- [Donate to Wikipedia](#)

## ▼ Interaction

- [Help](#)
- [About Wikipedia](#)
- [Community portal](#)
- [Recent changes](#)
- [Contact Wikipedia](#)

## ► Toolbox

## ► [Print/export](#)

## ▼ Languages

Article [Discussion](#)

Read [Edit](#) [View history](#)

Search



## Crowdsourcing

From Wikipedia, the free encyclopedia

**Crowdsourcing** is the act of outsourcing tasks, traditionally performed by an employee or [contractor](#), to an undefined, large group of people or community (a [crowd](#)), through an open call.

Jeff Howe, one of first authors to employ the term, established that the concept of crowdsourcing depends essentially on the fact that because it is an open call to an undefined group of people, it gathers those who are most fit to perform tasks, complex problems and contribute with the most relevant and fresh ideas to benefits from their inputs.

For example, the public may be invited to develop a new technology, carry out a design task (also known as **community-based design**<sup>[1]</sup> and [distributed participatory design](#)), refine or carry out the steps of an algorithm (see [human-based computation](#)), or help capture, systematize or analyze large amounts of data (see also [citizen science](#)).

The term has become popular with businesses, authors, and journalists as shorthand for the trend of leveraging the mass collaboration enabled by [Web 2.0](#) technologies to achieve business goals. However, both the term and its underlying business models have attracted controversy and criticisms.

### **Contents** [\[hide\]](#)

- [1 History](#)
- [2 Overview](#)
  - [2.1 Web-based crowdsourcing](#)
  - [2.2 Collaboration](#)

# Amazon Mechanical Turk

All HITS | HITS Available To You | HITS Assigned To You

Search for  containing  that pay at least \$  for which you are qualified

## All HITS

1-10 of 2080 Results

Sort by:

[Show all details](#) | [Hide all details](#)

1 2 3 4 5 > [Next](#) >> [Last](#)

### Identify Arabic Dialect in Text

[View a HIT in this group](#)

**Requester:** [Chris Callison-Burch](#)      **HIT Expiration Date:** Dec 31, 2010 (2 weeks 3 days)      **Reward:** \$0.05  
**Time Allotted:** 15 minutes      **HITS Available:** 13917

### Find first name, last name and email address for Parent Teacher Association board members

[View a HIT in this group](#)

**Requester:** [James Gunter](#)      **HIT Expiration Date:** Dec 27, 2010 (1 week 6 days)      **Reward:** \$0.04  
**Time Allotted:** 60 minutes      **HITS Available:** 5489

### Judge the Quality of Search Results

[View a HIT in this group](#)

**Requester:** [Dolores Labs](#)      **HIT Expiration Date:** Dec 21, 2010 (6 days 23 hours)      **Reward:** \$0.08  
**Time Allotted:** 60 minutes      **HITS Available:** 4638

### Preference Judgements between Search Engine Results

[View a HIT in this group](#)

**Requester:** [jaime arguello](#)      **HIT Expiration Date:** Dec 23, 2010 (1 week 2 days)      **Reward:** \$0.04  
**Time Allotted:** 5 minutes      **HITS Available:** 2736

# CAPTCHA, reCAPTCHA



WIKIPEDIA  
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia

## Interaction

- Help
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia

## Toolbox

## Print/export

## Languages

[Log in / create account](#)

Article

**Discussion**

Read

**Edit**

View history

Search

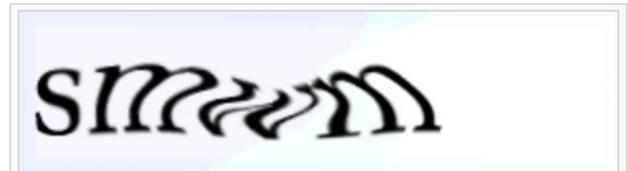


## CAPTCHA

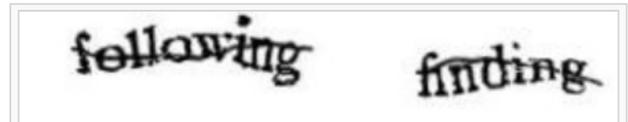
From Wikipedia, the free encyclopedia

A **CAPTCHA** or **Captcha** (pronounced /ˈkæptʃə/) is a type of **challenge-response** test used in **computing** to ensure that the response is not generated by a computer. The process usually involves one computer (a **server**) asking a user to complete a simple test which the computer is able to generate and grade. Because other computers are unable to solve the CAPTCHA, any user entering a correct solution is presumed to be human. Thus, it is sometimes described as a **reverse Turing test**, because it is administered by a machine and targeted to a human, in contrast to the standard **Turing test** that is typically administered by a human and targeted to a machine. A common type of CAPTCHA requires the user to type letters or digits from a distorted image that appears on the screen.

The term "CAPTCHA" (based upon the word **capture**) was coined in 2000 by **Luis von Ahn**, **Manuel Blum**, Nicholas J. Hopper, and **John Langford** (all of **Carnegie Mellon University**). It is a **contrived acronym** for "**C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part." Carnegie Mellon University attempted to trademark the term <sup>[2]</sup> but the trademark application was abandoned on 21 April



Early CAPTCHAs such as these, generated by the EZ-Gimpy program, were used on **Yahoo!**. However, technology was developed to read this type of CAPTCHA<sup>[1]</sup>



A modern CAPTCHA, rather than attempting to create a distorted background and high levels of warping on the text, might focus on making **segmentation** difficult by adding an angled line



# Games with a Purpose (1)

- By Lenore Blum and Luis von Ahn.
- The users enjoy playing the games.
- The players perform basic tasks that cannot be automated.
- The players work simply by playing the game.
- The data are generated as a by-product of the game.
- If the game is enjoyable, it brings human resources and saves financial resources.
- The game popularity brings more game sessions and thus more annotated data.

# Games with a Purpose (2)

- Key issues:
  - mechanisms that make the game enjoyable,
  - mechanisms that guarantee the game output is correct, i.e. quantity vs. quality
    - correctness
    - fidelity
    - usability
    - applicability

If we get the players' agreement on the output then we can expect more reliable results.

Sign In

remember me

[forgot password?](#)

# Play the Games, Change the Web.

When you play a game at Gwap,  
you aren't just having fun.

[Learn More](#)

[Register](#)

Gender  
Guesser

Prizes

Best  
Images

## ESP Game Concentrate...

What do you see  
in an image?  
Do you have ESP?



