VIOLETTA KOSESKA

Institute of Slavic Studies, Polish Academy of Sciences Warszawa, Poland
`amaz@inetia.pl`

SEMANTICS, CONTRASTIVE LINGUISTICS AND PARALLEL CORPORA

Abstract

In view of the ambiguity of the term "semantics", the author shows the differences between the traditional lexical semantics and the contemporary semantics in the light of various semantic schools. She examines semantics differently in connection with contrastive studies where the description must necessary go from the meaning towards the linguistic form, whereas in traditional contrastive studies the description proceeded from the form towards the meaning. This requirement regarding theoretical contrastive studies necessitates construction of a semantic interlanguage, rather than only singling out universal semantic categories expressed with various language means. Such studies can be strongly supported by parallel corpora. However, in order to make them useful for linguists in manual and computer translations, as well as in the development of dictionaries, including online ones, we need not only formal, often automatic, annotation of texts, but also semantic annotation — which is unfortunately manual. In the article we focus on semantic annotation concerning time, aspect and quantification of names and predicates in the whole semantic structure of the sentence on the example of the "Polish-Bulgarian-Russian parallel corpus".

Keywords: contrastive studies, online dictionary, parallel corpora, direct approach to semantics, semantic interlanguage, Petri nets, semantic annotation.

1 Semantics

Due to the understanding of "semantics" as a discipline of studies dealing with the meaning, i.e. assignment of linguistic symbols to extra-language objects in a broad sense, even the term "meaning" itself acquired "various meanings", depending on the individual semantic schools. As we know, TRADITIONAL LINGUISTICS defined semantics as *the study of meaning of words*. In CONTEMPORARY LINGUISTICS, as Kazimierz Polański wrote, the focus shifted *from words to the sentence* (Polański, 1999, p. 341) This implies that words mean something in the whole structure of a sentence or an utterance, but not outside the sentence (utterance), i.e. separately.

"Even if — writes Polański — we use a single word as a sentence or utterance, this will be a sentence or utterance consisting of a single word". It is worth emphasizing that the above applies to semantic phenomena expressed formally in the language rather than to extra-linguistic phenomena, see Item 2 below. As one of the experts and scientific advisors assisting the development of the theoretical concept of the Bulgarian-Polish Contrastive Grammar (from now on, the BPCG), see e.g. (BPCG 1990), K. Polański was of the opinion that SEMANTIC PHENOMENA IN A NATURAL LANGUAGE can be described in the most precise way using logical theories and the conceptual apparatus of mathematical logic — a view I completely agree with, see (Studia, 1984).

2 Direct approach semantics

As I have repeatedly stressed, the semantic volumes of the BPCG are based just on selected logically-mathematical theories: namely, quantification theory or contemporary process theory, known as "Petri net" theory. Semantics is understood here like in the "direct approach semantics" works of B. Russell and H. Rasiowa, and in later works on situation semantics by Barwise and Perry (see Russell, 1967; Rasiowa, 1975; Barwise & Perry, 1983; Cooper, 1996). Scholars subscribing to this semantic trend were interested a fact which, though in principle known, was not fully realized by many: that THE SAME LINGUISTIC FORMS (WORDS, EXPRESSIONS, SENTENCES) CAN BE CARRIERS OF QUITE DIFFERENT INFORMATION CONTENTS, see in more detail (Koseska, 2013). While in other semantic schools the meaning of a sentence is defined with help of two abstract objects (in case of Frege (Frege 1892), these are truth and falsity), the theory with a direct approach to semantics discussed here introduces into the definition of the meaning of a sentence also the notion of a SITUATION, and DEFINES THE MEANING OF A SENTENCE AS A SET OF ABSTRACT SITUATIONS (Barwise & Perry, 1983; Cooper, 1996). Standard and classic examples of theories with direct approach to semantics are Bertrand Russell's denotation and description theories (Russell, 1967), used in the second volume of the Bulgarian-Polish Contrastive Grammar (Koseska & Gargov, 1990) and model theory of first order predicate logic, which, being extensional, is applied to a natural language with considerable limitations, see (Rasiowa, 1975) and the exemplary sentence: *Maria kończy studia* [Mary is finishing her studies], which is understood as true information in opposition to the sentence: *Mówi się, że Maria kończy studia* [Mary is said to be finishing her studies], which is understood as information that might be true, but its truth is not certain.

