MARK KIT$^1$, A & VIOLETTA KOSESKA-TOSZEWA$^{1,2}$

$^1$Language Interface Inc., New York
$^2$Institute of Slavic Studies, Polish Academy of Science, Warsaw

$^A$mark.kit@langint.com ; $^B$amaz@inetia.pl

DIALOG BETWEEN A LEXICOGRAPHER AND A TRANSLATOR

Abstract

The discussion between the authors of the paper concerns the most pressing issues encountered in natural language semantics, as well as in corpus linguistics and computational linguistics. A broad range of knowledge, allowing linguists and information scientists to work together, is required in these areas. The paper describes some primary problems of human and machine translation caused by gaps between different fields of knowledge. The authors suggest that interdisciplinary approach is required when it comes to contrastive studies in linguistics.

Keywords: reversibility of dictionaries, lexical structures, multilingual dictionaries, perfective aspect, semantic tags, parallel corpora, corpus linguistics, computational linguistics, contrastive linguistics.
MACIEJ PIASECKI
G4.19 Research Group, Institute of Informatics, Wroclaw University of Technology
maciej.piasecki@pwr.wroc.pl ; http://nlp.pwr.wroc.pl

SELF-ORGANISING LOGIC OF STRUCTURES AS A BASIS FOR A DEPENDENCY-BASED DYNAMIC SEMANTICS MODEL

Abstract

We present Self-organising Logic of Structures (SLS), a semantic representation language of high expressive power, which was designed for a fully compositional representation of discourse anaphora following the Dynamic Semantics paradigm. The application of SLS to the description of possible meanings of Polish multiple quantifier sentences is discussed. Special attention is paid to the phenomena of: cardinality dependency/independency of Noun Phrase quantifiers and variety of quantification. Semantic representation based on several formal operators is proposed. They can be combined in many different ways, if one takes a purely theoretical perspective. However, in the paper we show that this huge number is practically reduced in the language use and is governed by several constraints motivated by the analysis of Polish language data. The Hypothesis of Local Range of Cardinality Dependency is formulated as an alternative to representations based on quantifier rising technique. SLS provides a multi-layered language description of inter-linked representation of several semantic phenomena: quantification, reference, presupposition and anaphora.

Keywords: Dynamic Semantics, compositionality, reference, quantification, anaphora, Polish, variableless semantic representation.
SVETLA KOEVA¹ & DIANA BLAGOEV¹
¹Institute for Bulgarian Language. Bulgarian Academy of Sciences

THE DICTIONARY WRITING SYSTEM LEXIT
AND ITS APPLICATION IN BILINGUAL LEXICOGRAPHY¹

Abstract

The paper presents LexIt — a web-based Dictionary Writing System designed for the creation and enlargement of dictionaries. A wide range of dictionaries — monolingual, specialised, multilingual, thesauri, can be configured with the system, as it enables users to model, copy, and save templates describing the structure of any dictionary. The paper describes a model for a bilingual dictionary containing phonetic, grammatical, stylistic, semantic, and collocational information and its configuration within LexIt. A bilingual dictionary is chosen for a demonstration, as its entries have a complex structure that includes various elements reflecting the correspondences between source and target languages.

Keywords: Dictionary Writing System, bilingual lexicography.

¹ LexIt is developed with the financial support of the Fund for scientific research at the Bulgarian Ministry of Education, Youth and Science, contract DTP 02-53/2009.
ABOUT CERTAIN SEMANTIC ANNOTATION
IN PARALLEL CORPORA

Abstract

The semantic notation analyzed in this work is contained in the second stream of semantic theories presented here — in the direct approach semantics. We used this stream in our work on the Bulgarian-Polish Contrastive Grammar. Our semantic notation distinguishes quantificational meanings of names and predicates, and indicates aspectual and temporal meanings of verbs. It relies on logical scope-based quantification and on the contemporary theory of processes, known as “Petri nets”. Thanks to it, we can distinguish precisely between a language form and its contents, e.g. a perfective verb form has two meanings: an event or a sequence of events and states, finally ended with an event. An imperfective verb form also has two meanings: a state or a sequence of states and events, finally ended with a state. In turn, names are quantified universally or existentially when they are “undefined”, and uniquely (using the iota operator) when they are “defined”. A fact worth emphasizing is the possibility of quantifying not only names, but also the predicate, and then quantification concerns time and aspect. This is a novum in elaborating sentence-level semantics in parallel corpora. For this reason, our semantic notation is manual. We are hoping that it will raise the interest of computer scientists working on automatic methods for processing the given natural languages. Semantic annotation defined like in this work will facilitate contrastive studies of natural languages, and this in turn will verify the results of those studies, and will certainly facilitate human and machine translations.

