

UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

**International Scientific Conference
on the Automatic Processing of Natural-Language
Electronic Texts “NooJ’2015”**

NOOJ 2015

Abstracts

June 11–13, 2015, Minsk, Belarus

Minsk
UIIP NASB
2015

УДК 004.91

International Scientific Conference on the Automatic Processing of Natural-Language Electronic Texts “NooJ’2015” : Abstracts (11–13 June, 2015, Minsk, Belarus). – Minsk : UIIP NASB, 2015. – 80 p.
ISBN 978-985-6744-89-4.

This volume contains the abstracts of the International conference “NooJ 2015”. The research presented covers different aspects of natural language processing using NooJ, including formalizing such levels of linguistic phenomena as syllabification, phonemic and prosodic transcription, multiword units and discontinuous expressions, local and structural syntax; transformational syntax and paraphrase generation, semantic analysis and machine translation, etc.

Abstracts are published in the form presented by authors.

У дадзеным зборніку прадстаўлены тэзісы дакладаў Міжнароднай канферэнцыі “NooJ 2015”. Разглядаюцца розныя аспекты апрацоўкі натуральнай мовы з выкарыстаннем лінгвістычнага асяроддзя распрацоўкі NooJ, улічваючы фармалізаваўне такіх напрамкаў лінгвістычнага аналізу як склададзяленне, фанетычная і прасадычная транскрыпцыі, устойлівыя выразы і дыскрэтныя слоўныя канструкцыі, лакальны і структурны сінтаксісы, трансфармацыйны сінтаксіс і перафразаванне, семантычны аналіз і машынны пераклад і г. д.

Тэзісы друкуюцца ў выглядзе, пададзеным аўтарамі.

Scientific Editors:

DSc in Engineering B.M. Lobanov,
PhD in Engineering Yu.S. Hetsevich

ISBN 978-985-6744-89-4

© United Institute of Informatics
Problems of the National Academy
of Sciences of Belarus, 2015

ADDITION OF PHONETIC TRANSCRIPTIONS TO BELARUSIAN MODULE OF NOOJ

S. Lysy, A. Hiuntar, Yu. Hetsevich

United Institute of Informatics Problems of the NAS of Belarus, Minsk

e-mail: stanislau.lysy@gmail.com

To learn, study or perform a text processing for one or another language full and thorough description of the language is required. The authors of this article have noted that while much has been done in the development of different areas of language processing with NooJ, yet little attention has been paid to issues related to phonetic language features [1].

This article will describe a way to represent phonetic level of language for Belarusian module of NooJ. This will be done in two ways: via creating a dictionary including phonetic transcriptions and via developing morphological NooJ grammars for creating a phonetic transcription for orthographic words.

For the first part of this aim, a software tool which allows to quickly transform both single words and whole texts into phonetic transcription will be used [2, 3]. This software tool can generate three kinds of transcription: Cyrillic, simple Latin and International Phonetic Alphabet [4]. Apart from that there will be developed and implemented an algorithm, which adds phonetic transcription in three forms listed above for every word in the Belarusian dictionary.

For the second part of this aim, a morphological NooJ grammars will be developed. In Belarusian, one letter may be represented by different allophones depending on their surrounding letters or position in the word. The most common sound changes in Belarusian are assimilation, elision and positional fortition. For example, in the word *дуб* “dub – eng. oak”, the last letter *Б* changes into the sound [p] as a result of end-word fortition. These sound changes will present in the grammar as following: all the graphemes, which are surrounded by other particular graphemes will be given as an output the allophone match, for instance, grapheme *Б* from the example above will be marked by *P* as a corresponding allophone.

The main purpose of this paper is description structure of the Belarusian language using NooJ, which will help in introducing and learning the norms of the literary pronunciation of this language. Moreover the results can be useful in dealing with other educational and linguistic problems.

References

1. Silberztein, M. Nooj Manual / M. Silberztein [Electronic resource]. – 2014. – Mode of access : <http://www.nooj4nlp.net>. – Date of access : 22.12.2014.

2. Transcription Generator [Electronic resource]. – 2014. – Mode of access : <http://corpus.by/transcriptionGenerator>. – Date of access : 22.12.2014.

3. The system of generation of phonetic transcriptions for input electronic texts in belarusian / Yu. Hetsevich [et al.] // Pattern Recognition and Information Processing : Proceedings of The 12th Intern. Conf., 28–30 May, Minsk, Belarus. – Minsk : UIIP NASB, 2014. – C. 81–85.

4. International Phonetic Association [Electronic resource]. – 2005. – Mode of access : <http://www.internationalphoneticassociation.org>. – Date of access : 22.12.2014.

