How to Present NLP Topics to Children?

Zuzana Nevěřilová, Adam Rambousek

Natural Language Processing Centre
Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic
xpopelk@fi.muni.cz, rambousek@fi.muni.cz

Abstract. This paper deals with important, but underestimated aspect of research – presenting to the general public. Amongst many outreach activities, Natural Language Processing Centre takes part in Masaryk University project of courses for children 9 to 14 years old. We describe specifics to think of when presenting NLP topics to children, and use cases of previous and planned courses.

Key words: research publicity, NLP research presentation

1 Introduction

Presenting science to the public is sometimes underestimated since it is not considered to be scientific. Nevertheless, good presentation to the public can attract new students, industry partners, or influence fundings.

In this paper, we present one particular case – presentation for children attending the Masaryk JUniversity[1]. Masaryk JUniversity offers courses for children between 9 and 14, its main aim is to attract potential students to the university. Natural Language Processing Centre (NLPC) cooperates in the project from its beginning. This paper presents the topic we presented during two seasons and a new topic we plan to present in the forthcoming season.

The result of the work is a short set of recommendations for future presentations of NLP to the public.

2 Presenting NLP research to general public

NLPC participates in various outreach activities to present results of NLP research and publicly available tools. These events may have different audience groups, e.g. high school students interested in computer science studies, or general public during European Researchers' Night. However, they usually do not have deep linguistic knowledge and do not realize the scope and meaning of NLP research.

Difficulty with linguistic and text processing research presentation is often the lack of physical object to demonstrate. The first important issue is to

[1] http://www.mjuni.cz

Aleš Horák, Pavel Rychlý, Adam Rambousek (Eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2016, pp. 153–159, 2016.

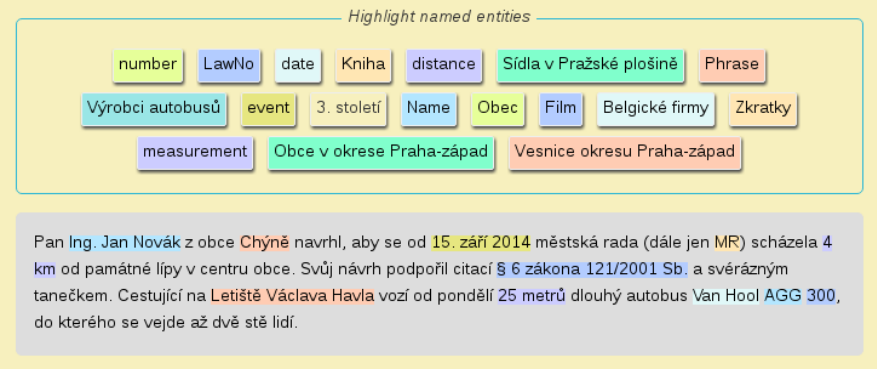

Fig. 1: Czech Named Entity Recognition with visually highlighted named entities.

Fig. 2: Word games during European Researchers' Night.

describe the tasks of NLP research and explain them using tools that people use in everyday life [1,2], e.g. spelling checkers, predictive typing, voice recognition, or machine translation. When presenting research results or NLP tools to public, try to use visually appealing graphical representation as much as possible. Not just to attract attention of the audience, but mainly because the information presented with the help of pictures enhance the chance of remembering [3,4]. For example, see visualization of named entity recognition in Figure 1.

During some events, visitors have opportunity to try our tools hands-on. For such occasion, we have created several computer games with the linguistic themes, e.g. finding the word by its related words, or computer asking

questions to guess the player's word. Although not directly related to research in NLPC, we have created two language-related board games for visitors who like to play with words – semantic network with various word relations (see Figure 2) and modified crossword [2].

3 Presenting science to primary school children

Masaryk JUniversity offers courses for children aged between 9 and 14 years old and the capacity is usually limited to 170 participants. Thanks to the requirements, children who are interested in science and more keen to study in general are enrolled in the courses. On the other hand, age of 9 to 14 is still quite wide range, and skills and development of children may vary a lot [5,6].

Generally, there are several guidelines to use when preparing courses for children in preadolescent or early adolescent age [7,8]:

- Children are curious and not shy to ask, be prepared for a lot of questions. Sometimes even off-topic.
- Children are competitive and most of them like to move. Break the lecture after approximately 30 or 40 minutes and plan some team game when the children can get up.
- It is the age of changing emotionality, so try to avoid sensitive subjects.
- Do not forget that children do not share the same long-term experience. E.g. when we were presenting predictive writing tools, some children did not know keypad mobile phones.