- [Today's](#)
- [All Time](#)
- Top 10**
- 6 [Prodigy](#) **139 k**
- 7 [hamilton](#) **117 k**
- 8 [Lottie](#) **112 k**
- 9 [Livenlearn](#) **93 k**
- 10 [soybean](#) **79 k**

# Remarkable gwap.com statistics

“A total of 13,630 people played the ESP Game during this time [August 9 - December, 2003], generating 1,271,451 labels for 293,760 different images. Over 80% of the people played on more than one occasion. Furthermore, 33 people played more than 1,000 games (this is over 50 hours playing the game).“

Luis von Ahn, 2004



Let's turn text annotation into games

# Annotation

vs.

# Game

<ul style="list-style-type: none"><li>▪ linguists</li></ul>	<b>goal, team</b>	<ul style="list-style-type: none"><li>▪ crowd of Internet users, <b>motivation needed</b></li></ul>
<ul style="list-style-type: none"><li>▪ theoretical linguistic framework</li></ul>	<b>input knowledge</b>	<ul style="list-style-type: none"><li>▪ no linguistic knowledge</li></ul>
<ul style="list-style-type: none"><li>▪ e.g. single, parallel</li></ul>	<b>strategy</b>	<ul style="list-style-type: none"><li>▪ [single double ...]-player game</li></ul>
<ul style="list-style-type: none"><li>▪ guidelines, annotation schemes</li></ul>	<b>instructions</b>	<ul style="list-style-type: none"><li>▪ what can (not) the players do</li></ul>
<ul style="list-style-type: none"><li>▪ editor</li></ul>	<b>environment</b>	<ul style="list-style-type: none"><li>▪ game implementation</li></ul>
<ul style="list-style-type: none"><li>▪ evaluation</li></ul>	<b>evaluation</b>	<ul style="list-style-type: none"><li>▪ evaluation</li></ul>

cheaper and faster

# Texts vs. video vs. audio

- observing images
- listening to tunes
- reading texts



*... In the year 1878 I took my degree of Doctor of Medicine of the University of London, and proceeded to Netley to go through the course prescribed for surgeons in the army. Having completed my studies there, I was duly attached to the Fifth Northumberland Fusiliers as Assistant Surgeon. The regiment was stationed in India at the time, and before I could join it, the second Afghan war had broken out. On landing at Bombay, I learned that my corps had advanced through the passes, and was already deep in the enemy's country. I followed, however, with many other officers who were in the same situation as myself, and succeeded in reaching Candahar in safety, where I found my regiment, and at once entered upon my new duties...*



# Text GWAPs so far

- **PhraseDetectives** – anaphora resolution
- **Jinx** – word sense disambiguation
- **Verbosity** – collecting common-sense facts
- **OntoGames** – semantic content annotation

# Phrase Detectives (1)



## USERPROFILE

**bvh**  
4 this week  
4 decisions  
0 agreements  
0 extras

4 this month  
8 all time

Level: Trainee

Your rating: **36%**

**CASE OPEN**  
11 tasks remaining

1 completed case

[EDIT PROFILE](#) | [LOGOUT](#)

 To se mi libí  Toto se libí 15 lidem.



## Instructions

## NAME THE CULPRIT

Has the phrase shown in **orange** been mentioned before in this text or is it a property? Use your mouse to select the **closest phrase(s)** if it has been mentioned before.



PHRASE-DETECTIVES

## Acoustic Kitty (Wikipedia)

**Acoustic Kitty** was a CIA project launched by the Directorate of Science & Technology in the 1960s attempting to use cats in spy missions.



Not mentioned before

This is a property 

Done 

-  Comment on this phrase
- ▶▶ Skip this one
- ▶▶ Skip - closest phrase can't be selected
- ▶▶ Skip - closest phrase is no longer visible
- ▶▶ Skip - error in the text

## SEARCHCLUES

Words like **they**, **him**, **her** and **it** are likely to refer to something else in the text. Try to find the closest mention of this phrase.

Words like **they** or **them** could refer to more than one thing in the text so select more than one phrase if necessary.

Always look for the **closest previous mention** of the phrase to score maximum agreement points.

Feedback 

# Phrase Detectives (2)



### USERPROFILE

**bvh**  
7 this week  
7 decisions  
0 agreements  
0 extras

7 this month  
11 all time

Level: **Trainee**

Your rating: **36%**

**CASE OPEN**  
8 tasks remaining

1 completed case

[EDIT PROFILE](#) | [LOGOUT](#)

 To se mi líbí  Toto se líbí 15 lidem.



### Instructions

## DETECTIVES CONFERENCE

Another detective has made a decision about a phrase, either that it refers to another phrase, it has not been mentioned before, it is a property or it does not refer to anything. **Do you agree with them?**



PHRASE-DETECTIVES

### Acoustic Kitty (Wikipedia)

Acoustic Kitty was a CIA project launched by the Directorate of Science & Technology in the 1960s attempting to use cats in spy missions. **A battery** and a microphone were implanted into **a cat** and an antenna into its tail.

The phrase in blue is the **closest** phrase that refers to the phrase in orange.



Disagree

Agree



### SEARCHCLUES

Words like **they**, **him**, **her** and **it** are likely to refer to something else in the text. Try to find the closest mention of this phrase.

Words like **they** or **them** could refer to more than one thing in the text so select more than one phrase if necessary.

Always look for the **closest previous mention** of the phrase to score maximum agreement points.

Feedback



# Our contribution

- Play the Language: LGame portal

- text GWAPs

- coreference, language modelling



- Elysium

- an annotation editor for school kids

- morphology, syntax

- educational aspect for children's learning activities



# LGame - basic features

1. Playing the games requires no extra knowledge of the language in question.
2. The game rules are language independent.
3. Parallel data uploaded into the sessions.



Register



Registered: 296  
Logged-in: tester2

-  **Shannon game** - a game with words in the sentence ?
-  **Place the Space** - a game with words in the sentence ?
-  **PlayCoref** - a game with words in texts ?

Play!

Ranking

Play!

Ranking

Play!

