Hélène Wlodarczyk
Université Paris-Sorbonne
CELTA — Centre de Linguistique Théorique et Appliquée

FROM ONTOLOGICAL ATTRIBUTES
TO SEMANTIC FEATURE STRUCTURES
— EXPERIMENTAL RESEARCH ON ASPECT IN POLISH

Abstract. Our aim is the formalisation of semantic concepts using experimental Computer-aided Acquisition of Semantic Knowledge (CASK) method. This method is being used at CELTA with the SEMANA (SE-Mantic ANAlyser) software which was especially designed for this purpose. The case study we propose is the experimental research on Aspect in Polish: we show how although rich nevertheless intuitive notions resulting from years of traditional research can be formalised with computational methods of data analysis.

We use the notation close to that of Semantic Feature Structures as a representation language for describing the category of aspect in Polish taking into account not only the grammatical core of this category but also the whole semantic field (called aspectuality1 by A.V. Bondarko) regardless of its pertaining to various levels of expression (morphological, syntactical or lexical).

To model the Category of Aspect we treat ontological types of situations as conditions for two relevant aspect parameters we call Analysis of the situation (or Internal Aspect) and Control of the situation (or External Aspect). In order to cope both with the lexical diversity of aspectual morphemes (prefixes) and the grammatical (dichotomous and obligatory) character of aspect in Slavic languages, we have proposed to define the perfective aspect as a hypercategory.

The database of Polish Aspect was analysed using KDD statistical tools (Sauvet (2008)): thus, we obtained the first preliminary experience-based semantic definitions of the perfective and imperfective values of the category of aspect in a Slavic language.

References


1 In Russian *apektual’nost’,* (Agresll 1908).
Abstract. The analysis of the language confrontation issues presented in the paper shows the imperfection of the results of research where two or more languages are compared based on a formal inventory, i.e. the so-called morpho-syntactical features and values. The use of interlanguage as a language of consistent and simple notions helps overcome the formal barrier, and ensures that the individual confronted languages are always referred to the same meaning plane, known traditionally as tertium comparationis. The results of research on natural languages obtained based on a confrontation with a semantic interlanguage are comparable and have an equal status. The approach presented in the Part 2 motivates syntactical phenomena by semantic features of predicative units, as well as by the specific structure of the argument places opened by those units. The objects contained within those places are carriers of states in the sense of net theory, but describing that sphere solely by its relations to states and events does not exhaust the whole problem area. The paper presents an outline of an apparatus of analysis starting with the semantic plane, in which the elements of the interlanguage are classes of semantic predicates and the set of predicate-argument positions defined by positions at possibly simple predicates. The paper also shows problems connected with the condensation phenomena. As an effect of such phenomena, some elements of the semantic structure are only realized superficially, and functions of argument phrases might be sometimes neutralized.

Keywords: Semantic, Contrastive Studies, Interlanguage, Predicative unit, Experiencer, Agentive.
FORMAL DESCRIPTION OF TEMPORALITY
(PETRI NET APPROACH)

Abstract. In the paper two methods of representing temporal dependencies expressed in natural languages are given. The first one has been introduced by H. Reichenbach in 1948 is based on a linear representation of time, with events represented by points on the time scale. The second one is based on a net representation of states, events, and their succession introduced by C. A. Petri in 1962. The main difference between these two approaches consists in accepting by nets (i) partial ordering of events and states rather than their linear ordering, hence accepting their mutual independence, and (ii) a possibility of representing coexisting as well as mutually excluding states, hence accepting different histories in one model. Reichenbach’s representation can be viewed as a particular case of the Petri net one. Both representations use graphical means for modeling temporal phenomena.

