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

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
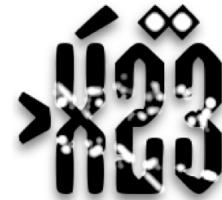
Université de Bourgogne Franche-Comté

Besançon, France

——

Publisher

Organization

Partner Institutions

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)

Anita Bartulović   (University of Zadar, Croatia)
Cecilia Olivieri   (X23 Science in Society, Italy)
Christian Boitet   (Université Joseph Fourier, Grenoble, France)
François Trouilleux   (Université Clermont Auvergne, France)
Giuseppe Giordano   (University of Salerno, Italy)
Héla Fehri   (University of Sfax, Tunisia)
Ignazio Mauro Mirto   (University of Palermo, Italy)
Izabella Thomas   (Université de Franche-Comté, France)
Johanna Monti   (University of Salerno, Italy)
Kristina Kocijan   (University of Zagreb, Croatia)
Laetitia Leonarduzzi   (Université d'Aix-Marseille, France)
Linda Mijić   (University of Zadar, Croatia)
Lorenzo Maggio Laquidara   (Duke Kunshan University, China [1] - X23 Science in Society, Italy [2])
Magali Bigey   (Université de Franche-Comté, France)
Marco Angster   (University of Zadar, Croatia)
Marika Mazzi Boém   (X23 Industry, Italy)
Marko Tadić   (University of Zagreb, Croatia)
Michelangelo Misuraca   (University of Calabria, Italy)
Philippe Lambert   (Université de Lorraine, France)
Samir Mbarki   (IbnTofail University, Morocco)
Stefania Maria Maci   (University of Bergamo, Italy)
Thierry Poibeau   (Laboratoire Lattice, CNRS, France)
Valentina Piunno   (University of Bergamo, Italy)
Walter Koza   (National University of General Sarmiento, Argentina)
Xavier Blanco   (Autonomous University of Barcelona, Spain)
Yuras Hetsevich   (National Academy of Sciences, Belarus)
Zoe Gavriilidou   (Democritus University of Thrace, Greece)

Contents

Automatic Grammatical Disambiguation in Belarusian and Russian Legal Domain

Valery Varanovich

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

Mikita Suprunchuk

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

Yauheniya
Zianouka

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

Yuras
Hetsevich

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

Abstract

The homonymy of language units at different levels is one of the most intractable problems in natural language processing [2; 4]. In 2023, the Belarusian NooJ team analyzed the homonymy of various types in the Belarusian language on the basis of composed legal domain corpus in NooJ [3]. Currently, we focus on one type – grammatical homonymy (coincidence of different word forms). It is a very common phenomenon in Slavic languages, including Belarusian [1]. For example, the forms of nominative and accusative cases of inanimate masculine nouns (край 'edge'), forms of genitive, dative, instrumental, and prepositional cases of feminine adjectives (маленькай 'small') and many others regularly match.

Recognizing grammatical homonymy is a core task for creating machine translation systems. Even when translating closely related languages, the case forms may vary: for example, case changes in the expression загадчык кафедры 'head of the department' while translating from Belarusian to Russian: the form of the genitive case is homonymous with the dative case, – заведующий кафедрой (the instrumental case).

In text-to-speech systems, the problem of stress arises in languages with non-fixed and mobile stress, including Belarusian. There are forms within the paradigm that differ only in stress: рукі [ru'ki] 'hand' is the genitive singular, and рукі ['ruki] is the nominative plural. Therefore, it is crucially important to identify which form of word is used in a particular sentence.

Previous research [3] revealed that grammatical homonymy is the most common among other types of homonymy in legal texts. This type covers about 25% of all cases, but it is already 78% in the total number of different words (lexemes). Within this work, we plan to use the full corpus of codes and conduct a detailed study, finally clarify the degree of prevalence of grammatical homonymy in Belarusian legal texts, and create NooJ grammars to solve grammatical homonymy.

Key words

Automatic text processing, Disambiguation, Grammatical homonymy, Legal texts corpus, NooJ

References

1. Barabash, O. V. (2015). Differentiation of homonymy and polysemy of legal terms (= in Russian: Барабаш, О. В.: Разграничение омонимии и полисемии юридических терминов). Rhema.=Рема (2), 39–51. https://cyberleninka.ru/article/n/razgranichenie-omonimii-i-polisemii-yuridicheskih-terminov, last accessed 2024/01/03.

2. Rakhilina, E. & Kobritsov, B. (2006). Semantic ambiguity as an application-oriented problem: word class tagging in the RNC (= in Russian: Рахилина, Е. В., Кобрицов, Б. П. Многозначность как прикладная проблема: лексико-семантическая разметка в Национальном корпусе русского языка). In: Laufier, N. (ed.) Kompjutiernaja lingvistika i intielliektualnyje tiekhnologii «Dialog 2006» (= Компьютерная лингвистика и интеллектуальные технологии: труды международной конференции «Диалог 2006» / под ред. Н. Лауфер), pp. 445–460. RGGU Publisher, Moscow.

3. Varanovich, V., Suprunchuk, M., Zianouka, Y., Dolgova, A. & Hetsevich, Y. (2023). Creation of a legal domain corpus for the Belarusian module in NooJ: texts, dictionaries, grammars. In: Rodrigo, M., Silberztein, M. (eds.) 16th International Conference NooJ 2022, pp. 151–162. Springer Cham, Rosario (2023).

4. Zalizniak, A. A. (2006). Ambiguity and ways of representing it. Jazyki slavianskikh kultur, Moscow. (= in Russian: Зализняк, А. А.: Многозначность и способы ее представления. Языки славянских культур, Москва.)