Ranking

**! A competition before Christmas 2010 !**

Logout

Profile

Contact

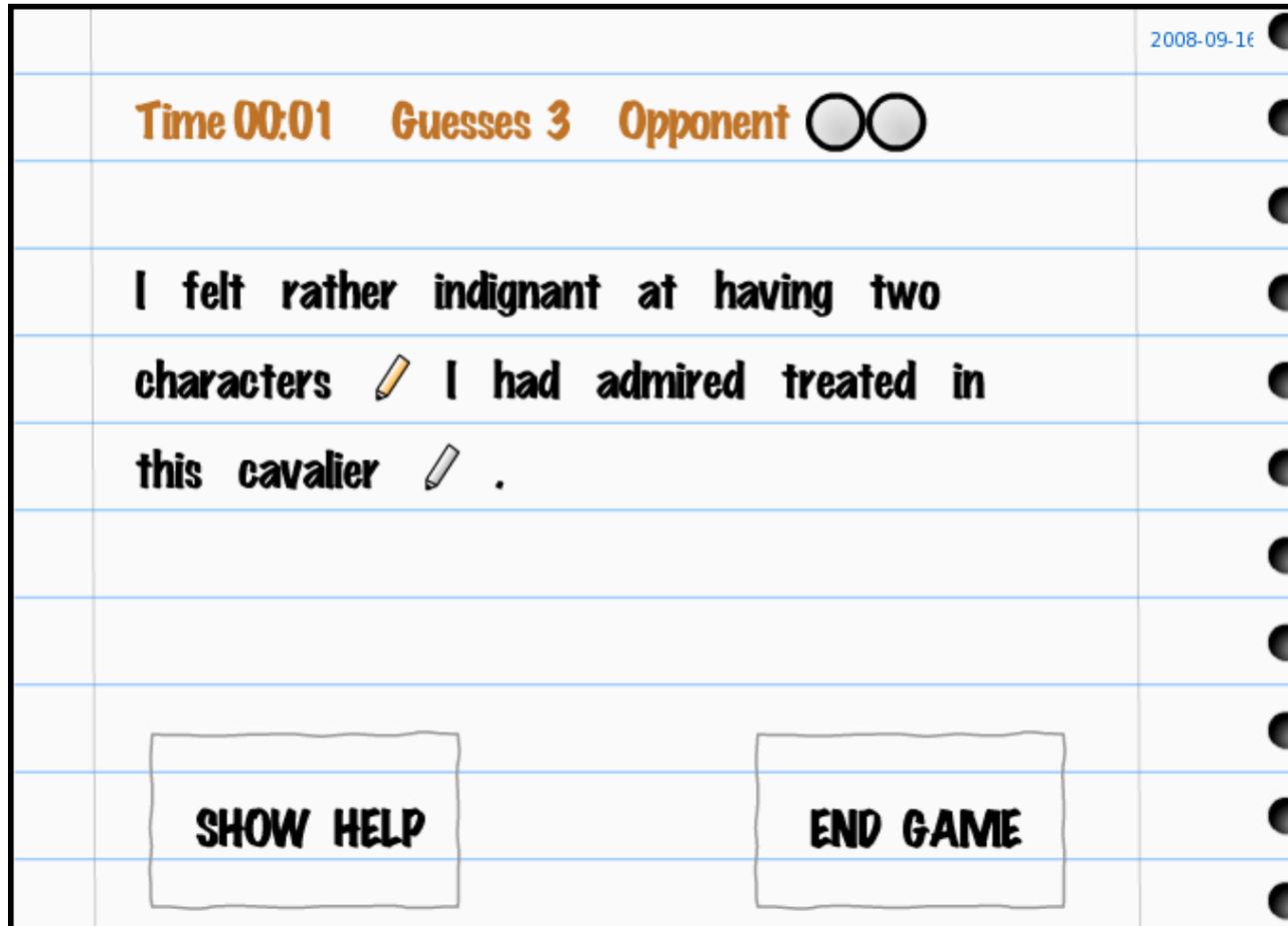
Academic column



# Shannon Game

- A game of intentionally hidden words in the sentence; players guess them.
- For two players only; could be for one, too.

# Shannon Game



The screenshot shows a game interface with a white background and blue horizontal lines. At the top right, there is a date '2008-09-16'. Below it, the game status is displayed: 'Time 00:01', 'Guesses 3', and 'Opponent' followed by two empty circles. The main text area contains the sentence: 'I felt rather indignant at having two characters I had admired treated in this cavalier manner.' The text is written in a black, monospace-style font. At the bottom, there are two buttons: 'SHOW HELP' on the left and 'END GAME' on the right, both in a simple, hand-drawn style.

2008-09-16

Time 00:01   Guesses 3   Opponent ○○

I felt rather indignant at having two  
characters I had admired treated in  
this cavalier manner .

SHOW HELP      END GAME

# Shannon Game

2008-09-16

Time 00:50    Guesses 3    Opponent ○○

I felt rather indignant at having two  
characters **whom** I had admired treated  
in this cavalier  .

SHOW HELP                      END GAME

# Shannon Game

2008-09-16

Time 01:28   Guesses 3   Opponent

I felt rather indignant at having two  
characters **whom** I had admired treated  
in this cavalier  .

SHOW HELP      END GAME

# Shannon Game

2008-09-16

**Results**

Player		-20 pts
Opponent		30 pts

---

**You lost this game.**

**PLAY AGAIN**

# Shannon Game – Scoring Func.

For each guessed word:

- 40 pts – if 1<sup>st</sup> guess correct
- 20 pts – if 2<sup>nd</sup> guess correct
- 10 pts – if 3<sup>rd</sup> guess correct
- -10 pts – if not guessed correctly

For 2 words: -20 ... 80 pts

# Shannon Game – Data Prep.

Arthur Conan Doyle: A Study in Scarlet

- free to use
- many translations
- free audio book

Only sentences without punctuation and of certain length selected.

Sentence segmentation, tokenization, (proper name recognition).



# Place the Space

- A game of intentionally hidden spaces in the sentence; players restore them.
- For a single player.



# Place the Space

Finish Time: 6

I have a mule and two horses  
sitting in the Eagle  
Ravine.

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

# Place the Space

## Results

Correct solution:

I have a mule and two horses waiting in the Eagle Ravine.

Player's solution:

I have amule and two horses waiting in the Eagle Ravine.

Score: 95

Play again

# Place the Space – Scoring Func.

$$\text{Score}(\text{Player}) = 100 * F_1(\text{Player}, \text{Correct})$$

Positions in text are given by the sequence of characters without spaces.

$$F_1 = 2 * P * R / (P + R)$$

*Precision* = # correctly guessed / # guesses

*Recall* = # correctly guessed / # correct



# Place the Space – Data Prep.

Arthur Conan Doyle: A Study in Scarlet

- only sentences without punctuation and of certain length selected
- sentence segmentation, tokenization



# PlayCoref

- A game of marking coreferential pairs in the document.
- For a single player or for two players.