Keywords: States, events, ordering, temporality, histories, Petri nets.
VIOLETTA KOSESKA¹, ANTONI MAZURKIEWICZ²
¹Institute of Slavic Studies, Polish Academy of Sciences, Poland
²Institute of Computer Science, Polish Academy of Sciences, Poland

NET-BASED DESCRIPTION OF MODALITY IN NATURAL LANGUAGE (ON THE EXAMPLE OF CONDITIONAL MODALITY)

Abstract. The intention of the present paper is to show how the Petri nets formalism can be applied for explaining not only temporal but also modal properties of sentences in natural languages. A special attention has been paid for distinguishing courses of actions with forking (that creates different, but coexistent courses) from branching (that creates different and mutually exclusive courses). It is argued that conditional sentences cannot be represented properly by means of logical implication; instead, for this representation the net description is proposed. Examples serve to show how Petri nets can be viewed as a universal tool (an intermediate language) for analyzing and comparing different natural languages.

Keywords: states, situations, events and histories, branchings and forks, implication and conditional, Petri nets.
KATARZYNA DROŻDŻ-ŁUSZCZYK, ZOFIA ZARON
Institute of Applied Polish Language Studies, Warsaw University, Poland

SEMANTIC INTERRELATIONS BETWEEN THE WORDS
MISTRZ AND UCZEŃ

Abstract. The paper presents the results of research into relation between the lexemes mistrz (master) and uczeń (apprentice or student). Some phrases (mój mistrz, mój uczeń) and syntactic features suggest a kind of convertive relation. Both of the lexemes are — for example — a three-actant: ktoś, czyj, w jakiej dziedzinie wiedzy (/sztuki) (someone, someone’s, in a field of science (/art). This feature is not sufficient to make a convertive relation. In Polish Language the mistrz-uczeń relation is a unilateral one. The relation depends on the apprentice who selects a master. The master confirms this only if uczeń (apprentice) is taught by the master and if the master is regarded by the apprentice as a skilled, knowledgeable person.

Keywords: master, apprentice, student, relationship, meaning.
DARJA FIŠER¹
TOMAŽ ERJAVEC²
¹Department of Translation, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia
²Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia

SEMANTIC CONCORDANCES FOR SLOVENE

Abstract. The paper presents the annotation of a Slovene language corpus at the semantic level. Manual annotation was performed in two cycles with an automatically generated semantic lexicon according to the wordnet model. The analysis of the results shows that nearly all polysemous words in the corpus can be assigned a sense from our wordnet but also that the task was quite challenging; in many cases, wordnet sense distinctions are too fine-grained even for human annotators to distinguish between them. This is why annotation with more coarse-grained senses could prove to be more successful.

Keywords: corpora, semantic annotation, semantic lexicon, Slovene language.
PETER ĎURČO¹
RADOVAN GARABÍK²
DANIELA MAJCHRÁKOVÁ²
MATEJ ĎURČO³
¹Univerzita sv. Cyrila a Metoda v Trnave, Trnava, Slovakia
²L. Štúr Institute of Linguistics, Slovak Academy of Sciences, Bratislava, Slovakia
³Austrian Academy Corpus, Austrian Academy of Sciences, Vienna, Austria

CONTRASTIVE DICTIONARY
OF GERMAN AND SLOVAK COLLOCATIONS²

Abstract. In the article we discuss ongoing work concerning a confrontational German-Slovak collocation lexical database. The database consists of two parts, a section of German collocations with Slovak equivalents and a section of Slovak collocations. Intended size of the database is several hundred words of different parts of speech (nouns in the first phase of the project) for each of the languages, together with their collocation profiles. The database uses MediaWiki engine and a wiki-based approach to article editing and collaborative work of a team of lexicographers.

References


² The lexical database has been supported by the grant agreement VEGA 1/0006/08 Konfrontačný výskum kolokácií v slovenčine a v nemečine. The study and preparation of these results have been partly supported by the EC’s Seventh Framework Programme [FP7/2007-2013] under the grant agreement 211938 MONDILEX. Parts of this article have been published in Ŏurčo et al. (2009).
Abstract. The paper discusses some problems related to entry classifiers in digital dictionaries. Information technologies offer great possibilities to linguists and lexicographers for the development of various dictionaries, especially for bi- and multilingual digital dictionaries. The authors' point of view is based on their experience from the development of the first Bulgarian-Polish Digital Dictionaries. The dictionaries are being developed in the framework of the joint research project “Semantics and Contrastive linguistics with a focus on a bilingual electronic dictionary” between IMI-BAS and ISS-PAS, coordinated by L. Dimitrova and V. Koseska. The experimental version of the Bulgarian-Polish electronic dictionary is prepared in WORD-format and consist approximately 20 thousand dictionary entries. The dictionary is used for creation of the lexical database (LDB) that is an entry point to the relational database (RDB) of the first Bulgarian-Polish online dictionary. The structure of the LDB allows synchronized and unified representation of the information for Bulgarian and Polish, which is a step towards the creation of online Polish-Bulgarian dictionary in the future.

