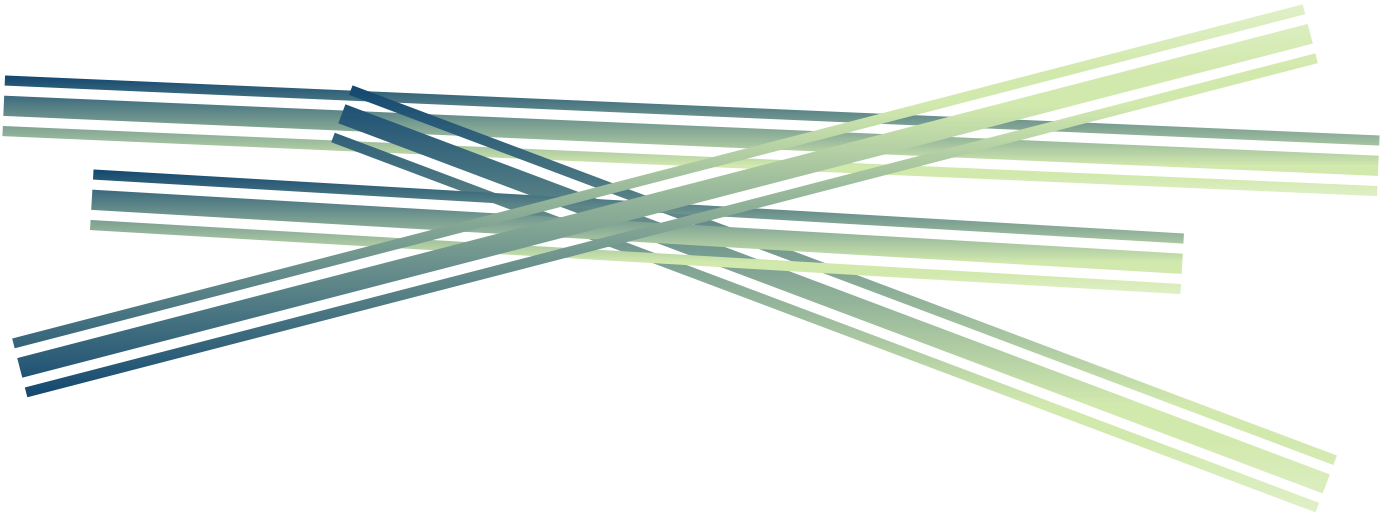

15th INTERNATIONAL CONFERENCE

NooJ 2021

Book of Abstracts

Virtual conference
Besançon, France
June 9-11, 2021

Magali Bigey
Annabel Richton
Max Silberztein
Izabella Thomas (Eds.)

Organization

14th Annual NooJ 2021 International Conference was organized as a Virtual conference.

Organizing Institutions

UNIVERSITE DE FRANCHE-COMTE, BESANÇON, FRANCE

ELLIADD LABORATORY
UNIVERSITE DE FRANCHE-COMTE

CRIT LABORATORY
UNIVERSITE DE FRANCHE-COMTE

Association Internationale des Utilisateurs de NooJ

Organizing Committee

Magali Bigey	ELLIADD, Université de Franche-Comté, France
Annabel Richeton	ELLIADD, Université de Franche-Comté, France
Max Silberztein	ELLIADD, Université de Franche-Comté, France
Izabella Thomas	CRIT, Université de Franche-Comté, France

Table of contents

Table of Contents

Preface **4**

Invited speakers, *Nastia Osidach & Hugues de Mazancourt* 7

Part 1: Lexical and Morphological Resources **8**

Treatment of the aspectual value of Kabyle language verb using NooJ: Distinction between preterit and intensive aorist, *Hamid Annouz* 9
Intensive frozen adverbs of the PECO class in Old French, *Xavier Blanco, Yauheniya Yakubovich* 10
The Western Armenian resources for NooJ, *Anaid Donabedian* 12
Romanian language formalization: building a lexical resource for Nooj, *Maria-Diana Manescu* 13
Formalization of the Italian negation system and sentiment analysis, *Mario Monteleone, Ignazio Mauro Mirto* 15
A NooJ module for Albanian language, *Odile Piton* 17
NooJ module for the Ukrainian language, *Olena Saint-Joanis* 18