Keywords: direct approach to semantics, semantic annotation, perfective aspect, imperfective aspect, event, state, Petri nets, parallel corpora, contrastive linguistics.
Abstract

In the article the author describes the experimental corpus of the Lithuanian local dialect of Puńsk in Poland (ECorp-of-Punsk). It is the first corpus of this type for the Lithuanian local dialect. The corpus consists of three subcorpora. The first one (referred to as fundamental) contains utterances given by Lithuanians in the local dialect, the second one — utterances given by Lithuanians in Polish, the third one — aligned Polish-dialectal texts. The texts recorded in the years 1986–2012 have been included in the Ecorp-of-Punsk resources.

Keywords: corpora, annotation, Lithuanian local dialect of Puńsk in Poland, experimental dialectal corpus.
DANUTA ROSZKO\textsuperscript{1, A} \& ROMAN ROSZKO\textsuperscript{1, B}

\textsuperscript{1}Institute of Slavic Studies, Polish Academy of Science, Warsaw
\textsuperscript{A}danuta.roszko@ispan.waw.pl ; \textsuperscript{B}roman.roszko@ispan.waw.pl

EXPERIMENTAL POLISH-LITHUANIAN CORPUS
WITH THE SEMANTIC ANNOTATION ELEMENTS

Abstract

In the article the authors present the experimental Polish-Lithuanian corpus (ECorpPL-LT) formed for the idea of Polish-Lithuanian theoretical contrastive studies, a Polish-Lithuanian electronic dictionary, and as help for a sworn translator. The semantic annotation being brought into ECorpPL-LT is extremely useful in Polish-Lithuanian contrastive studies, and also proves helpful in translation work.

Keywords: corpora, parallel and comparable corpora, annotation, Polish, Lithuanian.
FROM THE PROBLEMS OF DICTIONARIES
AND MULTI-LINGUAL CORPORA

Abstract

The article describes the work on a number of dictionaries being developed by the Corpus Linguistics and Semantics Group of the Institute of Slavic PAS. They include “Contemporary Bulgarian-Polish Dictionary”, “Bulgarian-Polish Online Dictionary” and “Russian-Bulgarian-Polish Dictionary”. The dictionaries differ in the numbers of entries, as well as in the different degrees of their connection with parallel corpora being elaborated under the “Clarin” project. All the discussed dictionaries are similar with respect to their use of traditional, syntactic classifiers and of semantic classifiers, introduced for the first time in the existing lexicographical practice. Thanks to the “Polish-Bulgarian-Russian Corpus”, the Group has managed to verify the results of contrasting Polish and Bulgarian in the light of scope-based logical quantification. Thanks to the Russian material added to the trilingual corpus, the researchers have managed to confirm the fact that from the viewpoint of “incomplete quantification” Russian and Polish (synthetic languages) behave similarly, and are opposed to the analytic Bulgarian.

Keywords: Bilingual dictionary, trilingual dictionary, online dictionary, traditional classifier, syntactic classifier, semantic classifier, bilingual parallel corpora, trilingual parallel corpora, linguistic quantification, “incomplete quantification”.