2.1 It is worth stressing here that contemporary linguistic schools combine semantics with mathematical logic in relation to a sentence rather than its fragments, out of which a sentence can only be built. The branch of contemporary linguistics understood in this way decisively distinguishes between semantics and semantic syntax concerning the argument-predicate structure of the sentence (i.e., its semantic segments having different syntactic order in different languages). By way of example, the predicate: *x jest chory* [x is sick] becomes a sentence only after assigning an argument value to x or assigning a quantifier to x — see, respectively, *Jan jest chory* [John is sick] and $(ix)P(x)$, where i is the iota-operator, and (ix) jest chory [(ix) is sick] is a sentence. In some linguistic works, the above sentence with

the iota-operator together with x is erroneously treated as a "predicate" rather than a sentence. Another problem which can be observed e.g. in the syntactic volume of the academic grammar of Polish is the fact that in its part authored by S. Karolak a "predicate" is not understood as a propositional function, like in the part authored by M. Grochowski. There is no need to convince the reader that ambiguity of such a commonly described notion as a "predicate" still confuses some scholars. The BPCG assumes that a "predicate" is to be understood as a propositional function, i.e. the way a "predicate" is understood by M. Grochowski, see the volume of the Contemporary Polish Grammar (Gramatyka, 1984) devoted to the syntax.

3 Semantics versus contrastive studies — Theoretical contrastive studies
Since in the BPCG we understand semantics as above, why in relation to contrastive studies do we treat semantics differently? Contrastive studies are known to be a branch of synchronous linguistics with both theoretical and practical applications. When contrastive studies deal with analysis of differences and similarities between languages for practical (didactic or translational purposes), we speak of them as of a branch of APPLIED linguistics, connected first of all with teaching of foreign languages. In turn, we speak of THEORETICAL contrastive studies only in the case when the studies concern universal language notions, when they employ methods of linguistic research aimed at isolating, in an equal way, what is common and what is different in the studied languages. From the viewpoint of research methods employed, as well as the use of synchronous approach, theoretical contrastive studies are close to TYPOLOGICAL studies, but differ from them in the purpose of description: typological studies lead to classification of languages, while contrastive studies — to systemic analysis of the contrasted languages. Without doubt, also in this case a good solution would be an INTERLANGUAGE, allowing for equal comparison of the meanings and forms of the studied languages. However, development of such a tertium comparationis is not an easy task (Koseska, Korytkowska & Roszko, R., 2009), see also (Koseska, 2013; Selinker, 1972).

3.1 The process of developing a semantic interlanguage can be divided in a number of stages:

1. Selection of a universal semantic language category — e.g. definiteness/indefiniteness, time, communicant, etc.

2. Selection of logically-semantic theory to be used for developing the notion system of the interlanguage, e.g. logical quantification, network-based description of time in a natural language, etc.

   Defining notions in accordance with the selected theory, see below.

   Developing a terminological vocabulary starting from the notions of the semantic interlanguage.

Ad. 1. Selection of semantic language categories must absolutely take into consideration the specifics of Slavic languages. By example, the well-known and popular in linguistics logically semantic Reichenbach's theory, which describes time in a natural language, concerns in fact English, which has no grammatical category

of aspect. Hence it does not distinguish between the meanings of time and aspect — and without such a distinction, the description of the semantic issues of time in each of the Slavic languages where the grammatical category of aspect has become distinguished will be both incomplete and false.