Keywords: digital dictionary, entry classifiers, digital corpus, semantics and contrastive studies.
Abstract. The paper shortly describes the first Bulgarian-Polish digital corpus. The corpus is collected with the main purpose to ensure the selection of the entries for the first experimental electronic Bulgarian-Polish dictionary. The texts were collected concurrently and do not have a connection with national monolingual or other corpora. This bilingual corpus supports the lexical database of the first experimental online Bulgarian-Polish dictionary. The use of digital multilingual corpora in contrastive studies is briefly discussed.

Keywords: multilingual electronic corpus, parallel and comparable corpora, corpus annotation, bilingual digital dictionary, Bulgarian, Polish.
DIANA BLAGOEV
Institute for Bulgarian Language, Bulgarian Academy of Sciences, Bulgaria

ELECTRONIC CORPORA AND BULGARIAN NEW-WORD LEXICOGRAPHY

Abstract. The article substantiates the need to apply a corpus-based approach when describing neological lexis in a lexicographic way. It introduces the specialized lexicographic electronic corpus, used as an empirical basis for the compilation of the latest academic Bulgarian dictionary of new words. The procedure for semi-automatic extraction of neologisms from an electronic corpus applied in compiling the lexical frame of this dictionary is described in the article.
Keywords: neologisms, new-word lexicography, corpus linguistics.
MATCHING FRAMENET FRAMES WITH POLISH SENSE DIVISIONS: THE CASE OF Jechać

Abstract. The paper presents a proposal of a novel sense division for the Polish lexeme jechać, arising from an attempt to match the dictionary sense with an appropriate FrameNet frame. Polish lexicography proposes a single sense for all cases of humans going on land by some means of transportation, in order to contrast the verb jechać with the verb iść ‘to go on foot’. However, all factual examples of use of the verb jechać can easily be matched with one of three relevant frames: “Operate_vehicle”, “Ride_vehicle” and “Travel”, at least from the wider context. The paper shows how a parallel sense division can be established using purely linguistic, i.e. syntactic and semantic, criteria.

Keywords: Frame, frame semantics, sense division, Polish motion verbs.
JANUSZ S. BIEŃ
Department of Formal Linguistics, University of Warsaw, Poland

FACILITATING ACCESS
TO DIGITALIZED DICTIONARIES IN DJVU FORMAT

Abstract. One of the best formats for scanned documents is DjVu. An essential feature of the format is the hidden text layer, usually containing the results of Optical Character Recognition. Another important feature is the ability to store (and serve over Internet) the documents as a collection of individual pages.

From the very beginning the DjVu format has been used also for dictionaries, in particular there are several Polish dictionaries available in this format. So the question is how to search efficiently the text layer in such large multi-volume works. For this purpose we intend in particular to adapt Poliqarp (Polyinterpretation Indexing Query and Retrieval Processor), a GPLed corpus query tool developed in the Institute of Computer Science of Polish Academy of Sciences. Some preliminary experiments are described in the talk.

In our “quick and dirty” approach we treat every page as a single document with the metadata consisting of the name of the document index and the name of the file with the page content. For every word, instead of grammatical tags, we provide its localization on the page in the form of the line number and its position in the line. All the data taken together allow to link the search results to the appropriate fragments of the original scans.

We mention also another approach to the problem, exemplified by djvu-xfgrep program.