VIOLETTA KOSESKA-TOSZEWA1.A, JOANNA SATOŁA-STAŚKOWIAK1
& WOJCIECH SOSNOWSKI1

1Institute of Slavic Studies, Polish Academy of Science, Warsaw
Aamaz@inetia.pl
PAWEŁ KĘDZIA¹, MACIEJ PIASECKI¹, EWA RUDNICKA¹ & KONRAD PRZYBYCIEŃ¹
¹G4.19 Research Group, Institute of Informatics, Wrocław University of Technology
http://nlp.pwr.wroc.pl

AUTOMATIC PROMPT SYSTEM IN THE PROCESS OF MAPPING PLWORDNET ON PRINCETON WORDNET

Abstract

The paper offers a critical evaluation of the power and usefulness of an automatic prompt system based on the extended Relaxation Labelling algorithm in the process of (manual) mapping plWordNet on Princeton WordNet. To this end the results of manual mapping – that is inter-lingual relations between plWN and PWN synsets – are juxtaposed with the automatic prompts that were generated for the source language synsets to be mapped. We check the number and type of inter-lingual relations introduced on the basis of automatic prompts and the distance of the respective prompt synsets from the actual target language synsets.

Keywords: semi-automatic prompt system, mapping procedure, plWordnet, WordNet, Polish, interlingual relations.
CONTEMPORARY CONTRASTIVE STUDIES OF POLISH, BULGARIAN AND RUSSIAN NEOLOGISMS VERSUS LANGUAGE CORPORA

Abstract

In the field of Slavonic linguistics contrastive studies of neologisms occupy little place, the newest words are insufficiently described and classified. The aim of this article is to draw attention to the need for contrastive description of the newest lexis and checking exclusively one of many possibilities of obtaining Polish, Bulgarian and Russian neologisms. Language corpora, as this possibility is in question, are not the only source from which the author obtains her research material, yet a growing interest in corpora has inspired her to also use this method. The author wants to show the reader to what degree language corpora can help in building the thesaurus of Polish, Bulgarian and Russian neologisms. Making an attempt to confront a collection of neologisms of contemporary Polish, Bulgarian and Russian language, the author points out the need to standardize the description (identical for each of the analysed languages\textsuperscript{a}), which she intends to propose in another publications on neologisms in Polish, Bulgarian and Russian language. The application of contrastive method to three different but related languages from the Slavonic group will help, in her opinion, to discover more mechanisms of new words coming into existence and examine the newest derivative processes and their productivity.

Keywords: neologisms, neosemantisms, internationalisms, language corpora, parallel corpora, contrastive studies, Polish language, Bulgarian language, Russian language.

\textsuperscript{a}More on the same subject in other articles by the author — cf. e.g. Satol-Staśkowiak (2012, 2013).
LUDMILA DIMITROVA\textsuperscript{1,\textit{A}} & VIOLETTA KOSESKA-TOSZEW\textsuperscript{2,\textit{B}}

\textsuperscript{1}Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia
\textsuperscript{2}Institute of Slavic Studies, Polish Academy of Science, Warsaw
\textsuperscript{A}ludmila@cc.bas.bg ; \textsuperscript{B}amaz@inetia.pl

BULGARIAN-POLISH LANGUAGE RESOURCES
(CURRENT STATE AND FUTURE DEVELOPMENT)

Abstract

The paper briefly reviews the first Bulgarian-Polish digital bilingual resources: corpora and dictionaries, which are currently developed under bilateral collaboration between IMI-BAS and ISS-PAS: joint research project “Semantics and contrastive linguistics with a focus on a bilingual electronic dictionary”, coordinated by L. Dimitrova (IMI-BAS) and V. Koseska (ISS-PAS).

Keywords: digital bilingual resources, parallel corpus, aligned corpus, lexical data base, paper and digital dictionary, online dictionary, digital entry classifiers.
GRIGORIJ CHETVERIKOV\textsuperscript{1,A} \& IRYNA VECHIRSKA\textsuperscript{1,B}

\textsuperscript{1}Faculty of Computer Sciences, Kharkiv National University of Radio electronics, Kharkiv, Ukraine
\textsuperscript{A}chetvergg@gmail.com ; \textsuperscript{B}ira_se@list.ru

FORMALIZING THE PROCESS
OF CONSTRUCTING CHAINS OF LEXICAL UNITS

Abstract

The paper investigates mathematical aspects of describing the construction of chains of lexical units on the basis of finite-predicate algebra. Analyzing the construction peculiarities is carried out and application of the method of finding the power of linear logical transformation for removing characteristic words of a dictionary entry is given. Analysis and perspectives of the results of the study are provided.