Ad. 2. The interlanguage should be developed BASED ON THEORIES NOT LEADING TO THEORETICAL CONTRADICTIONS. For example, when creating the basic semantic units used in the interlanguage to describe the language category of definiteness/indefiniteness, we can use reference theory, but also theory of definite description and quantification. However, a simultaneous use of both theories is not recommended, since its leads to internal contradictions in the notion system of the interlanguage. This can be observed in works that do not distinguish between the selected notions, e.g. such as reference and definite description. Already based on Volume 2 of the Bulgarian-Polish Contrastive Grammar (Koseska & Gargov, 1990) we can see that a description that takes as starting point Bulgarian formal language means is totally different from a description originating from Polish formal language means. This is determined even by the more expanded morphological plane of the means for expressing the notions of definiteness and indefiniteness in Bulgarian compared to Polish, see also (Koseska & Mazurkiewicz, 1988). Hence it would be a serious methodological error to replace the interlanguage by one of the languages being contrasted, together with its metalanguge, Nevertheless, this is how that issue is treated in most of the works we know, where the description of the language proceeds from the form to the contents. The latter approach is the basis of most grammars, which describe one language (most often, a foreign one) using another (native) language.

4 Direction of description

Distinguishing semantics in its relation to contrastive studies is connected with the DIRECTION OF DESCRIPTION of language phenomena and with GOING FROM THE MEANING TOWARDS THE FORM rather than, as is the custom in traditional contrastive descriptions and descriptive presentations of a single language, from the form towards the contents.

4.1 In order to develop a description going from the meaning towards the forms in two or more languages, one should precisely DISTINGUISH BETWEEN A FORM AND ITS MEANING. For example, the use of the term "definiteness" in the cases when the so-called "definite article" expressed INDEFINITENESS, i.e. UNIVERSALITY, was an obvious error in interpreting the meaning of the article, and followed from failure to distinguish between the form of the article and its meaning. In our works, the definiteness/indefiniteness category has been defined as a category with semantic opposition: uniqueness — non-uniqueness, whereby definiteness is understood solely as UNIQUENESS OF AN ELEMENT OR A SET (satisfying the predicate), and by indefiniteness — NON-UNIQUENESS (both EXISTENTIALITY AND UNIVERSALITY) (Koseska, 1982; Koseska, Gargov, 1990).

4.2 The interlanguage for contrasting Polish and Bulgarian within the semantic category of definiteness/indefiniteness used in our works is based on the theoretical assumption about the quantificational character of this category. The basic

notion of uniqueness (of an element and of a set) in that interlanguage could be written down using a linguistic iota-operator construction (in the text, shortly "iota-operator"), that of existentiality — with help of an existential quantificational expression (in the text, shortly existential quantifier), and of universality — with help of a universal quantificational expression (in the text, shortly "universal quantifier). For a description of the definiteness/indefiniteness category using the logical theory of scope-based quantification, see (Bellert, 1971; Barwise & Cooper, 1981; Cooper, 1996; Grzegorczyk, 1972, 1976; Koseska, 1982; Koseska & Gargov, 1990; Descles, 1999; Roszko, R., 2004; Koseska, 2006; Roszko, D., 2014).

4.3 The methodology presented above is applied in SEMANTIC ANNOTATION, which we impose by hand in PARALLEL CORPORA. In Bulgarian, as I have already written in many works, the most typical morphological means for expressing UNIQUENESS AND UNIVERSALITY in the nomen group, is deemed to be the article. Its absence, or morphological Ø, is an exponent of existentiality, or pure predication. The ambiguity of the Bulgarian article is a good illustration of the difficulties encountered by a scholar studying this category when classifying natural language expressions. As I have already mentioned, in Bulgarian the same form of the article expresses both uniqueness and universality (or, respectively: definiteness and indefiniteness). In the already quoted book (Koseska-Toszewa, 1982), I put forward a hypothesis concerning development of the meaning of the Bulgarian article. In my opinion, the article initially expressed uniqueness of an element (object), and then started expressing also uniqueness of a set, which later, as a result of equalling two completely different semantically-logical structures, i.e. structures with universal and with unique quantification, led to homonymy and to the article expressing also universality. I later confirmed the above observations, based first of all on semantically-logical aspects of the definiteness category, on the historical language material from *Kodeks Supraski* [the Supraśl Code], where the Bulgarian article does not appear yet in universally quantified nominal structures, but in uniquely quantified nominal expressions, meaning satisfaction of a predicate by either a single element of a set or by the whole set treated as the only one, see (Zaimov, 1982, p. 5–9).