# PlayCoref



# PlayCoref

On the very day that I had come to this conclusion, I was standing at the Criterion Bar, when some one tapped me on the shoulder, and turning round I recognized young Stamford, who had been a dresser under me at Barts. The sight of a friendly face in the great wilderness of London is a pleasant thing indeed to a lonely man

Player	Opponent
Sentences: 2	Sentences: 0
Pairs created: 0	Pairs created: 0

4:57

Next  
Finish

Adding pairs  
Deleting pairs

---

# PlayCoref

The screenshot displays the PlayCoref interface. At the top, a text passage is shown with coreference annotations. The text is: "On the very day that I had come to this conclusion, I was standing at the Criterion Bar, when some one tapped me on the shoulder, and turning round I recognized young Stamford, who had been a dresser under me at Barts The sight of a friendly face in the great wilderness of London is a pleasant thing indeed to a lonely man". The word "The" is highlighted in blue. Orange lines connect the pronouns "I" and "me" to their respective referents. A yellow line connects "Stamford" to "who".

Below the text, the interface includes a control panel with the following elements:

- Player**: Sentences: 2, Pairs created: 6
- Opponent**: Sentences: -, Pairs created: 0
- Next** button
- Finish** button
- Adding pairs** button (with a plus sign icon)
- Deleting pairs** button (with a minus sign icon)
- Timer**: 3:44

A progress bar is visible at the bottom of the interface.

# PlayCoref

On the very day that I had come to this conclusion, I was standing at the Criterion Bar, when some one tapped me on the shoulder, and turning round I recognized young Stamford, who had been a dresser under me at Barts. The sight of a friendly face in the great wilderness of London is a pleasant thing indeed to a lonely man in old days. Stamford had never been a particular crony of mine, but now I hailed him with enthusiasm, and he in his turn appeared to be delighted to see me.

Player Sentences: 3 Pairs created: 13

Opponent Sentences: - Pairs created: 0

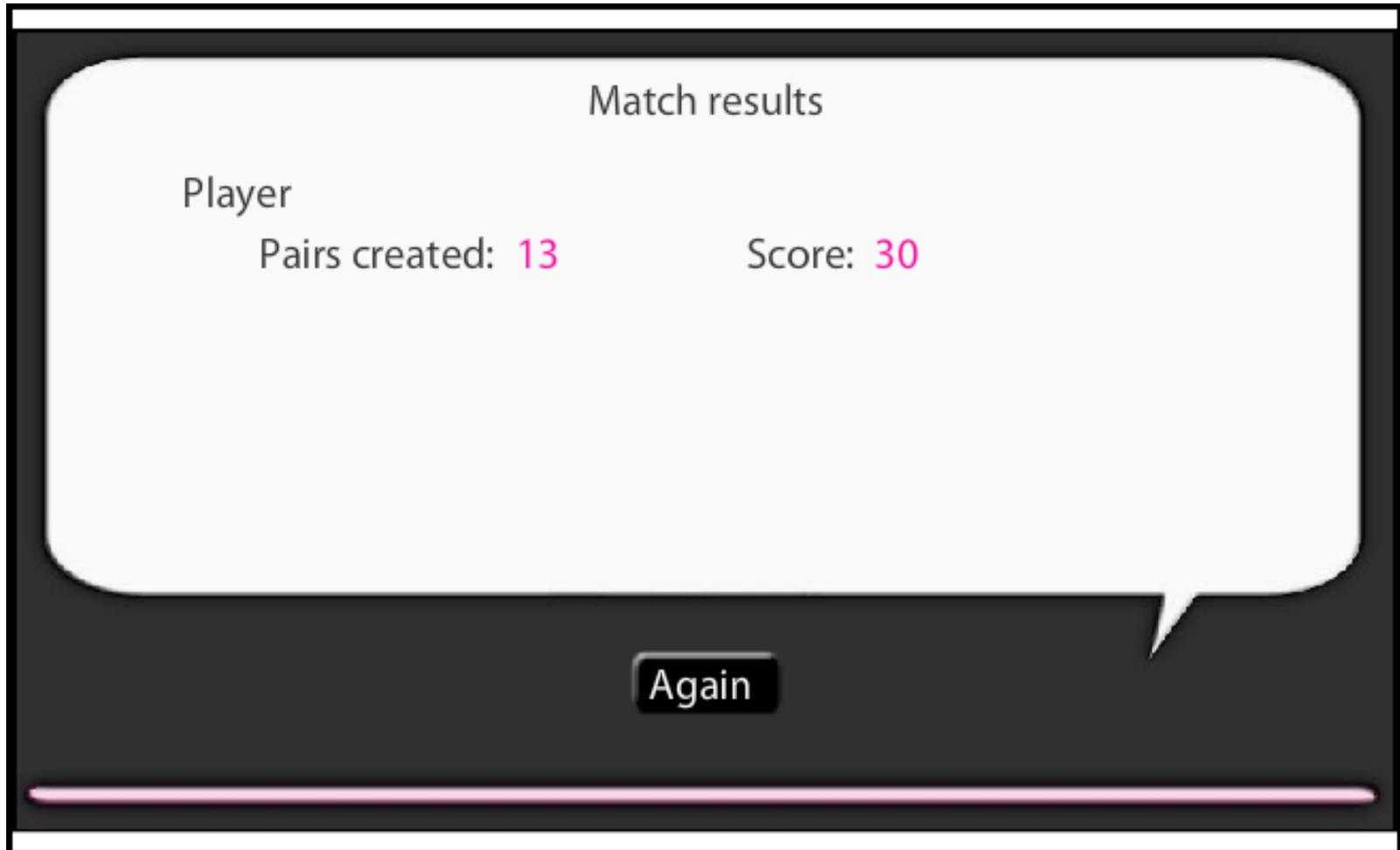
2 : 36

Next Finish

Adding pairs Deleting pairs

The image shows a screenshot of the 'PlayCoref' interface. At the top, there is a decorative blue and white pattern. Below it, the title 'PlayCoref' is displayed in a large, blue, sans-serif font. The main part of the interface is a dark grey area containing a white speech bubble with a drop shadow. Inside the speech bubble, a paragraph of text is displayed in a monospaced font. The text is: 'On the very day that I had come to this conclusion, I was standing at the Criterion Bar, when some one tapped me on the shoulder, and turning round I recognized young Stamford, who had been a dresser under me at Barts. The sight of a friendly face in the great wilderness of London is a pleasant thing indeed to a lonely man in old days. Stamford had never been a particular crony of mine, but now I hailed him with enthusiasm, and he in his turn appeared to be delighted to see me.' Several words are highlighted in pink: 'I', 'me', 'Stamford', 'who', 'man', 'mine', 'him', 'he', and 'see'. Blue lines connect these highlighted words, representing coreference links. For example, one line connects the 'I' at the start of the sentence to the 'I' in 'I was standing'. Another line connects the 'me' in 'tapped me' to the 'me' in 'under me'. A yellow line connects 'Stamford' to 'who'. A blue line connects 'man' to 'Stamford'. Below the speech bubble, there is a control panel with a dark grey background. On the left, it shows 'Player Sentences: 3' and 'Pairs created: 13'. In the center, there is a digital timer showing '2 : 36'. On the right, it shows 'Opponent Sentences: -' and 'Pairs created: 0'. There are two buttons: 'Next' and 'Finish'. To the right of the buttons, there is a vertical slider control with a white bar, and the text 'Adding pairs' and 'Deleting pairs' next to it. At the bottom of the interface, there is a horizontal pink progress bar.