Part 2: Syntactic and Semantic Resources **19**

Syntactic analysis of sentences containing Arabic psychological verbs, *Asmaa Amzali, Asmaa Kourtin, Mohammed Mourchid, Abdelaziz Mouloudi and Samir Mbarki* 20
The Recognition and the automatic translation of donation or deprivation verbs, *Hajer Cheikhrouhou* 21
Formalization of clause-subordinate transformations in Quechua, *Maximiliano Duran* 23
Lexicon-grammar tables for modern Arabic frozen expressions, *Asmaa Kourtin, Asmaa Amzali, Mohammed Mourchid, Abdelaziz Mouloudi and Samir Mbarki* 25
The nominal ellipsis in complex head of Spanish. Phenomenon description and computer-based model proposal, *Walter Koza, Hazel Barahona* 26
Meaning extraction from strappare causatives in Italian, *Ignazio Mauro Mirto, Mario Monteleone* 27
Using syntactic grammar to export Nooj morphological annotation: A case study of the morphological annotation of Indonesian texts, *Prihantoro* 29
Transformational analysis of auxiliary verbs, *Max Silberztein* 30

Part 3: Corpus Linguistics and Discourse Analysis **31**

The contribution of NooJ to digital surveillance: The assimilation of the major issues that characterized the U.S. presidential elections, *Nana Ama Ampomah Awuah* 32

The designation and storytelling of the culprits during the French Yellow Vests movement. A study of 2018-2019 books of grievances, *Marion Bendinelli* 33
Uses and potential of mobile devices in francophone sub-Saharan Africa, *Magali Bigey, Ibrahim Maïdakouale* 34
Sensitivity to fake-news: reception analysis with NooJ and ATISHS, *Magali Bigey, Justine Simon* 35
Creation of a legal domain corpus for the Belarusian NooJ module: texts, dictionaries, grammars, *Yuras Hetsevich[1], Yauheniya Zianouka, Valerii Varanovich, Mikita Suprunchuk, Tsimafei Prakapenka, Dmitrii Dzenisiuk* 36
Negation usage in Croatian Parliament, *Kristina Kocijan, Krešimir Šojat* 38
From laws and decrees to a legal ontology, *Ismahane Kourtin, Aziz Mouloudi, Samir Mbarki* 39
Festival volunteers, committed festival-goers or the legacy of cultural practice, *Stéphane Laurent* 41
How to locate traces of subjectivity in diplomatic discourse with Nooj? – The example of the French Ministry of Foreign Affairs, *Annabel Richeton* 42
Terms and oppositions: What unstructured texts tell us, *Giulia Speranza, Maria Pia Di Buono, Johanna Monti* 43
The debates on the advent of the fifth generation of mobile telephony (5G), *Yu Xia* 45

Part 4: Natural Language Processing Software Applications 47

Paraphrasing tool using NooJ Platform, *Amine Alassir ISG, Sondes Dardour, Héla Fehri* 48
Recognition and analysis of complex questions in standard Arabic using NooJ, *Essia Bessaies, Slim Mesfar, Henda ben Ghazela Riadi* 49
Answer validation in question answering system, *Essia Bessaies, Slim Mesfar, Henda ben Ghazela Riadi*. 50
The use of NooJ's functionalities to build an application for Arabic acquisition, *Ilham Blanchete, Mohammed Mourchid* 51
Linguistics, applied research and NLP: using NooJ in a technical-operational context. Case-study, analysis and perspectives, *Nicolas Boffo, Philippe Lambert* 53
Construction of an educational game "VocabNooJ", *Hela Fehri[1], Lazhar Arroum[2], Sameh Ben Aoun[2]* 54
Automatic analysis of finding predicates from a Lexicon-Grammar proposal, *Javiera Jacobsen, Mirian Muñoz, Walter Koza, Francisca Saiz* 55
Arabic spelling error detection and correction using NooJ, *Rafik Kassmi[1], Samir Mbarki[2] and Abdelaziz Mouloudi[3]* 56
Automatic identification and disambiguation of abbreviations in the medical domain, *Walter Koza, Ninoska Godoy, Constanza Suy, Romanet Contreras, Sofía Koza, Fernanda Aguirre, Martín Díaz* 57
Automatic detection and generation of argument structures of the medical domain, *Walter Koza, Constanza Suy Álvarez* 58
Geoparsing with NooJ Italian toponym resolution for environmental crimes, *Raffaele Manna, Annarita Magliacane, Antonio Pascucci, Wanda Punzi Zarino, Vincenzo Simoniello* 60
Integrated NooJ environment for Arabic linguistic disambiguation improvement using MWEs, *Dhekra Najar, Slim Mesfar, Henda Ben Ghezela* 62
Approach to the automatic treatment of gerunds in Spanish and Quechua: A pedagogical application of NooJ, *Andrea Rodrigo[1], Maximiliano Duran[2], María Yanina Nalli[3]* 63
Automatic generation of intonation marks and prosodic segmentation in Belarusian, *Yauheniya Zianouka, Dzmitry Dzenisiuk, David Latyshevich, Yuras Hetsevich* 65

