



**Sveučilište u Zadru**  
Universitas Studiorum  
Jadertina | 1396 | 2002 |

---

# THE 17<sup>th</sup> NOOJ INTERNATIONAL CONFERENCE 2023

## Book of Abstracts

**Zadar, May 31 – June 2, 2023**

---

## Editors

---

**Linda Mijić**

Department of Classical Philology  
University of Zadar  
Zadar, Croatia

**Anita Bartulović**

Department of Classical Philology  
University of Zadar  
Zadar, Croatia

**Marijana Tomić**

Department of Information Sciences  
University of Zadar  
Zadar, Croatia

**Laura Grzunov**

Department of Information Sciences  
University of Zadar  
Zadar, Croatia

**Kristina Kocijan**

Department of Information and Communication Sciences  
Faculty of Humanities and Social Sciences  
University of Zagreb  
Zagreb, Croatia

**Max Silberztein**

Université de Franche-Comté  
Besançon, France

## Publisher

University of Zadar  
For the Publisher: Dijana Vican, rector

**ISBN 978-953-331-421-1**

<b>A Prototype of Indonesian Multi-Level Tagger: SANTI-Network</b>	
Prihantoro .....	25
<b>NooJ Dictionary for Rromani Language: Importing of a Published Dictionary to the NooJ System</b>	
Masako Watabe .....	27
<b>Deciphering the Nomenclature of Chemical Compounds in NooJ</b>	
Kristina Kocijan, Krešimir Šojat, Tomislav Portada .....	29
<b>Croatian Cognition Verbs in Machine Sentence Processing</b>	
Marta Petrak, Bojana Mikelenić, Marko Orešković .....	31
<b>NooJ Dictionary of Croatian and English Internet Slang</b>	
Ivan Cota .....	33
<b>Latin Pronouns, Numbers and Prepositions in the NooJ Tool</b>	
Anita Bartulović, Linda Mijić.....	35
<b>SYNTACTIC &amp; SEMANTIC RESOURCES.....</b>	<b>37</b>
<b>Disambiguation Grammars for the Ukrainian Module</b>	
Olena Saint-Joanis .....	38
<b>A Proposal for the Processing of the Nucleus Verb Phrase of Pronominal (SVNPr) Verbs in Spanish</b>	
Andrea Rodrigo, Rodolfo Bonino, Silvia Reyes .....	40
<b>A Rioplatense Spanish Date Grammar Using the NooJ Platform</b>	
Mariana González .....	42
<b>Parafrasário: A Variety-Based Paraphrasary for Portuguese</b>	
Anabela Barreiro, Ida Rebelo, Cristina Mota .....	44
<b>CORPUS LINGUISTICS &amp; DISCOURSE ANALYSIS .....</b>	<b>46</b>
<b>The Limitations of Training Corpus-Based Methods in NLP</b>	
Max Silberztein .....	47
<b>Syntactic-Semantic Analysis of Perception Verbs in the Croatian Language</b>	
Daša Farkaš, Kristina Kocijan .....	49
<b>The Automatic Translation of Arabic Psychological Verbs Using NooJ Platform</b>	
Asmaa Amzali, Asmaa Kourtin, Mohammed Mouchid, Abdelaziz Mouloudi, Samir Mbarki .....	51
<b>Automatic Disambiguation of the Belarusian-Russian Legal Parallel Corpus in NooJ</b>	

Valery Varanovich, Mikita Suprunchuk, Yauheniya Zianouka, Yuras Hetsevich, Nastassia Yarash .....	53
<b>Advances in the Automatic Treatment of Newspaper Articles on Economics Journalism Using NooJ</b>	
Carmen González .....	55
<b>Sentiment Analysis of Texts Written in Arabic: Addressing the Issue of Negation</b>	
Mohamed El Ammari, Azeddine Rhazi, Salim Rami .....	57
<b>A Linguistic Approach for MDU-Based Segmentation</b>	
Chahira Lhioui, Malek Lhioui, Mounir Zrigui .....	59
<b>Embracing a Plant-Based Diet: A NooJ Analysis</b>	
Isabella Cossidente, Alessandra D'Agostino .....	61
<b>Differences Between Hate Comments and Insult Comments Directed to Men and Those Directed Towards Women</b>	
Martina Galović, Damir Puškarić .....	63
<b>Explicit Language in English Song Lyrics: Should We Be Worried?</b>	
Mila Bikić, Valerija Bočkaj .....	65
<b>Comparison of the Representation of Male vs. Female Athletes in Croatian News Portals Using a NooJ Syntax Grammar</b>	
Klara Kozolić, Krešimir Štimac .....	67
<b>Immigrant in the Light of Language Production</b>	
Barbara Vodanović .....	69
<b>Semantic Analysis of Migrants' Self-Entrepreneurship Ecosystem Narratives</b>	
Cecilia Olivieri, Jie Sheng, Lorenzo Maggio Laquidara, Agathe Semlali ...	71
<b>Reviewing the Position of the “Other” in Croatia’s “Non-European” Collections in the Second Half of the 20th Century Using NLP</b>	
Martina Bobinac .....	74
<b>Engaging with the Agenda 2030</b>	
Stella Nunzia Costanza, Antonio Pagano, Antonio Duca .....	76
<b>NATURAL LANGUAGE PROCESSING APPLICATIONS .....</b>	<b>78</b>
<b>NooJ Grammars for Morphophonemic Continuity and Semantic Discontinuity Title</b>	
Mario Monteleone .....	79