# PlayCoref



# PlayCoref – Scoring Function

*Score(Player A) =*

*$\lambda_1 * F_1(\text{Player A}, \text{Player B})$*

*+  $\lambda_2 * F_1(\text{Player A}, \text{ACR or Manual})$*

*+  $\lambda_3 * \min(12, \text{sentences})/12$*

*(and multiplied by 100;  $\Sigma(\lambda_i) = 1, \lambda_i > 0$ )*

# PlayCoref – Scoring Function

OUR morning 's exertions had been too much for my weak health, and I was tired out in the afternoon. After Holmes' departure for the concert, I lay down upon the sofa and endeavoured to get a couple of hours' sleep. It was a useless attempt. My mind had been too much excited by all that had occurred, and the strangest fancies and surmises crowded into it.

Player Sentences: 4 Pairs created: 3

Opponent Sentences: - Pairs created: 0

4:25

Next Finish

Adding pairs Deleting pairs

The screenshot displays a text passage with coreference annotations. The text is: "OUR morning 's exertions had been too much for my weak health, and I was tired out in the afternoon. After Holmes' departure for the concert, I lay down upon the sofa and endeavoured to get a couple of hours' sleep. It was a useless attempt. My mind had been too much excited by all that had occurred, and the strangest fancies and surmises crowded into it." The words "my", "I", "It", and "My" are highlighted in pink. Green lines connect "my" to "I" and "I" to "It". A blue box highlights "My". The interface includes a control panel with "Player" and "Opponent" statistics, a timer showing "4:25", and buttons for "Next" and "Finish". A vertical slider on the right is labeled "Adding pairs" (pink) and "Deleting pairs" (grey).

# PlayCoref – Scoring Function

•OUR morning •'s •exertions •had •been too much for •my weak •health,  
•and •I •was •tired out in the •afternoon. •After •Holmes' •departure for  
the •concert, •I •lay down upon the •sofa •and •endeavoured to •get a  
•couple of •hours' •sleep. •It •was a useless •attempt. •My •mind •had  
•been too much •excited by all •that •had •occurred, •and the strangest  
•fancies •and •surmises •crowded into •it.

Player Sentences: 4 Pairs created: 3

Opponent Sentences: - Pairs created: 0

3:53

Next Finish

Adding pairs  
Deleting pairs

The screenshot displays a text passage with coreference annotations. The text is: "OUR morning 's exertions had been too much for my weak health, and I was tired out in the afternoon. After Holmes' departure for the concert, I lay down upon the sofa and endeavoured to get a couple of hours' sleep. It was a useless attempt. My mind had been too much excited by all that had occurred, and the strangest fancies and surmises crowded into it." The annotations include: "OUR" in blue, "my" in pink, "I" in blue, "After" in blue, "My" in pink, and "It" in blue. A blue line connects "my" to "My". The interface also shows a control panel with "Player" and "Opponent" statistics, a timer at 3:53, and "Next" and "Finish" buttons. A vertical slider is labeled "Adding pairs" (pink) and "Deleting pairs" (grey).



# PlayCoref – Data Preparation

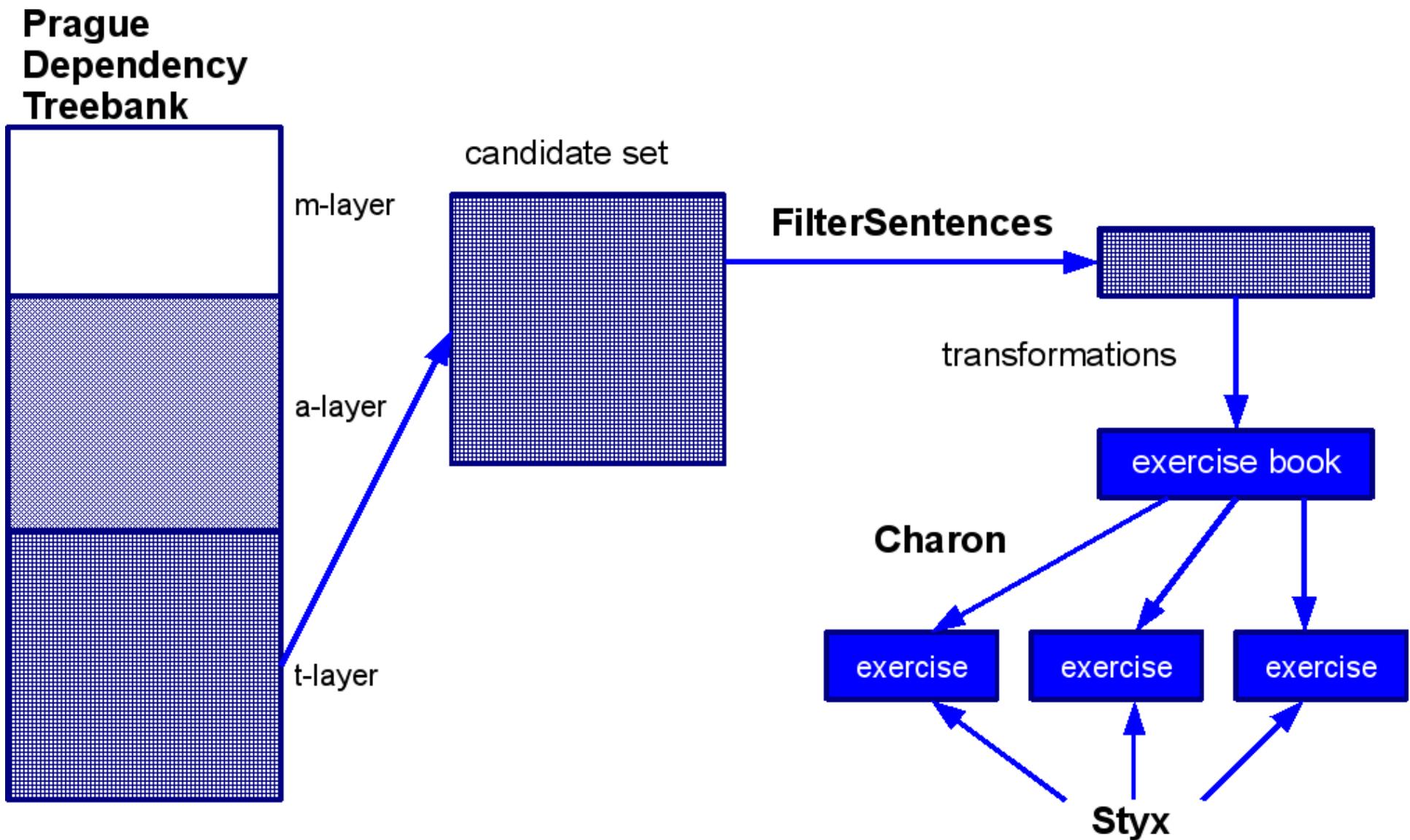
## Arthur Conan Doyle: A Study in Scarlet