When presenting NLP topics and tools to children, we have discovered several specific issues:

- Children happily destroy your software tools by simply trying what is possible, e.g. really large input values or unexpected options. As a general rule in software development, never trust user input and expect the unexpected.
- Some children are bilingual and do not realize it. For example, when asked for sample words, some children automatically responded in other languages than Czech.
- Speech recognition software have issues recognizing children's voices, because the recognition model is trained on adults.

4 Predictive writing presentation

In 2015 and 2016, we presented a well defined topic of NLP: predictive writing. The topic was selected since most users have personal experience with predictive writing, however, we assumed not many users thinking about the language technology behind the application.

The presentation outline was as follows:

[2] Original word ciphers from puzzle hunts Dnem and Krtčí norou, adapted to board games by Vít Suchomel.

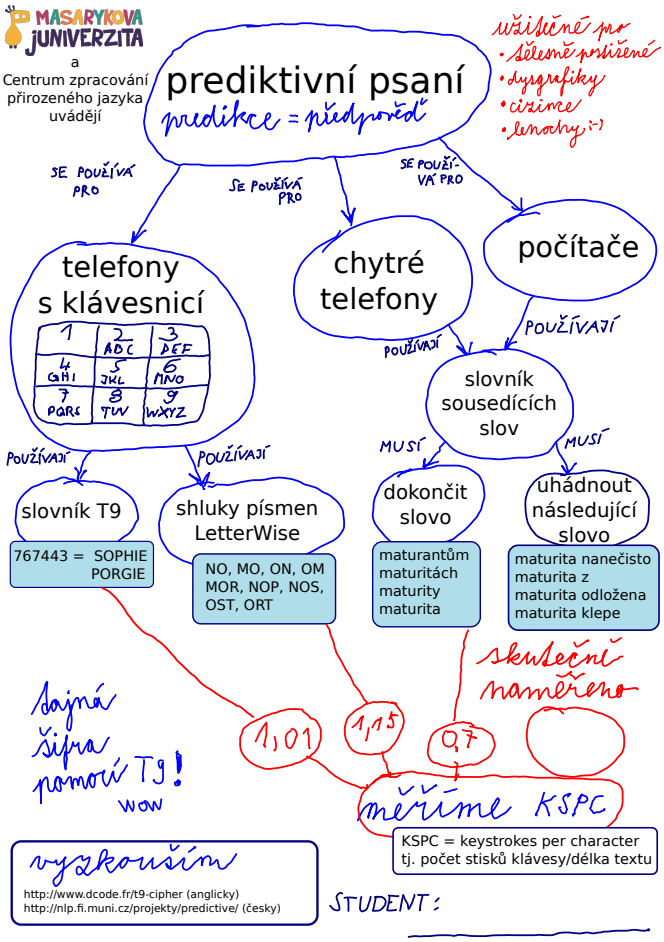

Fig. 3: Handout for predictive writing presentation.

– Predictive writing is something the audience is familiar with. We taught the term, we explained what can be predicted (word ending, next word, typo).
– Several systems exist depending on the hardware (restricted keyboards, touchscreens).
– The key software feature is a *dictionary* of words and/or n-grams and their *frequencies*.
– Physical activity: try to guess a word from a T9 typing.
– How to evaluate the quality of predictive writing software: keystrokes per character (KSPC, [9]), keystrokes on correction etc.
– Physical activity: In teams, try to compose as many words as possible from magnetic letters.
– Assistive technology: how can predictive writing help?
– Bonus material: switching among different user models in order to reduce KSPC.

We provide the audience a printed material (a single A4 sheet with a simple mindmap and links to applications, see Figure 3). We also let the children work with the Czech predictive writing demo[3].

4.1 Risks, pitfalls, and evaluation

We were unsure about the presentation intelligibility. The presentation was tested on one author's child, however, we would appreciate support from the university.

The web demo load was not tested before the presentation. The input length was not restricted. The web demo was running slowly because of the excessive web server load.

Some children were arguing that no predictive writing is needed if they used automatic speech recognition (ASR). However, no web demo contains children's voice models.

During presentation, we brought old mobile phones, since not all children have ever met mobile phones with keyboard.

The presentation was evaluated positively by the audience [10]. The topic was considered very easy or easy in two cases, just right in two cases. The presentation was considered intelligible, especially, the only comment to the presentation was "Surprisingly, I understood everything".

5 Semantic Network Presentation

For the year 2017, we have chosen a new topic. The main reason is that the number of children familiar with T9 prediction drops rapidly. The main risk we see is that unlike predictive writing, users do not have everyday experience with semantic networks.