Keywords: digitalization, DjVu, dictionaries, Poliqarp, djvu-xfgrep
IDiom VARIABILITY IN CROATIAN:
THE CASE OF THE CONTAINER SCHEMA

Abstract. In cognitive linguistics most idioms are considered to be motivated by various cognitive mechanisms which link the meaning of the idiom with the meanings of its constituents. One of these mechanisms is the CONTAINER image schema. In Croatian it is reflected in idioms containing the preposition  "in". The CONTAINER schema serves to structure abstract conceptual domains like SITUATIONS, EVENTS and STATES. For example, being in a difficult situation is conceptualized as being in a container. In addition to motivating the idioms with the constituent "in", the CONTAINER schema also constrains their variability. This means that variations are not unpredictable, but are dependent on the underlying cognitive mechanism. The aim of the paper is to show that the Croatian idioms describing difficult situations vary their lexico-syntactic structure systematically to reflect the CONTAINER image schema. Based on the data from the Croatian National Corpus, we will show that the idioms share a common element, the construction "in + NP", which constitutes the fixed core of each of the expressions and serves as the basis for variant realizations. The lexical and syntactic variations reflect the different ways in which the relation between the trajector (a person) and the landmark (a container-like object) is conceptualized. The variability of the expressions raises the issue of the criteria used in establishing the citation form in monolingual dictionaries of idioms. Keywords: idiom variability, conceptual motivation, the CONTAINER image schema, Croatian
MATEUSZ-MILAN STANOJEVIĆ
BARBARA KRYŽAN-STANOJEVIĆ

Faculty of Humanities and Social Sciences, University of Zagreb, Croatian

LEVELS OF CONSTRUCTIONAL MEANING:
THE CONFLUENCE OF THE DATIVE AND MIDDLE VOICE
IN POLISH AND CROATIAN

Abstract. In Croatian and Polish various constructions with the reflexive marker se/się may or may not involve a noun in the dative case. In Croatian one may say govori se o ovome problemu ‘this problem is discussed’ as well as stalno im-DAT se govori o tom problemu ‘they are being told about this problem all the time’. Other examples include, for instance, Kto wie, co się zdarzy za dziewięć miesięcy (Polish) ‘Who knows what will happen in nine months’ as opposed to A jeżeli zdarzy im-DAT się coś złego? ‘And what if something bad happens to them?’. In this paper we will discuss the way in which the se/się construction interacts with the dative case in the construction of meaning. A corpus study was conducted on the IPI PAN corpus of Polish (http://korpus.pl/) and the Croatian National Corpus (http://www.hnk.ffzg.hr) to find examples where the se/się construction coincided with the dative construction. The results show that there are two basic semantic groups: the allative/competitor group and the transfer group, which partially corresponds to semantic groups found for various dative senses (Stanojević and Tužman Vuković forthcoming). In the allative/competitor group the dative serves as an abstract goal, and the se/się construction marks the self-movement of the agent (i.e. the fact that it has internal energy). As opposed to that, in various transfer subsenses the se/się construction is grammaticalized to defocus the agent, and the dative gradually changes its role from a potentially affected recipient (as in stalno im-DAT se govori o tom problemu ‘they are being told about this problem all the time’) to a completely affected experiencer (Meni-DAT kad se plače plačem ‘When I feel like crying I cry’; Wszystko można, tylko człowiekowski-DAT się nie chce ‘Anything can be done, but a person simply doesn’t feel like it’). In these senses both the dative and the se/się construction are grammaticalized in respect to their other senses, and are hence semantically bleached. Therefore, in those senses new constructional meaning occurs, which is not present in any senses of the two components taken alone: dative as the experiencer of its internal change of state. Constructional meaning is possible only in the bleached senses, which are less
detailed in respect to the “basic”, diachronically older senses.