- analysis up to the tectogrammatical layer using TectoMT modules; it includes simple named entity recognition
- manual annotation of coreference (automatic procedures – too low accuracy)

# Elysium

- to involve school kids in annotation
- offer them a language independent annotation editor
  - grammar practicing
  - not as enthusiastically as playing games
  - they do not have other choice

# Elysium as a part of the STYX<sup>⚡</sup> system (1)



Charon

File Set **Filter** Help

- Only sentences containing a subject
- Sentences not containing a nominal predicate
- Sentences not containing a verbal predicate
- Sentences not containing attributes
- Sentences not containing adverbials
- Sentences not containing objects
- Sentences not containing complements

V loňském  
Katalogy je  
Separátní p  
Hodně čes  
Tato fólie v

Hodnota odvedeného díla dosáhla téměř 100000 korun .

Náš student se za poslední čtyři roky změnil .

Z jeho vzdělání zmizely zbytečné předměty .

Poněkud více je jich na inženýrském stupni .

Plat profesora opravdu není velký .

U docentů a asistentů je to ještě horší .

```

graph TD
    Root(( )) --- Pred[dosáhla Pred]
    Root --- Atr1[díla Atr]
    Root --- Obj[téměř 100000 korun Obj]
    Atr1 --- Atr2[odvedeného Atr]
    Obj --- Adv[téměř Adv]
    Obj --- Obj2[100000 Atr]
  
```

Styx

File Task

Showing 963 sentences (out of 1223 sentences)

V loňském prosinci přijel k ostravské firmě Daklar .  
Rozdíl do regulované ceny byl hrazen z dotací .

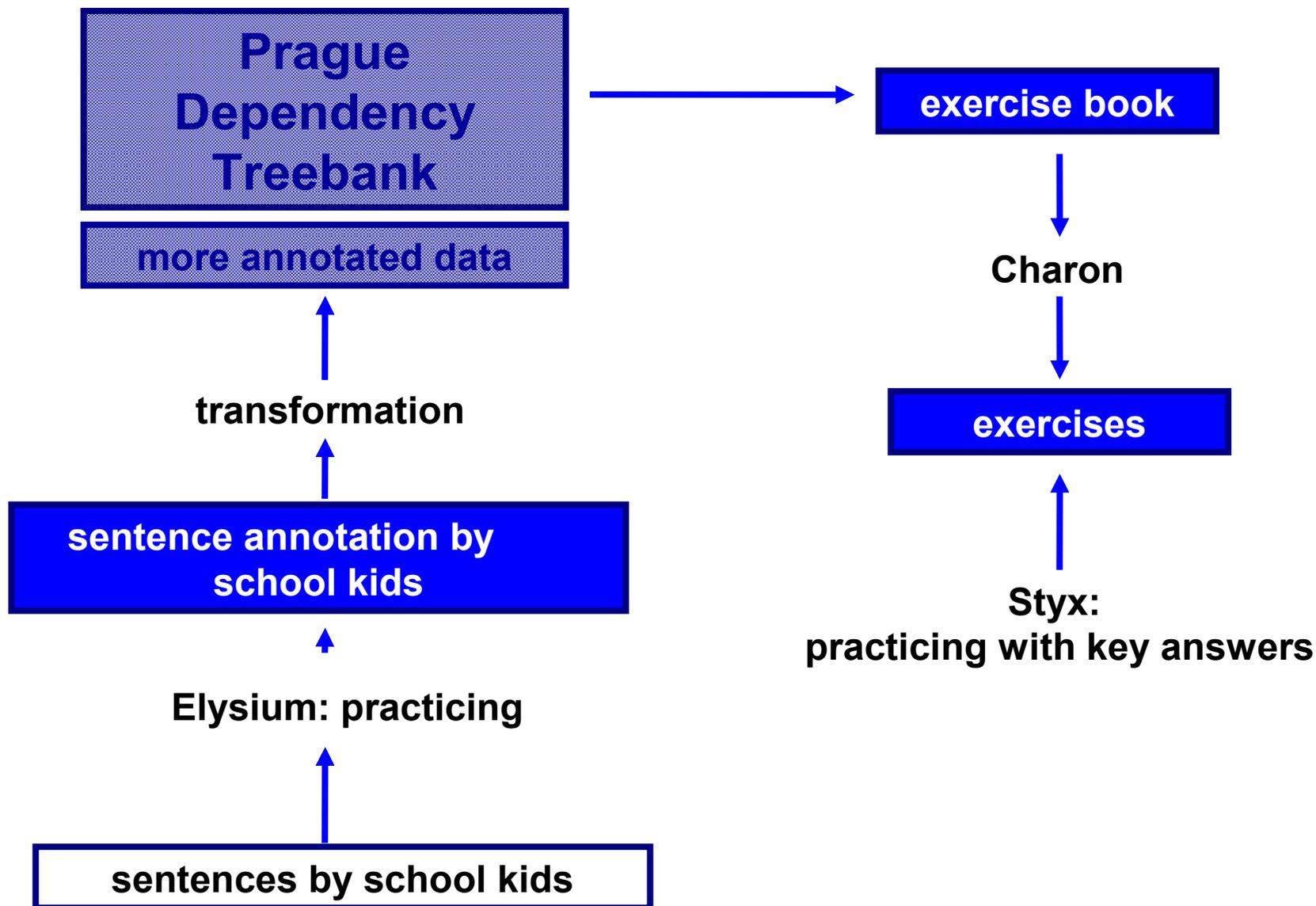
word form	lemma	part of speech	gender	number	case
Rozdíl	rozdíl	noun	masculine	singular	nominative

Rozdíl do regulované ceny byl hrazen z dotací .

