The Innovation Bakery™

X23 Science in Society, an independent research center and social enterprise is pleased to introduce

THE 18TH NOOJ INTERNATIONAL CONFERENCE, 2024

Book of Abstracts

Bergamo (Italy), June 5th to 7th, 2024

INFO AT: **https://nooj2024.x-23.org**

Editors

Agathe Semlali
X23 Science in Society, Italy
Treviglio, Italy

Cecilia Olivieri
X23 Science in Society, Italy
Treviglio, Italy

Lorenzo Maggio Laquidara
Duke Kunshan University, China [1]; X23 Science in Society, Italy [2]
Kunshan, Suzhou, China [1]
Treviglio, Italy [2]

Jiwon Shin
X23 Science in Society, Italy
Treviglio, Italy

Mario Monteleone
University of Salerno
Salerno, Italy

Max Silberztein
Université de Bourgogne Franche-Comté
Besançon, France

—

Publisher

X23 Science in Society **(www.x-23.org)**

Organization

Partner Institutions

Organised by
X23 Science in Society

Commissioned by
NooJ Association

and kindly Hosted by
Department of Foreign Languages, Literatures and Cultures
University of Bergamo, Italy

Organising Committee

Agathe Semlali	(X23 Science in Society, Italy)
Cecilia Olivieri	(X23 Science in Society, Italy)
Jiwon Shin	(X23 Science in Society, Italy)
Lorenzo Maggio Laquidara	(Duke Kunshan University, China [1]; X23 Science in Society, Italy [2])
Marika Mazzi Boém	(X23 Industry, Italy)
Mario Monteleone	(University of Salerno, Italy)
Max Silberztein	(Université de Bourgogne Franche-Comté, France)

Scientific Committee

Max Silberztein (Program Committee Chair)	(Université de Bourgogne Franche-Comté, France)
Agata Jackievicz	(Université Paul Valéry, France)
Agathe Semlali	(X23 Science in Society, Italy)
Agnès Tutin	(Université de Grenoble, France)
Anabela Barreiro	(INESC-ID, Portugal)
Andrea Rodrigo	(University of Rosario, Argentina)

Automatic Processing of Free Word Combinations Containing Quantitative Expressions in NooJ

Yauheniya Zianouka

United Institute of Informatics Problems
Minsk, Belarus
evgeniakacan@gmail.com

David Latyshevich

United Institute of Informatics Problems
Minsk, Belarus
david.latyshevich@gmail.com

Yuras Hetsevich

United Institute of Informatics Problems
Minsk, Belarus
yuras.hetsevich@gmail.com

Abstract

The core tasks of speech synthesis and recognition involve a set of operations for text processing, including text normalization and prosodic segmentation. Previous research has focused on syntagmatic delimitation of Belarusian literary texts using NooJ syntactic grammars [1–3]. Their realization is based on the use of rules-based approaches and grammatical patterns for analysing language properties by means of the NooJ environment [4].

Another important aspect is the automatic processing of word combinations containing quantitative expressions (QEMUs) [5]. The goal of normalization is to standardize the representation of these elements (a numeral quantifier – a number or a numeral – and a symbol or word that describes it) to facilitate their annotation and analysis. The study of such units is complicated by the peculiarities of the Belarusian language as an inflectional language [6]. For example, Belarusian has diverse grammatical categories with different morphological forms. In our case, an adjective/numeral must agree with a noun in six cases, number, and gender. It also can be difficult to identify dependencies between words and phrases, as these dependencies are often expressed through morphological forms and inflectional endings.

To accomplish this research, we will create a corpus of scientific-journalistic texts for audio guides. The next step is to extract all quantitative expressions (years, dates, periods, Latin and Roman numerals, percents, and area measurements) using NooJ and create a set of syntactic grammars for expanding them into verbal implementation and separate syntagma. These resources will be integrated into Belarusian NLP applications for various areas, such as syntactic parsing, prosody delimitation, text-to-speech synthesis, and more, serving as a means of voicing audio guides.

Keywords

Prosodic segmentation, Quantitative expression, a Corpus, Automatic processing.

References

1. Hetsevich, Y., Okrut, T. & Lobanov, B. (2016). Grammars for the Sentence into Phrase Segmentation: Punctuation Level. In: T. Okrut, Y. Hetsevich, M. Silberztein, H. Stanislavenka (eds.). International Scientific Conference on the Automatic Processing of Natural-Language Electronic Texts "NooJ'2015": Abstracts (11–13 June, 2015, Minsk, Belarus), Revised selected papers, 74-82. Springer International Publishing.
2. Zianouka Ya., Hetsevich Y., Latyshevich D. & Dzenisiuk Z. (2022). Automatic Generation of Intonation Marks and Prosodic Segmentation for Belarusian NooJ Module. In: Magali Bigey, Annabel Richeton, Max Silberztein, Izabella Thomas (eds.). 15th International Conference, NooJ 2021, Besançon, France, June 9-11, 2021, Revised Selected Papers, 231-242. Besançon: Springer, Cham.

3. Zianouka, Ya, Hetsevich Yu., Suprunchuk M. & Latyshevich D. (2022). Prosodic Segmentation of Belarusian Texts in NooJ. In: Mariana González, Silvia Susana Reyes, Andrea Rodrigo, Max Silberztein (eds.). 16th International Conference, NooJ 2022, 50-62. Universidad Nacional de Rosari, Rosario : Springer Cham.
4. A Linguistic Development Environment. http://www.nooj4nlp.org/. Last accessed 2022/07/21.
5. Borodina, Ju. & Hetsevich, Yu. (2016). Using NooJ to Process Satellite Data. In: ed. T. Okrut, Y. Hetsevich, M. Silberztein, H. Stanislavenka (eds.). Automatic Processing of Natural-Language Electronic Texts with NooJ: 9th International Conference, NooJ 2015, Minsk, Belarus, June 11-13, 2015, Revised Selected Papers, pp. 182-190. Springer International Publishing.
6. Skopinova, A., Hetsevich, Y. & Borodina, J. (2015). Converting quantitative expressions with measurement unit into an orthographic form and convenient monitoring for Belarusian. In: Monti, J., Silberztein, M., Monteleone, M., Di Buono, M.P. (eds.) Formalising Natural Languages with Nooj 2014, pp. 175-186. Cambridge Scholars Publishing, UK, Newcastle.