**Keywords:** dative, middle voice, grammaticalization, Croatian, Polish.
MAKSIM DUŠKIN
Institute of Slavic Studies, Polish Academy of Sciences, Poland

CONCERNING EXPONENTS OF ADNUMERAL APPROXIMATION IN POLISH AND RUSSIAN

Abstract. This article attempts to present the most important results of an analysis of Polish and Russian exponents of adnumerative approximation (e.g. Pol. około, przeszło, Rus. более, примерно, etc.). “Approximation” is defined here as a mechanism of numeric attribution, based on designating a segment in a series of numbers, instead of designating one particular point. This definition encompasses several different ways to designate a segment. All of those ways have separate exponents in Polish and Russian. Analysis indicates that exponents of the same sense can differ in terms of connectivity, semantics and stylistics. These differences should be taken into consideration while establishing Polish and Russian translational equivalents.

Keywords: Approximation, Approximate, Polish, Russian.
Abstract. The aim of this work is a syntactic analysis of Bulgarian constructions translated from sentences with uninflected perfect participles of the poem Pan Tadeusz ("Mr Thaddaeus") by Adam Mickiewicz. The analysis is based on the observation known from many other dissertations that if syntactic constructions, whose centres are both uninflected and inflected participles with regard to the meaning and syntactic function, cannot correspond in the translation to syntactic constructions with a participle as the centre, they correspond to clauses in precisely defined types of compound sentences.

Keywords: semantic layer, concurrent character, interpretation of the sequence of events, indicator of connection, translating
FROM THE WORKS ON THE BULGARIAN-POLISH
DICTIONARY OF VERBO-NOMINAL ANALITYCAL
CONSTRUCTIONS

Abstract. The following article discusses the work on the Bulgarian-Polish dictionary of verbo-nominal analytical constructions. The goal of the authors was to create a dictionary that would address contemporary lexicographical issues and thus be of use to both interpreters and linguists.

Keywords: lexicography, verbo-nominal analytical constructions, normative problems
EWA MICZKA
Université de Silésie, Katowice

LES STRUCTURES SITUATIONNELLES ET INFORMATIONNELLES DE DISCOURS

Abstract. The article regards relations between situational and information structures of discourse. Analyzing the possible configurations between these two types of structures, the author aims to present their role in discourse comprehension — the process which implicates creation of discourse representation.

The situational structures are defined as a sequence of frames Goffman (1991). Each frame permits to conceptualize one event forming a part of information introduced in discourse. The author proposes to apply the notion of cognitive event and their typology introduced by R. Langacker to describe the variations of situational structures of discourse.

The information structures are defined as hierarchically organized thematic-rhematic structures. The author distinguishes three levels in their thematic part represented by: global theme, theme of group of sentences and theme of sentence. The rhematic part is divided in two levels: the first one contains rhematic groups, the second rhemes of sentences.

The author focuses her attention on the highest level of information structure and describes the relations between the units of situational structures — frames and cognitive events — and choices regarding the global theme of discourse.

Keywords: Situational structures, information structures, theme, rheme, cognitive event, frame.

References

EWA GWIAZDECKA
Université de Varsovie, Warsaw, Poland

QUELLE DESCRIPTION POUR LE PRÉVERBE POLONAIS ?

Abstract. In this paper, we are trying to provide the semantic description for Polish verbal prefix. We are particularly interested in the prepositional "history" of this operator and the way the diachronic relation contributes to aspectuality. We use topological and quasi-topological operators for description of spatiality, temporality and activities encoded by preposition and verbal prefix. The topological intervals serve to represent basic aspectual situations: state, process and event. These aspects are regarded as a property of the whole predicative relation understood as an application (in the formal sense) of a predicate to its arguments. Moreover, we claim that in the binary predicative relation, $P_2 T^2 T^1$, the compositionality between verbal prefix and the second argument $T^2$ is of aspectual relevance. It follows that aspect encoded by prefixation does not focus on the verb only, but it can determine a spatial place or an object. We argue that in the majority of the cases, verbal prefix encodes an achievement understood as a closure of the right boundary of the process associated with the predicate, which coincides with the focus on some topological zone. Achievement is understood qualitatively and does not always implies the "completion" of action, but according to the meaning of the verbal prefix, it can indicate "achievement of the beginning", "achievement of the end phase", etc.

Keywords: aspect; Polish verbal prefix; preposition; achievement; state; process; event; topology; semantics.