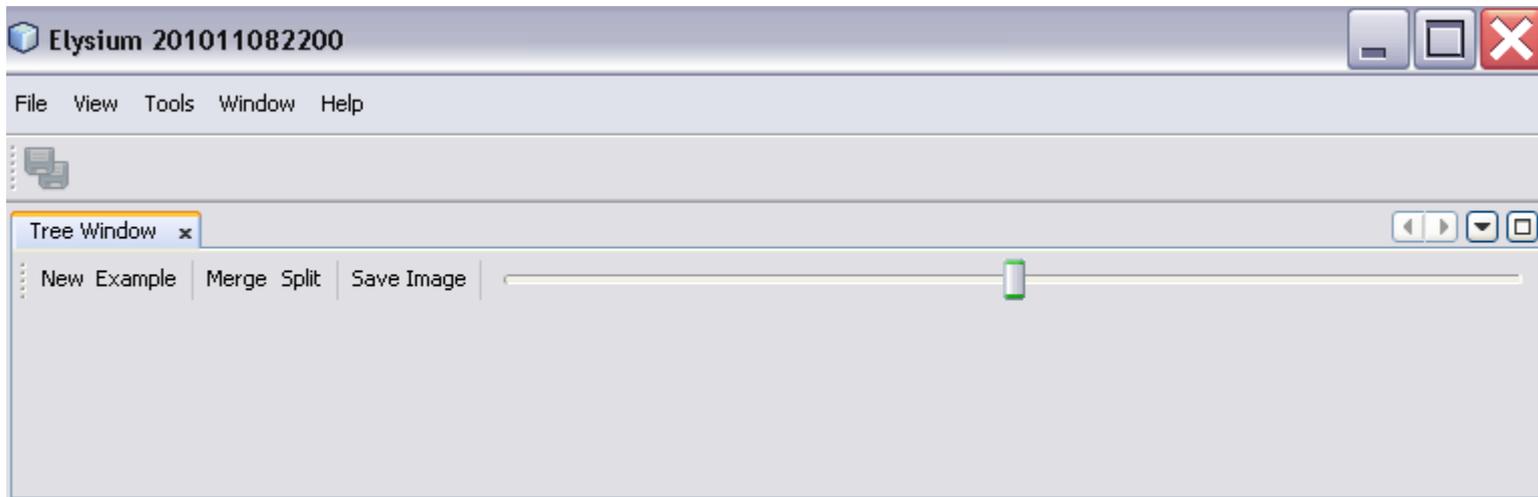
```

graph TD
    Root(( )) --- Sb[Rozdíl Sb]
    Root --- Pred[byl hrazen Pred]
    Root --- Adv[z dotací Adv]
    Sb --- Atr[do ceny Atr]
    Atr --- Atr2[regulované Atr]
  
```

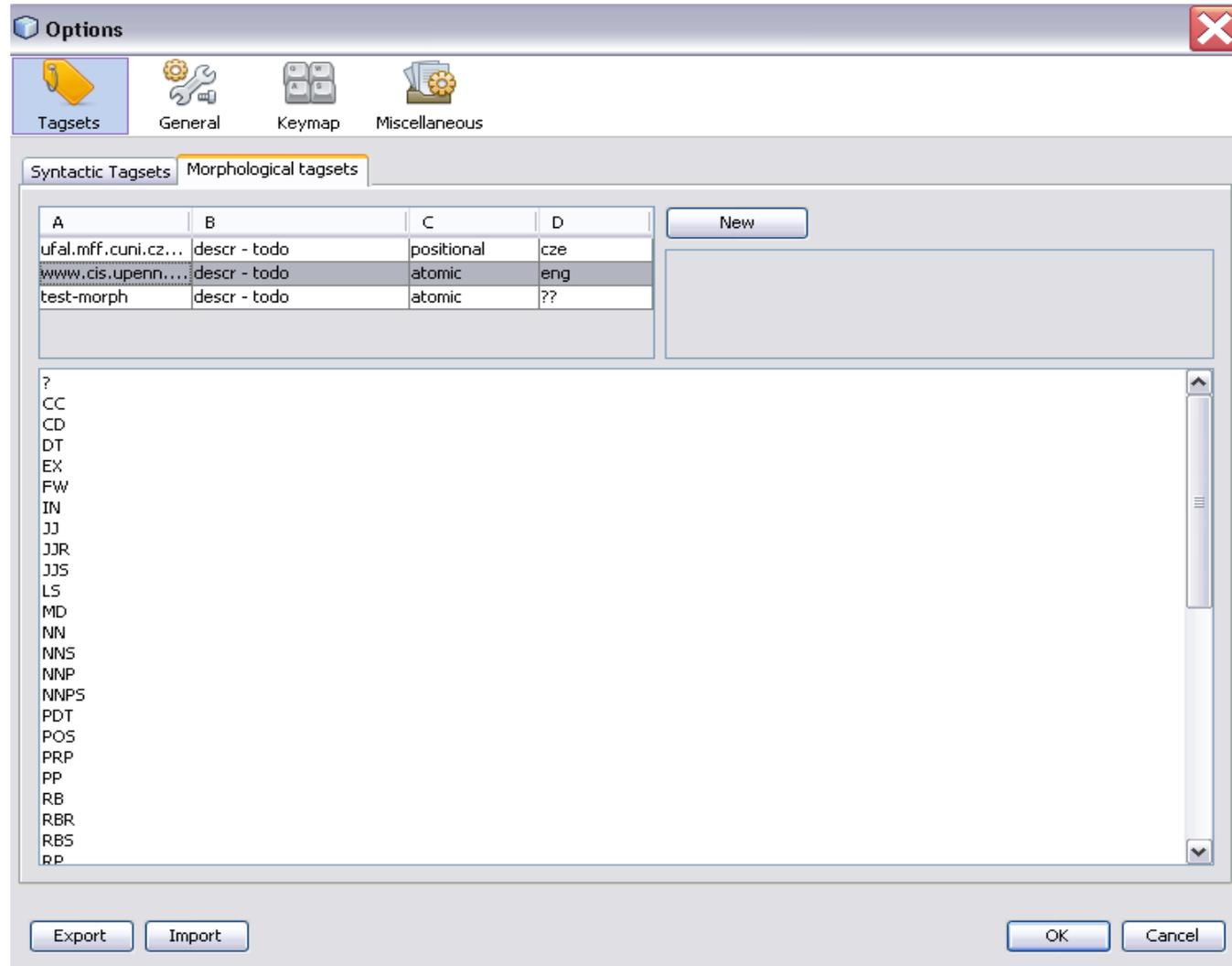
# Elysium as a part of the STYX<sup>⚡</sup> system (2)



# Elysium (1)



# Elysium - Tools->Options->Tagsets



# Elysium - New (3)

Kids are looking forward to a visit to Prague in December.