# Automatic Disambiguation of the Belarusian–Russian Legal Parallel Corpus in NooJ

---

Valery Varanovich  
Mikita Suprunchuk  
Yauheniya Zianouka  
Yuras Hetsevich  
Nastassia Yarash

United Institute of Informatics Problems  
Minsk, the Republic of Belarus  
ssrlab221@gmail.com

## Abstract

Homonymy is still the central problem of automatic text processing at the lexical level. Automatic (rarely semi-automatic) solving of lexical ambiguity was first formulated within the field of science and technology related to the creation of machine translation systems. To date, this is a critical problem of improving the quality of systems for various branches of computational linguistics (Varanovich, 2009).

There are two dominant classes of ambiguity resolution mechanisms (Bouarroudj et al., 2022):

- (1) Automatic one, implying a fully computerised solution to this problem.
- (2) Interactive (dialogic, semi-automatic) one, supposing a joint solution by a person and a computer.

It means that the user has a set of alternatives from which he should choose one option.

We are conducting work on creating a legal texts corpus in Belarusian and Russian, which is used in various software products (speech synthesis, machine translation, spell checker, etc.). A trilingual Belarusian–Russian–English dictionary of legal terms was created (Hetsevich et al., 2021) within the project. Also, in the process of forming the corpus, contextual dictionaries are compiled, which can become the basis for high-priority dictionaries in the NooJ system. The dictionaries reflect contexts that indicate the preferred translation of a certain term from Russian into Belarusian (Barabash, 2015).

We are planning to create high-priority dictionaries for each of the 26 law codes of Belarus. It is assumed that a lot of diagnostic contexts will be the same in different codes (*настоящий* ‘this’ = *зэты* ‘this’ (code), not *настоящий* ‘this’ = *сапраўдны* ‘real’ (code)), but we hypothesize that in some cases the contexts for the same values will be different, and it is also possible that in different codes, one polysemous word (or a homonym) will have different meanings.

Here are some words with several meanings which were found during our work: *данный* – ‘1) гэты, this; 2) дадзены, given’; *отпуск* – ‘1) выдача (тавараў), issuance, supply; 2) адпачынак (даць), holiday, vacation’. These meanings (translations) could be chosen according to their neighbor words. As NooJ is an effective tool for solving word ambiguity, we plan to use it for compiling syntactic grammars. They will search for homonyms by analyzing the context (the sequence of words) and form a list of various lexical units for different domains. This will assist in identifying terms in different thematic domains, which is very important for compiling special vocabularies for indicated fields.

## Key words

*Disambiguation, homonym, automatic text processing, dictionary, legal texts corpus*

## References

- [1] Barabash, O. V. (2015) Razgranicheniie omonimii i polisemii juridichieskikh tierminov (= Барабаш О. В., Разграничение омонимии и полисемии юридических терминов). *Rhema. Рема*, 2, 39–51. Available at: <https://cyberleninka.ru/article/n/razgranichenie-omonimii-i-polisemii-yuridichieskikh-terminov>.
- [2] Bouarroudj, W., Boufaïda, Z., Bellatreche, L. (2022) Named Entity Disambiguation in Short Texts over Knowledge Graphs. *Knowledge and Information Systems*, 64(2), 325–351. Available at: <https://doi.org/10.1007/s10115-021-01642-9>.
- [3] Hetsevich, Y. et al. (2021) Creation of a Legal Domain Corpus for the Belarusian NooJ Module: Texts, Dictionaries, Grammars. In: Bigey, M. et. al. (eds.) 15th International Conference NooJ 2021: Book of Abstracts, 36–37. Available at: <http://www.nooj-association.org/nooj2021/Abstracts.pdf>.
- [4] Silberztein, M. (2003–) NooJ Manual. Available at: <http://www.nooj-association.org>.
- [5] Varanovich, V. V. (2009) Slovar lieksichieskikh valientnostiej v sistiemie russko-bielorusskogo mashinnogo pierievoda (= Воронович, В. В. Словарь лексических валентностей в системе русско-белорусского машинного перевода). In: Третьи чтения, посвященные памяти В. А. Карпова: сб. науч. ст., 108–111. БГУ, Минск. Available at: <http://elib.bsu.by/handle/123456789/10645>.