Keywords: lexical units, chain of lexical units, semantic classification, linear logical transformation, kernel of linear logical transformation.
WEB-APPLICATION FOR THE PRESENTATION OF BILINGUAL CORPORA (FOCUSBING ON BULGARIAN AS ONE OF THE TWO PAIRED LANGUAGES)

Abstract

This paper briefly presents a web-application for the presentation of bilingual aligned corpora focusing on Bulgarian as one of the two paired languages. The focus is given to the description of the software tools and user interface. The software is developed in IMI-BAS and will be hosted on a server there. Some examples of the usage of the web-application for the presentation of a Bulgarian-Polish aligned corpus are included.

Keywords: parallel corpus, aligned corpus, concordance, linguistic annotation, lemmatization, POS-tagging, web-interface, web-application.
Abstract

We present outcomes of an undertaking on building a parallel Treebank for the Georgian and the Ukrainian languages, which is a “side product” of the GRUG multilingual Treebank project. The GRUG acronym stands for the German-Russian-Ukrainian-Georgian Treebank intended for contrastive studies and translation memory systems. The monolingual Ukrainian and Georgian parallel sentences were syntactically annotated manually using the Synpathy tool. Tagsets for both languages follow an adapted version of the German TIGER guidelines with necessary changes relevant for the Georgian and the Ukrainian grammar formal description. An output of the monolingual syntactic annotation is in the TIGER-XML format. Alignment of monolingual resources into a bilingual Georgian-Ukrainian Treebank was done by the Stockholm TreeAligner software.

Keywords: Treebanks, Annotation, Syntactic Parse, Tree-to-Tree Alignment, Translation Equivalents.
DOMINIKA TOPA-BRYNIARSKA
Institut des Langues Romanes et de Traduction de l’Université de Silésie, Sosnowiec
dominika.topa@us.edu.pl

LES STRUCTURES ONTOLOGIQUES DE L’ÉDITORIAL
SOCIO-POLITIQUE : UNE APPROCHE SOCIO-COGNITIVE

Abstract

Ontologic structures in a socio-political editorial — a socio-cognitive approach. The aim of the present study is to analyze the conceptualization of ontological structures in a special type of journalistic writing represented by a socio-political editorial. According to the socio-cognitive approach (Goffman, 1991; Langacker 1986), the author discusses the problem of discursive mechanisms related to the persuasive function of language and classification. Therefore, in basis of the gathered linguistic corpus, which constitutes of forty French editorials, treating about university and public service pension reforms, culled from the web pages of different journals and magazines (2007–2008), the author intends to describe the distribution of two semantic roles such as agent and patient (by means of the synthetic scheme presented in this paper) in the context of the persuasiveness of media communication. The way of constructing the semantic roles of participants both influences and modifies the readers’ opinions and their perception of the world.

Keywords: ontological structures, persuasive function of language, editorial, gatekeeping, cognitive action chain model, Goffman’s frame analysis, semantic role of agent, semantic role of patient.
WOJCIECH SOSNOWSKI
Institute of Slavic Studies, Polish Academy of Science, Warsaw
wpsosnow@uw.edu.pl

