

## **CLARIN-PL workshops and lectures in Gdańsk, 18–19 May 2018**

JOANNA REDZIMSKA  
DANUTA STANULEWICZ  
MAGDALENA WAWRZY尼亚K-ŚLIWSKA

*Received 25.06.2018,  
accepted 29.06.2018.*

### **1. CLARIN-PL**

CLARIN-PL is part of CLARIN (Common Language Resources and Technology Infrastructure), a European research infrastructure. Its aim is to facilitate work with large collections of texts in the humanities and social sciences (see <<http://clarin-pl.eu>>, <<http://clarin-pl.eu/en/home-page/>>). The partners of CLARIN-PL include the following institutions:

- Grupa Technologii Językowych G4.19 Politechniki Wrocławskiej, Wrocław University of Science and Technology;
- The Linguistic Engineering Group, Polish Academy of Sciences;
- Polish-Japanese Academy of Information Technology, Institute of Computer Science;
- Instytut Sławistyki PAN, Polish Academy of Sciences;
- University of Łódź;
- University of Wrocław.

The website of CLARIN-PL (<<http://clarin-pl.eu>>) hosts numerous language resources, including, *inter alia*:

- Paralela, a Polish-English parallel corpus, available at <<http://paralela.clarin-pl.eu/>>;
- ChronoPress: Chronologiczny Korpus Polskich Tekstów Prasowych (1945-1954), a corpus of Polish press texts 1945-1954, available at <<http://chronopress.clarin-pl.eu/>>;
- Słowa Dnia (Words of the Day), the most frequent words in the Polish press, available at <<http://slowadnia.clarin-pl.eu/#/default/1060>>;
- Słowosieć (Pl Wordnet), available at <<http://plwordnet.pwr.wroc.pl/wordnet/>>;
- Walenty, a valency dictionary of the Polish language, available at <<http://walenty.ipipan.waw.pl/>>;
- Spokes, conversational data resources, available at <<http://spokes.clarin-pl.eu/>>.

On the CLARIN-PL website, one may also find tools for analyzing language, including, *inter alia*,

- Nowy Morfeusz, a tool used for morphological analysis, available at <<http://sgjp.pl/morfeusz/>>;
- Inforex, a system used for editing of annotated corpora, available at <<https://inforex.clarin-pl.eu/>>;
- Mapa Literacka, used to recognize references of geographical names, available at <<http://litmap.clarin-pl.eu/>>;
- Transkrypcja Fonetyczna, a tool used for phonetic transcription, available at <<http://mowa.clarin-pl.eu/transcriber/>>;
- Chunker, a tool used for syntactic analysis, available at <<http://ws.clarin-pl.eu/chunker.shtml>>;
- Mowa, a speech processing tool, available at <<http://ws.clarin-pl.eu/chunker.shtml>>;
- Parser, used to analyze the Polish language, available at <<http://ws.clarin-pl.eu/parser.shtml>>.

## 2. The lectures and workshops

On 18-19 May 2018, the Institute of English and American Studies, University of Gdańsk, hosted the workshops and lectures “CLARIN-PL w praktyce badawczej: Cyfrowe narzędzia

do analizy języka w naukach humanistycznych i społecznych [CLARIN-PL in research practice: IT tools used in language analysis in the humanities and social sciences]”. It should be added that it was the 7th cycle of the CLARIN-PL workshops and lectures.<sup>1</sup>

The participants – university lecturers and students as well as employees at IT companies from all over the country – had a splendid opportunity to attend numerous workshops and lectures offered by specialists in natural language processing and IT technology.

The participants were warmly welcomed by the host, Prof. Maciej Michalski, the Dean of the Faculty of Languages, and by Dr Maciej Piasecki, representing CLARIN-PL. Then they were introduced to the topic of the workshops, listening to a lecture presenting the possibilities offered by CLARIN-PL:

- (1) “CLARIN – infrastruktura naukowa technologii językowych – wprowadzenie [CLARIN – the research infrastructure of language technologies: An introduction” by Maciej Piasecki.

The lectures and workshops that followed were divided into three sections, one of which was designed for students of Natural Language Processing, a specialty offered to MA students by the Institute of English and American Studies, University of Gdańsk. It should be added that other students also participated in the workshops.

During the two days of the conference, CLARIN-PL offered the following workshops and lectures:

- (2) “Zasoby leksykalne CLARIN-PL: wielki relacyjny słownik semantyczny – Słowosieć, Słowosieć polsko-angielska, Słownik znakowany wydźwiękiem emocjonalnym, Walenty – słownik walencyjny [CLARIN-PL’s lexical resources: the great relational semantic dictionary Słowosieć, the Polish-English wordnet,

---

<sup>1</sup> See our report of the CLARIN-PL workshops and lectures at the University of Łódź which were held in February 2017 (Redzimaska, Stanulewicz and Wawrzyniak-Śliwska 2017).

