

***CLARIN-PL in research practice: digital language analysis tools for the humanities and social sciences***

The second series of lectures and workshops, **May 18-20, 2015**

Wrocław University of Technology, Congress Centre, building D20, 8 Janiszewskiego Street,  
room 114

**CLARIN-PL and Wrocław Centre for Networking and Supercomputing invite you to participate in workshops on digital language analysis tools. You will learn in practice how to perform quantitative and qualitative analysis of natural language data. The workshops have been designed for researchers in the humanities and social sciences, and for translators**

Workshop leaders come from Wrocław University of Technology, University of Łódź, Institute of Computer Science of Polish Academy of Sciences, Polish-Japanese Academy of Information Technology, University of Wrocław, and Instytut Języka Polskiego of Polish Academy of Sciences

Anna Andrzejczuk, PhD; Maciej Eder, PhD; Elżbieta Hajnicz, DSc; Paweł Kędzia, MSc, Eng.; Jan Kocoń, MSc, Eng; Danijel Korzinek, MSc, Eng.; Krzysztof Marasek, DSc; Michał Marcińczuk, PhD, Eng; Marek Maziarz, PhD; Marcin Oleksy, PhD; Agnieszka Patejuk, MA; prof. Adam Pawłowski, DSc; Piotr Pęzik, PhD; Maciej Piasecki, PhD, Eng.; Adam Przepiórkowski, DSc; Ewa Rudnicka, PhD; Tomasz Walkowiak, PhD, Eng.; Marcin Woliński, PhD

CLARIN-PL is a Polish consortium, part of CLARIN, a Europe-wide Common Language Resources & Technology Infrastructure. Its mandate is the dissemination of language resources and automated language processing tools for applications in research based on the collection and analysis of large amounts on textual source material.

The majority of the workshops will discuss how research in the humanities and social sciences can leverage language tools and resources available via CLARIN-PL. You will learn about software systems for text classification, the construction of domain-specific lexicons from specialist texts, the recognition of proper names, and index construction; such systems are useful, e.g. in lexicography. You will find out how to build text corpora and make them widely available.

There will also be workshops meant for translators: how to use language corpora, in particular to extract and analyse phraseology. We will look at a collection of parallel texts in Polish and

English, study the concept of phraseological equivalence and its verification by corpus analysis.

We cordially invite all scientists interested in exploring new methods, techniques and tools in their work; no prior exposure to corpus linguistics is necessary. You may discover new avenues of research in literary studies, linguistics, lexicography, and translation and interpreting. Participation is free of charge.

If you are interested in participating in this series of workshops, send us email at **clarin-pl@pwr.edu.pl** by **May 8, 2015**. We regret that space is limited. Precedence will be give to those whom space limitations prevented from taking part in the first series of workshop. Other applicants will be accepted by the first-come-first-served rule. You must bring your own laptop in order to take full advantage of the practical nature of the workshops.

If you are interested in the events on one of three days, note it in your application.

There will be a third series of workshops in Cracow on June 17-19, 2015.

## **PROGRAM**

### **MONDAY, May 18**

#### **Research Infrastructure**

**9:00 – 10:00** Lecture

**CLARIN: Language Resources & Technology Infrastructure: depositing and disseminating resources and language tools for Polish**

Speakers: **Tomasz Walkowiak, PhD, Eng.; Maciej Piasecki, PhD, Eng.**

#### **Corpus tools**

**10:00 – 10:45** Lecture

**Collecting, annotating, and sharing corpora**

Speakers: **Marcin Oleksy, PhD; Jan Kocoń, MSc, Eng.**

**10:45 – 11:00** Coffee break

**11:00 – 12:30** Workshop

**Collecting, annotating, and sharing corpora**

**12:30 – 13:30** Lecture

### **Tools for automatic analysis of textual references**

Speakers: **Michał Marcińczuk, PhD, Eng.; Jan Kocoń, MSc, Eng.**

**13:30 – 14:15** Lunch break

**14:15 – 15:45** Workshop

**Tools for automatic analysis of textual references**

**15:45 – 16:30** Lecture

**Speech corpora and tools for their processing**

Speakers: **Krzysztof Marasek, DSc; Danijel Korzinek, MSc, Eng.**

**16:30 – 16:45** Coffee break

**16:45 – 17:30** Workshop

**Speech corpora and tools for their processing**

**TUESDAY, May 19**

**9:00 – 9:45** Lecture

**ChronoPress (Chronological Corpus for Polish Press) and its application in research**

Speaker: **Prof. Adam Pawłowski, DSc**

### **Dictionary tools**

**9:45 – 10:45** Lecture

**PIWordNet 3.0 – Polish lexical semantic network and its use in analysis of meaning**

Speakers: **Marek Maziarz, PhD; Paweł Kędzia, MSc, Eng.; Maciej Piasecki, PhD, Eng.**

**10:45 – 11:00** Coffee break

**11:00 – 12:00** Workshop

**PIWordNet 3.0**

**12:00 – 12:30** Lecture

**Bilingual WordNet – possibilities for application in translating**

Speaker: **Ewa Rudnicka, PhD**

**12:30 – 13:30** Workshop

**Bilingual WordNet**

**13:30 – 14:15** Lunch break

**14:15 – 15:00** Lecture

**Tools for automatic extraction of collocation dictionaries and for assessment of collocations**

Speakers: **Maciej Piasecki, PhD, Eng.; Marek Maziarz, PhD**

**15:00 – 16:00** Workshop

**Tools for automatic extraction of collocation dictionaries and for assessment of collocations**

**16:00 – 16:15** Coffee break

**16:15 – 17:00** Lecture

**Reference and parallel corpora in translation**

Speaker: **Piotr Pęzik, PhD**

**17:00 – 18:00** Workshop

**Reference and parallel corpora in translation**

**WEDNESDAY, May 20**

**Tools for textual analysis**

**9:00 – 9:30** Lecture

**Morfeusz 2 – analyser and inflection generator for Polish**

Speakers: **Marcin Woliński, PhD, Anna Andrzejczuk, PhD**

**9:30 – 10:15** Workshop

**Building domain dictionaries for Morfeusz 2**

**10:15 – 10:30** Coffee break

**10:30 – 11:05** Lecture

**A Polish Valence Dictionary**

Speaker: **Elżbieta Hajnicz, DSc**

**11:05 – 11:45** Workshop

**A Polish Valence Dictionary**

**11:45 – 12:00** Coffee break

**12:00 – 12:45** Lecture

**LFG syntactic parsing and its application**

Speaker: **Prof. Adam Przepiórkowski, DSc; Agnieszka Patejuk, MA**

**12:45 – 13:45** Workshop

## **LFG syntactic parsing and its application**

**13:45 – 14:30** Lunch break

**14:30 – 15:15** Lecture

**Conversational register of Polish, i.e., real-time discourse based on data obtained from Spokes**

Speaker: **Piotr Pezik, PhD**

**15:15 – 16:00** Workshop

**Conversational register of Polish**

**16:00 – 16:15** Coffee break

**16:15 – 17:15** Lecture

**Text classification and stylometric analysis system**

Speakers: **Maciej Piasecki, PhD, Eng.;** **Maciej Eder, PhD**

**17:15 – 18:00** Workshop

**Text classification and stylometric analysis system**