<b>Děti</b>
? <input type="checkbox"/>
?-----

<b>se</b>
? <input type="checkbox"/>
?-----

<b>těší</b>
? <input type="checkbox"/>
?-----

<b>na</b>
? <input type="checkbox"/>
?-----

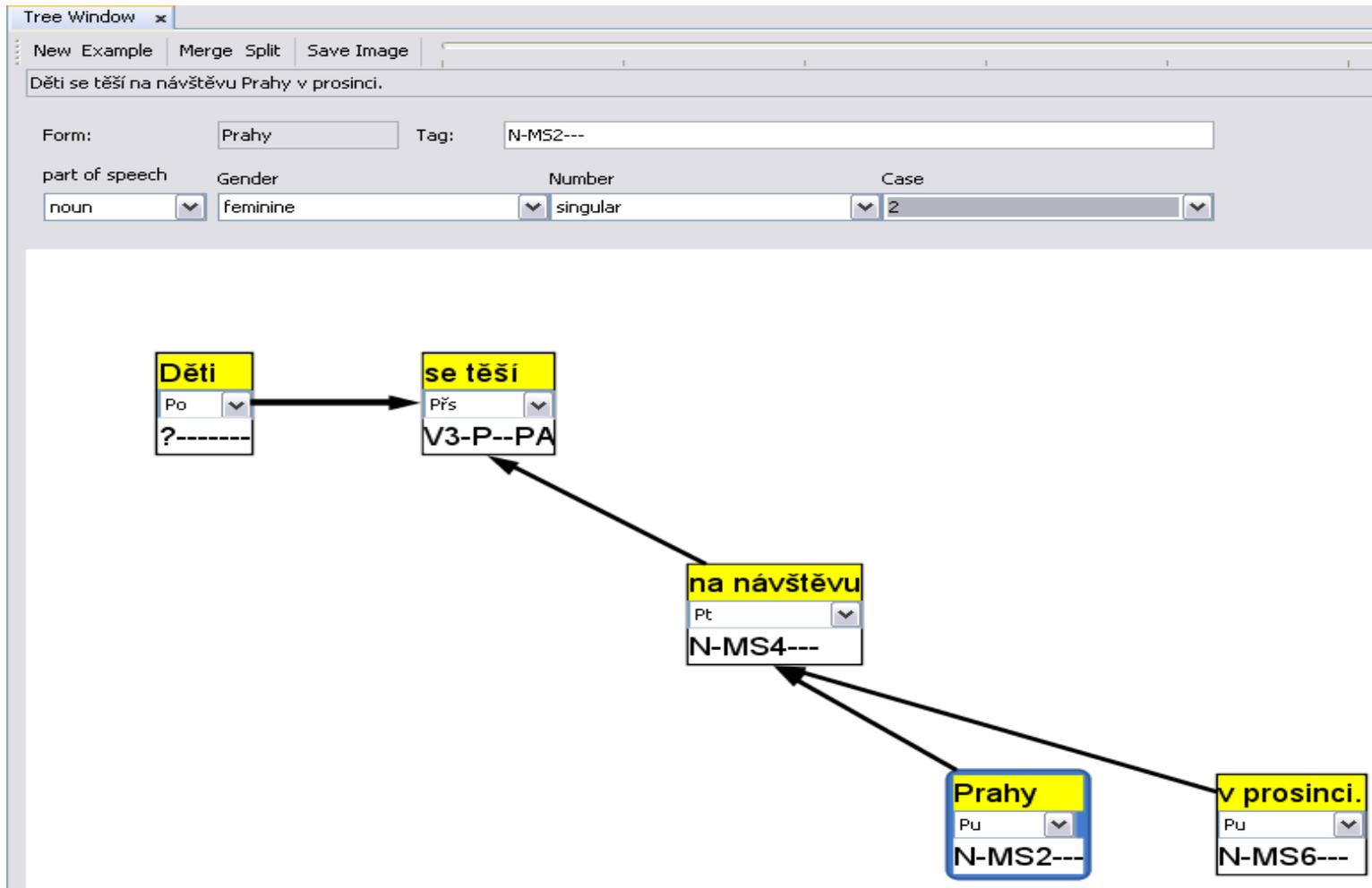
<b>návštěvu</b>
? <input type="checkbox"/>
?-----

<b>Prahy</b>
? <input type="checkbox"/>
?-----

<b>v</b>
? <input type="checkbox"/>
?-----

<b>prosinci.</b>
? <input type="checkbox"/>
?-----

# Elysium - analyses (4)





# A competition before Christmas 2010

**December 6-16, 2010**



“Although interesting as a topic, I would question the originality of the proposition and the possibility of the team to complete the research within the project.”

An anonymous reviewer, summer 2008

# Text GWAPs

- open-ended stories
- hard to predict success or failure in advance
- simply, they have to be tried ...
- Are we able to attract the Internet users to play our games? ... We are looking forward to your feedback 😊

# Links & References

- British National Corpus <http://www.natcorp.ox.ac.uk/>
- Brown Corpus [http://en.wikipedia.org/wiki/Brown\\_Corpus](http://en.wikipedia.org/wiki/Brown_Corpus)
- Elysium [http://ufal.mff.cuni.cz/styx/news\\_en.html](http://ufal.mff.cuni.cz/styx/news_en.html)
- GWAP <http://gwap.com>
- 100K ESP game sample  
<http://server251.theory.cs.cmu.edu/ESPGame100k.tar.gz>
- Jinx Seemakurty N., Chu J., von Ahn, L., Tomasic, A.: Word Sense Disambiguation via Human Computation, *KDD-HCOMP'10*, Washington, CD, USA, 2010.
- LGame <http://www.lgame.cz>
- OntoGames <http://ontogame.sti2.at/>
- PageHunt Page Ma H., Chandrasekar, R., Quirk, Ch., Gupta, A.: Hunt: Improving Search Engines Using Human Computation Games. *SIGIR'09*, Boston, MA, USA, 2009.
- Phrase Detectives <http://anawiki.essex.ac.uk/phrasedetectives/>
- STYX system <http://ufal.mff.cuni.cz/styx>
- Verbosity game <http://gwap.com>
- Web 1T 5-gram  
<http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13>