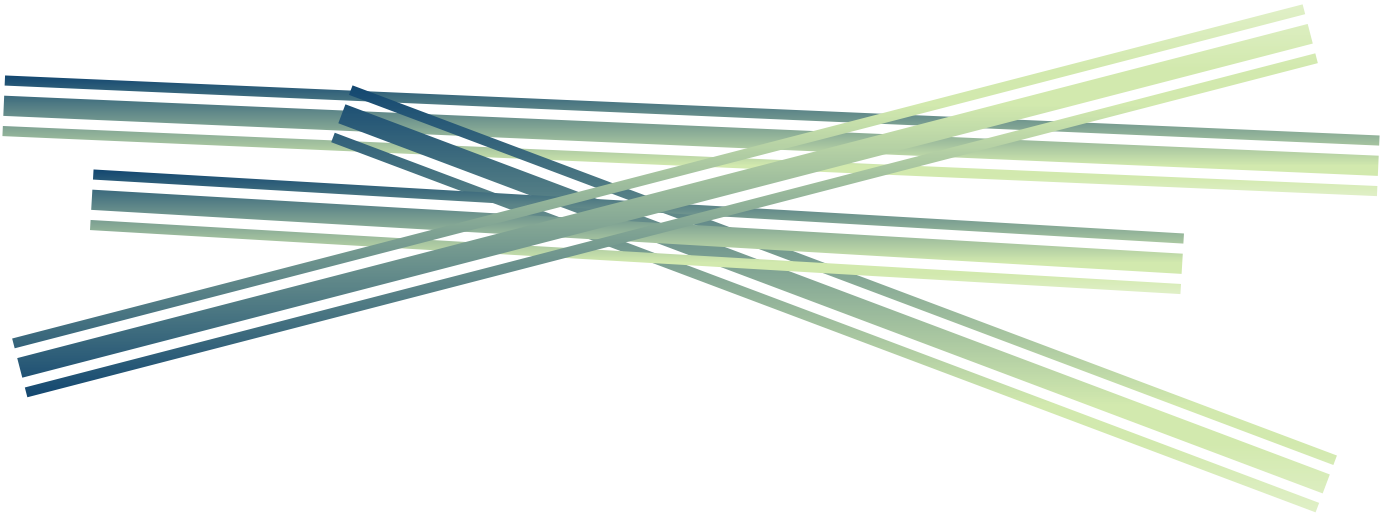

15th INTERNATIONAL CONFERENCE

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

Automatic generation of intonation marks and prosodic segmentation in Belarusian, *Yauheniya Zianouka, Dzmitry Dzenisiuk, David Latyshevich, Yuras Hetsevich*

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

Abstract

Automatically localizing intonation boundaries in a text is one of the main tasks of prosodic processors, considered as a mandatory unit in any speech recognition system. The syntagmatic articulation of the speech flow allocates minimal semantic units and reflects the structural and semantic components of utterances. The automatic selection of syntagmas is complicated by the lack of deep parsing, leading to the search for new approaches to the development of machine algorithms, methods and techniques by defining sequences of linguistic elements associated with certain semantic relationships.

To solve the problem of automatic delimitation in NooJ, we have collected a Belarusian text corpus from the medical domain. It comprises texts of news from medical online portals and consists of nearly 500 texts, 120.000 word forms, more than 8.000 sentences. This work is a continuation of a previous research in which we have analyzed sentence parts separated by punctuation and developed most punctuation marks for such sentences (up to 5 words, but the most frequent being three-word syntagmas). Now, we are planning to expand our study with texts in which the number of syntagmas in a sentence can significantly exceed the number of punctuation marks.

The delimitation of syntagmas is connected with the sentence structure, the word order, the presence of homogeneous members, the nature of word combinations and other linguistic parameters. All the mentioned components should be taken into account and noted in separate syntagmas during developing new syntactic and morphological NooJ grammars.

Hence, we hope to improve the synthetic speech generated by Belarusian text-to-speech systems by using prepared algorithms and grammars from Belarusian medical domain corpus in NooJ for the automatic generation of prosodic transcription of long sentences.

References

Dzenisiuk D. (2019). "Automatic Generation of Right Intonational Marks and Speech for Medical domain in Belarusian", Dz. Dzenisiuk, Yu. Hetsevich, A. Drahun, A. Bakunovich, J. Shynkevich, In: *International Conference NooJ 2019: Book of Abstracts*. Hammamet, Tunisia.

Okrut T. (2015). "Resources for Identification of Cues with Author's Text Insertions in Belarusian and Russian Electronic Texts", T. Okrut, Y. Hetsevich, B. Lobanov, Y. Yakubovich, In: *Formalising Natural Languages with NooJ 2014* / UK; ed. Johanna Monti, Max Silberztein, Mario Monteleone and Maria Pia di Buono. Newcastle: Cambridge Scholars Publishing, p.1 29-139.

Hetsevich Y. (2016). "Grammars for Sentence into Phrase Segmentation: Punctuation Level", Y. Hetsevich, T. Okrut, B. Lobanov, In: *Automatic Processing of Natural-Language Electronic Texts with NooJ: 9th International Conference*, Minsk, Belarus, June 11-13, ed. T. Okrut, Y. Hetsevich, M. Silberztein, H. Stanislavenka. — Springer International Publishing, p. 74-82.

Hetsevich Y. (2015). "Grammars for the Sentence into Phrase Segmentation: Punctuation Level", Y. Hetsevich, T. Okrut, B. Lobanov, In: *International Scientific Conference on the Automatic Processing of Natural-Language Electronic Texts* "NooJ'2015": June 11-13, Minsk, Belarus), ed. B.M. Lobanov, Yu.S.Hetsevich, p. 25