CONTENTS

PREFACE	5
Ben Ali H., Rhazi A., Aouini M. Translating Arabic Active Sentences into English Passive Sentences using NooJ Platform 7	
Benet V. Semantic Tags for NooJ Russian Dictionary.....	9
Blanco X. A Hierarchy of Semantic Labels for Spanish Dictionaries.....	10
Chernyshevich M., Stankevitch V. A Hybrid Approach to Extracting and Encoding Disorder Mentions from Clinical Notes.....	12
Collec Clerc V. Mixed Prolog and NooJ Approach in Japanese Benefactive Constructions.....	14
Buono di M.P. Semi-Automatic Indexing and Parsing Information on the Web with NooJ.....	16
Duran M. The Annotation of Compound Suffixation Structure of Quechua Verbs.....	18
Dzenisiuk D., Hetsevich Yu. Processing of Publication References in Belarusian and Russian Electronic Texts.....	20
Ghezaiel N., Haddar K. Study and Resolution of Arabic Lexical Ambiguity through the Transduction on Text Automaton.....	21
Hetsevich Yu., Borodina J. Using NooJ for the Processing of Satellite Data.....	23
Hetsevich Yu., Okrut T., Lobanov B. Grammars for the Sentence into Phrase Segmentation: Punctuation Level.....	25
Hiuntar A., Zahariev V. Grammars for Making Written Orthographic Words from Transcribed Spoken Language.....	26
Kaigorodova L., Hetsevich Yu., Nikalaenka K., Prakapovich R., Gerasuto S., Sychou U. Language Modelling for Robots-Human Interaction.....	28

Kirova M. Translating Spacial and Temporal Deixis in Near Languages: A Comparative Classification Approach with NooJ.....	30
Kocijan K., Librenjak S. Recognizing Verb-Based Croatian Idiomatic MWUs.....	31
Koshchanka U., Hetsevich Yu., Varanovich V., Tretyak A. Comparison of Lexical and Grammatical Base of Belarusian N-Korpus with Dictionary Properties' Definition File of Belarusian NooJ Module.....	33
Le Pesant D. Semantic Tagging of the Sentiment Words with NooJ.....	34
Loskutova A. Creation of Geographical Names Dictionary of Alaska Toponyms.....	35
Lysy S., Hiuntar A., Hetsevich Yu. Addition of Phonetic Transcriptions to Belarusian Module of NooJ.....	36
Maisto A., Guarasci R. Morpheme-Based Recognition and Translation of Medical Terms.....	38
Mesfar S., Najar D. How to Automatically Enrich Linguistic Resources Using NooJ: Application on Arabic Module.....	40
Monteleone M. Local Grammars and Formal Semantics: Past Participles Vs. Adjectives in Italian.....	41
Mota C., Carvalho P., Raposo F., Barreiro A. Paraphrasing Human Intransitive Adjective Constructions in Port4NooJ.....	43
Najar D., Mesfar S. A Large Terminological Dictionary of Arabic Compound Words.....	46
Okrut T., Lobanov B., Yakubovich Y. Context-Sensitive Homograph Disambiguation with NooJ in Belarusian and Russian Electronic Texts.....	48
Patsiomkin A., Hetsevich Yu. Semantic Analysis for Locating Expressive Means and Stylistic Devices in Authentic English Texts, Ranging and Classification.....	49
Pejar T., Kocijan K., Bekavac B. Normalization of Tweets in Croatian Language Using NooJ.....	51

Pelosi S. Morphological Relations for the Automatic Expansion of Italian Sentiment Lexicons.....	52
Reentovich I., Hetsevich Yu., Varanovich V., Kachan E., Kozlovskaya H. First One Million Corpora for Belarusian NooJ Module.....	54
Rodrigo A.F. A Proposal for the Treatment of Clitics in Rioplatense Spanish Verbs Using NooJ.....	56
Rusetski K., Ilyushchenia D., Nikalaenka K., Lysy S. Towards Building Ostis Technology-Based Semantic NLP Applications Using NooJ.....	58
Sazhok M., Robeiko V., Fedoryn D., Selyukh R., Yukhymenko O. Ukrainian Data and Knowledge Base and its Adaptation to NooJ.....	60
Seideh M.A.F., Fehri H., Haddar K., Ben Hamadou A. Named Entity Recognition from Arabic-French Herbalism Parallel Corpora.....	62
Silberztein M. Transformational Analysis of Transitive Sentences.....	64
Sovpel I. From Linguistic to Knowledge Processor.....	65
Veka A., Yakubovich Y. Automatic Translation from Belarusian into Spanish Based on Using NooJ's Linguistic Resources.....	66
Yamouni F. A French-Tamazight MT System for Computer Science.....	68