The presentation outline is planned as follows:

- How to explain meanings of words to humans and to computers? *Connection* with other meanings is a key feature of all explanations (except of deixis).
- We need *memory* to store information about meanings, so do computers (analogy). In order to understand, we also need to know how to use information stored in memory: we present simple *entailments*.
- Semantic networks connect words that have something common in their meanings: same meaning, subclass, opposite meaning, has part etc. The connection itself explains the word meaning. Semantic network is both a way of storing information in memory, and a recipe to entail new knowledge.
- Physical activity: try to guess a word in our semantic network.
- Practical use of semantic networks: e.g. search engines use synonyms and hypernyms for query expansion.

[3] `https://nlp.fi.muni.cz/projekty/predictive`

– Bonus material: when entailments do not work. Explain monotonicity and the default rule [11].

We provide a web application demo that uses Czech WordNet [12] as a semantic network. We can show what can be entailed, and check whether it is true (manual annotation). We also present how an expanded query may improve the search results (e.g. on Google Images).

6 Conclusions

In previous sections, we presented shortly two presentations, the former already presented to children aged between 9 and 14, the latter planned for presentation in January 2017. Our experience lead to a short list of recommendations that is intended to help other scholars to present NLP research and topics to the general public. Presentation to children have few specifics compared to presentation to adults: children have less (long term) experience, they appreciate physical activity, and they are often like to compete.

7 Recommendations on presenting NLP to the public

1. Choose an appropriate subtopic of NLP, do not assume that the audience knows the field. You can present arbitrary topics, even a "difficult" ones if you are able to explain the essence of it in one simple sentence. For example: Predictive writing in cell phones is based on frequencies of words and word sequences.
2. Simplify the terminology and stay consistent in it. Do not use linguistic terminology (use sentence composition instead of syntax, word sequence instead of n-grams etc.). Use metaphors and analogies.
3. Explain words that mean something else in your field than in the general domain (e.g. *dictionary* in the general domain contains lemmata, dictionary in NLP means list of word forms)
4. Test the presentation before you give it. Let the test person(s) interrupt your test presentation with questions, comments, and associations they make.
5. Design an attractive an easy-to-use UI to your tools:
   – Language of the UI consistent to the language you used
   – Provide the UI for later use if possible
   – Log all activity, let the audience give a feedback if possible
6. Offer physical activity, specifically when presenting to children. Use physical objects, not only software.
7. Provide printed material.
8. Offer bonus material.
9. Let one main message (explanation) to emerge from your presentation
10. Expect off-topic questions and prepare answers to them. In language technology presentations, people often ask about orthography, language learning. In predictive writing presentations, people ask about speech technologies. In semantic networks presentations, people ask about artificial intelligence, machine translation, philosophy.

Acknowledgments. This work has been partly supported by the Ministry of Education of CR within the LINDAT-Clarin project LM2015071.

References

1. Purnell, T., Raimy, E., Salmons, J.: Making linguistics matter: building on the public's interest in language. Language and Linguistics Compass **7**(7) (2013) 398–407
2. McKee, C., Zimmer, E., Fountain, A., Huang, H.Y., Vento, M.: Public outreach in linguistics: Engaging broader audiences. Language and Linguistics Compass **9**(9) (2015) 349–357
3. Kosara, R., Mackinlay, J.: Storytelling: The next step for visualization. Computer (5) (2013) 44–50
4. Carney, R.N., Levin, J.R.: Pictorial illustrations still improve students' learning from text. Educational psychology review **14**(1) (2002) 5–26
5. Zelazo, P.D.: The Oxford Handbook of Developmental Psychology, Vol. 1: Body and Mind. Volume 1. Oxford University Press (2013)
6. Macek, P., et al.: Adolescence. Portál (2003)
7. Charlesworth, R.: Math and science for young children. Cengage Learning (2015)
8. Hassard, J., Dias, M.: The art of teaching science: Inquiry and innovation in middle school and high school. Routledge (2013)
9. MacKenzie, I.: KSPC (keystrokes per character) as a characteristic of text entry techniques. In Paternò, F., ed.: Human Computer Interaction with Mobile Devices. Volume 2411 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2002) 195–210
10. Masarykova univerzita, Rektorát: Dětská MASARYKOVA jUNIVERZITA MjUNI: Zhodnocení 2. ročníku 2015/2016 (2016) Internal material.
11. Allen, J.: Natural Language Understanding (2nd ed.). Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, USA (1995)
12. Pala, K., Smrž, P.: Building Czech Wordnet. Romanian Journal of Information Science and Technology **2004**(7) (2004) 79–88