- the sentiment dictionary, the valence dictionary Walenty]” run by Agnieszka Dziob, Ewa Rudnicka, Tomasz Naskręt and Elżbieta Hajnicz;
- (3) “Tworzenie, przeglądanie i anotacja korpusów tekstów przy użyciu narzędzi: Korpusomat, DSpace, Inforex [Creating, using and annotating text corpora using tools including Korpusomat, DSpace, Inforex]” by Marcin Oleksy, Michał Marcińczuk and Zbigniew Gawłowicz;
  - (4) “Narzędzia i zasoby do przetwarzania mowy: korpusy, segmentacja, rozpoznawanie i analiza mowy; transkrypcja [Tools and resources for speech processing: corpora, segmentation, speech recognition and analysis; transcription]” run by Daniel Korżinek;
  - (5) “Analiza semantyczna tekstu, wydobywanie informacji (stylometria, proste statystyki tekstu, modelowanie tematyczne, rozpoznawanie nazw własnych, rozpoznawanie terminów, analiza wydzźwięku LEM, WebSty, TermoPL) [Semantic analysis of the text, information extraction (stylometry, simple text statistics, thematic modeling, recognition of proper names, recognition of terms, LEM sentiment analysis, WebSty, TermoPL)]” run by Maciej Piasecki, Tomasz Walkowiak and Agnieszka Mykowiecka;
  - (6) “Korzystanie z korpusów konwersacyjnych i równoległych: SPOKES i Paralela [Use of conversational and parallel corpora: SPOKES and Paralela]” run by Piotr Pęzik;
  - (7) “Analizy chronologiczne z wykorzystaniem webserwisu ChronoPress [Chronological analyzes using the web service ChronoPress]” run by Adam Pawłowski.

The students had an opportunity to participate in the following lectures and workshops which were run by Jan Wiczorek and other scholars who conducted the workshops enumerated above:

- (8) Narzędzia i zasoby do przetwarzania mowy: korpusy, segmentacja, rozpoznawanie i analiza mowy; transkrypcja [Tools and resources for speech processing: corpora, segmentation, speech recognition and analysis; transcription]”;
- (9) Analiza semantyczna tekstu, wydobywanie informacji (stylometria, proste statystyki tekstu, modelowanie tematyczne,

- rozpoznawanie terminów, analiza wydźwięku LEM, WebSty, TermoPL) [Semantic analysis of the text, information extraction (stylometry, simple text statistics, thematic modeling, recognition of terms, LEM sentiment analysis, WebSty, TermoPL)];
- (10) Tworzenie oraz zarządzanie korpusami tekstów (DSpace, Inforex, Korpusomat) [Creation and management of text corpora (DSpace, Inforex, Korpusomat)];
- (11) Zasoby leksykalne CLARIN-PL: wielki relacyjny słownik semantyczny – SłowoSieć, SłowoSieć polsko-angielska, Słownik znakowany wydźwiękiem emocjonalnym, Walenty – słownik walencyjny [CLARIN-PL's lexical resources: the great relational semantic dictionary SłowoSieć, the Polish-English wordnet, the sentiment dictionary, the valence dictionary Walenty].

### 3. A final word

The workshops offered by CLARIN-PL provided their participants – especially the students of Natural Language Processing as well as other students of the University of Gdańsk – with an excellent opportunity to get acquainted with the different language resources and tools available on its website. Undoubtedly, the students will use these resources and tools in their projects, including preparing their MA theses.

### References

- CLARIN-PL. Available at <<http://clarin-pl.eu/pl>>. Accessed 20.06.2018.
- Redzimska, Joanna, Danuta Stanulewicz, Magdalena Wawrzyniak-Śliwska (2017). “CLARIN-PL workshops in Łódź, 3–4 February 2017”. *Beyond Philology* 14/1: 251-257.

Joanna Redzimska  
ORCID iD: 0000-0001-7837-9397  
Instytut Anglistyki i Amerykanistyki  
Uniwersytet Gdański  
ul. Wita Stwosza 51  
80-308 Gdańsk  
Poland  
joared@wp.pl

Danuta Stanulewicz  
ORCID iD: 0000-0003-1792-3883  
Instytut Anglistyki i Amerykanistyki  
Uniwersytet Gdański  
ul. Wita Stwosza 51  
80-308 Gdańsk  
Poland  
danuta.stanulewicz@gmail.com

Magdalena Wawrzyniak-Śliwska  
ORCID iD: 0000-0002-6795-9481  
Instytut Anglistyki i Amerykanistyki  
Uniwersytet Gdański  
ul. Wita Stwosza 51  
80-308 Gdańsk  
Poland  
magdalenaws@ug.edu.pl