

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

CONTEXT-SENSITIVE HOMOGRAPH DISAMBIGUATION WITH NOOJ IN BELARUSIAN AND RUSSIAN ELECTRONIC TEXTS

T. Okrut¹, B. Lobanov¹, Y. Yakubovich²

¹United Institute of Informatics Problems of the NAS of Belarus, Minsk;

²Universitat Autònoma de Barcelona, Bellaterra, Spain

e-mail: tatberrie@gmail.com

When we read, we may encounter words having different phonological representations associated with singular orthographic representation. In speech synthesis, disambiguation of such words, or homograph disambiguation, serves an obstacle to overcome at the stage of text preprocessing. There are several major types of homographs we deal with in Belarusian and Russian: different lexemes of the same part of speech, different forms of the same lexeme and different lexemes of different parts of speech. Moreover, these homographic groups may be divided into the subgroups on the base of grammatical similarity of homographic word pairs [1]. Elements in such pairs differ at least in one grammatical feature, which influences the stress position in a word. For example, in the following homographic pairs of the group “different forms of the same lexeme” elements differ in number:

ГО́ДА (singular, “year”) – ГОДÁ (plural, “years”),
О́ЗЕРА (singular, “lake”) – ОЗЁРА (plural, “lakes”).

Such similarity allows developing of one context-sensitive disambiguation algorithm for a number of homographic pairs at once. The authors have already developed a Russian syntactic NooJ grammar for disambiguation of 58 homographs referring to the homographic subgroup “Singular nouns of masculine or neuter gender in genitive case – Plural nouns in accusative or nominative case”. Therefore, the goal of this research is to improve the grammar mentioned above and to develop a similar Belarusian disambiguation grammar using a context-sensitive approach.

References

1. Выращэнне амаграфіі з дапамогай NOOJ для больш чым 50 амографіаў рускай мовы / Т.І. Окрут [і інш.] // Контрастивные исследования и прикладная лингвистика: материалы Междунар. науч. конф., Минск, 29–30 окт. 2014 г. : в 2 ч. / М-во образования Респ. Беларусь, Минский гос. лингв. ун-т; редкол. : А.В. Зубов [и др.]. – Минск, 2014. – Ч. II. – С. 83–87.

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., Prakupovich 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