



Filozofická
fakulta
Faculty
of Philosophy

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice



Université de Franche-Comté, Besançon



The NOOJ Association

NOOJ 2016

International Conference

Book of Abstracts

České Budějovice (Czech Republic)

9-11 June 2016

Scientific Committee

Xavier Blanco (Autonomous University of Barcelona, Spain)

Yuras Hetsevich (United Institute of Informatics Problems, Belarus)

Svetla Koeva (University of Sofia, Bulgaria)

Peter Machonis (Florida International University, USA)

Slim Mesfar (University of Manouba, Tunisia)

Mario Monteleone (University of Salerno, Italy)

Johanna Monti (University of Sassari, Italy)

Karel Pala (Masaryk University, Czech republic)

Vladimír Petkevič (Charles University, Prague)

Jan Radimský (University of South Bohemia, Czech republic)

Max Silberztein (Université de Franche-Comté, France)

Marko Tadić (University of Zagreb, Croatia)

François Trouilleux (Université Blaise-Pascal, France)

Organization Committee

Jan Radimský (University of South Bohemia, Czech Republic)

Max Silberztein (Université de Franche-Comté, France)

Zuzana Nevěřilová (Masaryk University, Czech Republic)

Petr Kos (University of South Bohemia, Czech Republic)

Mario Monteleone (Università degli studi di Salerno, Italy)

NOOJ 2016 International Conference – Book of Abstracts

Edited by Jan Radimský

Published by the University of South Bohemia in České Budějovice (2016)

! Semantic Tagging of the Feeling Words in French. How to process the idioms	45
! Semantic tags for Russian language	46
√ Semi-automatic Part-of-Speech Annotating for Belarusian Dictionaries Enrichment in NooJ.....	47
Syntactic Analysis of Relative Clauses in Tamazight (Kabyle)	48
The Arabic translation of the French auxiliary: using the platform NooJ.....	50
! The goals of mentioning people on Twitter : analysing the uses of @.....	51
! Using NooJ and statistical data to analyze reading reception of a best-seller's audience in France.....	52
! Using Text Mining and Natural Language Processing to support Business Decision: towards a NooJ application	53

Semi-automatic Part-of-Speech Annotating for Belarusian Dictionaries Enrichment in NooJ

Yu. Hetsevich¹, V. Varanovich², E. Kachan¹, S. Lysy¹, I. Reentovich¹

¹United Institute of Informatics Problems, Minsk, Belarus

²Belarusian State University, Minsk

e-mails: yury.hetsevich@gmail.com, gamrat.vvv@gmail.com, evgeniakacan@gmail.com,
stanislau.lysy@gmail.com, mwshrewd@gmail.com,

Lexicology is the most flexible level of language, which is continuously changing, updating, improving and at the same time it responds to changes in the reality. Recently, in connection with the development of science, culture, technology and industry the vocabulary of the Belarusian language has changed, manifested in the appearance of new words and phrases denoting new objects, phenomena, concepts. New words reflect the current state of the vocabulary.

Our research is aimed to develop the algorithm for the Belarusian main dictionaries enrichment in NooJ on the basis of the first one-million corpus for the Belarusian NooJ module. From the broad list of possible subject categories the corpus focuses on fiction, historic, medical, scientific, sociological literature and etc., which is considered to be the finest source for searching unknown words of different domains [2]. NooJ as a linguistic development environment gives an opportunity to develop and expand high-quality linguistic algorithms for text processing at the level of lexicology [1].

The task is to work out a mechanism or algorithm for further processing (annotating different categories and paradigms according to flexion classes) of all extracted unknown words (neologisms, dialectal words, subneutral words, colloquial words, jargon words and various foreign words), and adding to the present dictionary on the basis of the Belarusian NooJ module.

So, for this purpose two specific algorithms for automatic generation of word paradigms based on the given grammatical information and automatic stress arrangement for all forms of entire word is planned to develop.

References

1. Max Silberztein. 2014. Nooj Manual. <http://www.nooj4nlp.net>. Date of access : 31.12.2014.
2. Reentovich, I. First One Million Corpora for Belarusian NooJ Module / I. Reentovich, Yu. Hetsevich, V. Voronovich, E. Kachan, H. Kozlovskaya // International Scientific Conference on the Automatic Processing of Natural-Language Electronic Texts "NooJ'2015" : Abstracts (11–13 June, 2015, Minsk, Belarus) / ed. B.M. Lobanov, Yu.S.Hetsevich. — Minsk : UIIP NASB, 2015. — C. 54-55.