FORMS OF ADDRESS AND THEIR MEANING IN CONTRAST IN POLISH AND RUSSIAN LANGUAGES

Abstract

Many studies in contemporary linguistics focus on investigating politeness and rudeness in language. This paper, however, has not been intended as a contrastive study of the phenomena in question. Language politeness and rudeness are conveyed by means of expressions of politeness and rudeness which are perceived as entrenched and recurring in specific situations. These expressions convey the expected meaning of politeness and rudeness accepted in the model of social behaviour. If one uses the explicative method such expressions could be reduced to the following formula ‘I inform you that I follow a verbal conduct defined as polite’. Owing to the emergence of parallel corpora of particular languages, it is nowadays easier to collect data for research on forms of address as well as on expressions of politeness in the first half of the 21st century. Investigating the meaning of forms of address, which are part of linguistic repertoire used to express politeness and rudeness should be regarded as an interesting area of research. It is the consequence of the increasing importance of intercultural communication, expansion of international cooperation, and formation of new standards of interpersonal communication aimed at achieving mutual understanding without resorting to violence. It is worth mentioning that currently there are no bilingual dictionaries which would include practical rules for using forms of address. Moreover, dictionaries (especially bilingual ones) also do not list classifiers of politeness, which becomes a shortcoming as regards the purposes of translation and teaching foreign languages. The aforementioned problems apply to print as well as computer dictionaries. A reliable list of forms of address and their meaning may become helpful in intercultural communication. It would be also important to create a Contemporary Dictionary of Expressions of Politeness and Rudeness in a paper as well as a computer version. Keywords: forms of address, language politeness, language rudeness, act of speech, bilingual dictionary, parallel corpora, classifiers of politeness, print dictionary, computer dictionary.
AGNIESZKA PLUWAK
Institute of Slavic Studies, Polish Academy of Science, Warsaw
agnieszka.pluwak@gmail.com

AN ANALYSIS OF THE WAR AGAINST THE DISEASE-METAPHOR IN SELECTED 2011 PRESIDENTIAL SPEECHES OF FELIPE CALDERON WITH THE APPLICATION OF THE NEURAL THEORY OF METAPHOR

Abstract

The following paper describes the background of the neural theory of language by Jerome Feldman and provides a set of related research tools by George Lakoff. It also offers an example of its application to metaphor analysis based on 2011–2012 speeches by the Mexican President, Felipe Calderon. A complex WAR AGAINST THE DISEASE-metaphor is decomposed and explained with respect to cognitive patterns involved. In its envoy, the article goes back to a famous linguistic question, whether one can judge from linguistic research onto the mental and neural processes involved in language creation and processing. It is also an attempt to indicate possible future directions in the interdisciplinary field of metaphor research.

Keywords: metaphor, neural theory of language, frame, neural circuit, domain transitivity.
LES ENSEMBLES RHÉMATIQUES COMME ÉLÉMENTS CONSTITUTIFS DU NIVEAU SUPRAPHRASTIQUE DE LA STRUCTURE INFORMATIONNELLE DU DISCOURS

Abstract

Rhematic sets as components of global information structures in discourse. The aim of this paper is to analyze the global organisation of rhematic part of information structures in discourse. The information structures are defined as hierarchically organized thematic-rhematic structures and, in their rhematic part, the author distinguishes two levels represented by: rhematic sets and rhemes of sentences. The author describes the configurations of rhematic sets in discourse and their internal structures.

Keywords: global information structure, rhematic set, rheme, levels of information structure.
ALLA LUCHYK
National University of Kyiv-Mohyla Academy, Kyiv, Ukraine

INTERPRETATION OF UKRAINIAN AND POLISH AVERBIAL WORD EQUIVALENTS FORM AND MEANING INTERACTION IN NATIONAL EXPLANATORY LEXICOGRAPHY

Abstract

The article proves the necessity and possibility of compiling dictionaries with intermediate existence status glossary units, to which the word equivalents belong. In order to form the Ukrainian-Polish dictionary glossary of this type the form and meaning analysis of Ukrainian and Polish word equivalents is done, the common and distinctive features of these language system elements are described, the compiling principles of such dictionary are clarified.

Keywords: word equivalents, adverbial equivalent, dictionary, glossary, dictionary entry, model, form of expression, semantics.
AFFIRMATION MODALITY IN BULGARIAN, MACEDONIAN AND SERBIAN

Abstract

In the case of affirmation modality the speakers transform their utterances by stressing or attributing a positive value as an additional component added to the semantic structure of a proposition. This type of affirmative polarization is triggered in opposition to negation or hypothetically negative contexts. The goal of the present paper is twofold: on the one hand to compare and contrast affirmative periphrastic constructions in Bulgarian, Macedonian and Serbian and, on the other hand, to ascertain what these constructions reveal regarding the organization of grammatical categories in general and the status of affirmation modality as a coherent and homogenous category with a linguistic validity.

Keywords: affirmation, modality, subjectivity, auxiliation, evidentiality.