Creation of a legal domain corpus for the Belarusian NooJ module: texts, dictionaries, grammars, *Yuras Hetsevich[1], Yauheniya Zianouka, Valerii Varanovich, Mikita Suprunchuk, Tsimafei Prakapenka, Dmitrii Dzenisiuk*

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

Belarusian State University, Minsk, Belarus
gamrat.vvv@gmail.com, tsimafei.prakapenka@gmail.com

Minsk State Linguistic University, Minsk, Belarus
ms@philology.by

Abstract

The current language situation in the Republic of Belarus is characterized primarily as state bilingualism. At the legislative level, two state languages are established – Belarusian and Russian. But despite the state bilingualism, the vast majority of legislative documents are implemented only in Russian. Thus, of the 26 codes of the Republic of Belarus, i.e. texts which are available on the National Legal Internet Portal pravo.by, 25 are officially adopted in Russian and only one in Belarusian. One of the main factors hindering the practical support of bilingualism in the legal sphere of the Republic of Belarus is the unresolved problem of high-quality and fast linguistic processing of large texts, which testifies to the relevance of high-quality machine translation. To handle the question of translating legislative documents into Belarusian, Speech Synthesis and Recognition Laboratory of UIIP NASB, in cooperation with specialists from Faculty of Social and Cultural Communications of BSU, have translated all codes of the Republic of Belarus into the Belarusian language using automatic services of corpus.by.

The next step is to collect all legislative codes of the Republic of Belarus in the Belarusian language on order to create a unified text corpus. This task is relevant for solving the following tasks. First of all, it is very important to perform primary analysis of legal domain corpus to find out the main linguistic peculiarities of this kind of corpus in comparison with Belarusian literary corpus. Secondly, we will be able to compose different types of dictionaries (Belarusian-Russian, Belarusian-English, Belarusian-English-Russian dictionaries). This question is very actual for Belarusian since there are very few translated Belarusian-foreign and foreign-Belarusian dictionaries of legal terminology – we are aware of 6 dictionaries that are different in their merits and significance for the ordering and development of Belarusian legal terminology. And the last – but not least – task is to develop special morphological and syntactical grammars for further prosodic analysis of legal texts. Automatic syntagmatic delimitation is still not solved for the Belarusian language. That is why developing NooJ grammars will assist the process of creating a system of prosodic marks (including punctuation and intonation marks) and further automatic segmentation of Belarusian texts of the legal domain.

References

National Legal Internet Portal of the Republic of Belarus (2019). *List of legal acts* [Electronic source]. Mode of access: https://pravo.by/document/?guid=3871&p0=H11800130. – Date of access: 18.07.2019.

Computational platform for electronic text and speech processing corpus.by (2019). [Electronic source]. Mode of access: http://corpus.by/. – Date of access: 12.07.2018.

Drahun A. (2019). "Semi-Automatic Proofreading of Belarusian and English texts", A. Drahun, Yu Hetsevich, A. Bakunovich, Dz. Dzenisiuk, J. Shynkevich, In: *International Conference NooJ 2019: Book of Abstracts*. –Hammamet, Tunisia.

Hetsevich Y. (2016). "Semi-automatic Part-of-Speech Annotating for Belarusian Dictionaries Enrichment in NooJ", Y. Hetsevich, V. Varanovich, E. Kachan, I. Reentovich, S. Lysy, In: *Automatic Processing of Natural-Language Electronic Texts with NooJ: 10th International Conference*, České

Budějovice, Czech Republic, June 9-11, Revised Selected Papers, ed. L. Barone, M. Monteleone, M. Silberztein, Springer, 2017, p. 101-111.

Reentovich I. (2016). "The First One-Million Corpus for the Belarusian NooJ Module", I. Reentovich, Y. Hetsevich, V. Voronovich, E. Kachan, H. Kozlovskaya, A. Tretyak, U. Koshchanka, In: *Automatic Processing of Natural-Language Electronic Texts with NooJ: 9th International Conference*, NooJ 2015, Minsk, Belarus, June 11-13, Revised Selected Papers, Springer, ed. T. Okrut, Y. Hetsevich, M. Silberztein, H. Stanislavenka, Springer International Publishing, p. 3-15.