



C.R.I.T.



Centre de Recherches
Interdisciplinaires
et Transculturelles
EA 3224

16th INTERNATIONAL NOOJ 2022 CONFERENCE

Book of Abstracts

Hybrid Conference

Rosario, Santa Fe, Argentina

June 14-16, 2022

Conference Venue

**ECU (Espacio Cultural Universitario, UNR)
Rosario, Santa Fe, Argentina**

**Silvia Susana Reyes
Andrea Rodrigo
Max Silberztein
Carolina Tramallino (Eds.)**

A linguistic approach for automatic analysis, recognition and translation of Arabic predicative nouns Cheikhrouhou Hajer & Imed Lahyani	31
Creation of parallel medical and social corpora for the machine translation and speech synthesis Mikita Suprunchuk, Nastassia Yarash, Yuras Hetsevich, Valerii Varanovich, Siarhey Gaidurau, Yauheniya Zianouka & Palina Sakava	33
Processing the discourse of insecurity in Rosario with the NooJ platform Andrea Rodrigo, Silvia Reyes & Mariana González	34
IMF's Impact on Argentina: using NooJ to automatically analyze the print media literature on Argentina's debt affairs Carmen González, Andrea Fernanda Rodrigo & Mariana González	35
Natural Language Processing Applications	36
The digital text workshop cloud, new solutions for super calculation environments Iliaria Veronesi, Rita Bucciarelli, Francesco Saverio Tortoriello, Andrea Rodrigo, Marianna Greco, Colomba La Ragione, Javier Julian Enriquez	37
Construction of an educational game "CONJ_NooJ" Hela Fehri & Nizar Jarray	39
From questions in natural language to SPARQL queries Ismahane Kourtin	40
Integrated NooJ environment for Arabic Linguistic Disambiguation using MWEs Dhekra Najjar & Slim Mesfar	41
The Use of NooJ Platform to Build a Formative Assessment for Arabic language Ilham Blanchete & Mohammed Mourchid	42
Using First Language Grammar to Support Second Language Grammar Acquisition in an Argentinian ESL Classroom by Resorting to the NooJ Platform Mariana González & Andrea Rodrigo	43
Extraction of phrasemes in the electrical field by NooJ Tong Yang	44
Towards a change of paradigm in language teaching in teacher training with computer tools Virginia Gonfiantini & Andrea Rodrigo	45

Mikita Suprunchuk

Minsk State Linguistic University, Zakharov str. 21, Minsk, Belarus

ms@philology.by

Nastassia Yarash

United Institute of Informatics Problems, Minsk, Belarus

anyarosh62@gmail.com

Yuras Hetsevich

United Institute of Informatics Problems, Surhanava str. 6, Minsk, Belarus

yuras.hetsevich@gmail.com

Valerii Varanovich

Belarusian State University, Kurchatova str. 5, Minsk, Belarus

gamrat.vvv@gmail.com

Siarhey Gaidurau

United Institute of Informatics Problems, Minsk, Belarus

Yauheniya Zianouka

United Institute of Informatics Problems, Surhanava str. 6, Minsk, Belarus

evgeniakacan@gmail.com

Palina Sakava

United Institute of Informatics Problems, Minsk, Belarus

polina.sakova.work@gmail.com

In the Republic of Belarus both Belarusian and Russian have the status of official language. Since there is bilingualism at the legislative level, it is necessary to provide the availability of texts for different purposes in both official languages. One of the important spheres is public education in medical and social fields.

Medical and social texts contain essential information. The first are intended to convey information between health professionals and scientists (for example, in articles, research papers) or to inform the population about public and personal health. The second group represents the historical and cultural heritage of Belarus and is aimed at acquainting country visitors with it.

Therefore, the need for a system of machine translation and speech synthesis in Belarusian and Russian is still relevant. There is a data layer of mentioned domains poorly covered with such systems, which, in turn, require large parallel corpora for their training. There are no large Belarusian-Russian parallel corpora on open access. So, the first step in the machine translation and speech synthesis development may be the creation of corresponding corpora. Using proofreading and word processing services and NooJ processor [1; 2], it is possible to work with big data to further use it for training machine systems.

Currently, we work at creating a corpora of medical and social domains. A medical corpus of 848 parallel texts in Belarusian (303,469 word forms), English (373,709 word forms) and Russian (330,996 word forms) has been compiled. A trilingual parallel corpus of social domain (based on parallel texts in Belarusian, English and Russian from the “KrokApp” travel audio guide [3]) is being created. Development of specialized syntactic grammars for the Belarusian NooJ module, which will allow the analysis of set phrases specific for medical and social texts, is planned.

References

- [1] Hetsevich, Y. [et al.] (2021). Creation of a legal domain corpus for the Belarusian NooJ module: texts, dictionaries, grammars. Bigey, M. [et al.] (Eds.). *15th International Conference NooJ 2021: book of abstracts*. Besançon. (36-37).
- [2] Drahun, A. (2019). Semi-Automatic Proofreading of Belarusian and English texts. Drahun, A. [et al.]. *International Conference NooJ 2019: Book of abstracts*. Hammamet, Tunisia.
- [3] KrokApp – personal audio guide in Belarus. <https://krokapp.by/about/>.