4.4 Let us return to the second important feature of semantics and contrastive linguistics, that it, distinguishing between language forms and their meanings. It should be stressed that WITHOUT DISTINGUISHING BETWEEN A LANGUAGE FORM AND ITS MEANING, CONTRASTING MATERIAL FROM SEVERAL LANGUAGES MAY LEAD TO COMMITTING NUMEROUS SUBSTANTIVE ERRORS AND DRAWING NON-SCIENTIFIC CONCLUSIONS. The above issues would be only postulates, were it not for the 12-volume academic Bulgarian-Polish contrastive grammar (BPCG, 1990–2007), which solved the above-mentioned problems.

5 Form, meaning and corpora

Distinguishing between language forms and their meanings is of key importance for the work on unilingual and multilingual corpora, which once more brings the problems of semantics and contrastive linguistics closer to the use of parallel corpora. Without distinguishing between language forms and their meanings, it is difficult to imagine what the application of parallel corpora and their benefits will

be. To distinguish between a language form and its meaning, it is not enough to write "language form" and "language form content". In order to know what content of the language form we have in mind, we need thorough research on the set of contents of a selected semantic language category that the given content belongs to. What is more, to contrast language material in parallel corpora of several languages, we need also to distinguish between contrastive and comparative studies. Contrastive studies are synchronous, while comparative ones — diachronic. This is important for selecting texts which are synchronous on the contrastive level, and in our case concern contemporary development of the languages being contrasted in the corpora.

6 Semantic category of time

6.1 The semantic category of time in Slavic languages selected here as an example also contains information about the aspect of verbum, and can be described in a detailed and precise way using the contemporary formal theory of processes known as Petri nets (Petri, 1962; Mazurkiewicz, 1986; Laskowski, 1986; Koseska & Mazurkiewicz, 1988). In my opinion, identifying the meaning of the Slavic aspect with the so-called "action type" is a fundamental substantive and methodological error. The Aktionsart teory, or semantic category of a verb, has been distinguished in German linguistics due to the way in which an action is to run in languages where no grammatical category of aspect has developed, like in all Slavic languages. The Aktionsart theory has allegedly given rise to various meanings concerning the forms (kinds) of actions, e.g. so-called inchoative ones (representing the beginning of an action), or desiderative ones (representing the wish to perform some action). The grammatical category of Aspect characterizes Slavic verbs only. This problem can be well illustrated using Bulgarian, which, in opposition to e.g. Polish or Russian, has a very expanded system of aspectually-temporal meanings, which allows us to understand why such renowned aspect specialists as S. Ivanchev (1971) insisted on consistent treatment of temporal and aspectual meaning of the verbum. This is because in Bulgarian the aorist form is derived from both perfective and imperfective verbs, and likewise the imperfectum form is derived from both imperfective and perfective verbs. It is difficult to understand and translate the meaning of aorist of imperfective verbs in isolation from the meaning of aspect combined with the temporal meaning of the verbal form. The meaning of aorist of imperfective verbs is translated to Polish using the praeterite form of imperfective verbs. However, take care! Also the Bulgarian imperfectum form of imperfective verbs is translated to Polish using the praeterite form of imperfective verbs, though in both cases we have to do with different aspectually-temporal meanings, see (Koseska, 1982, 1985, 1995, 1977, 2006). Both meanings differ in being a placeholder for a quantifier expression which occurs in the semantic structure of the sentence next to the verbum, and this is totally independent of the so-called kind of action. The imperfectum form of imperfective verbs is a placeholder first of all for either a "universal quantifier" or an "existential quantifier" in the semantic structure of the sentence, but rarely appears next to a "iota-operator expression". In turn, the aorist form of imperfective verbs is a placeholder ONLY for an iota — operator expression. In Polish sentences, the praeterite form of imperfective verbs is in one case a placeholder for either a

"universal or existential quantifier", and in the other case for a "iota-operator". In Polish we very often have to do with so-called incomplete quantification (a term coined by Ajdukiewicz), see (Ajdukiewicz, 1974), (Koseska & Gargov, 1990). Incomplete quantification in Polish sentences can be "completed" only with either a semantic paraphrase or a set of situations concerning selected sentences.

Examples:

1. Aorist of imperfective verbs.

   In Bulgarian the aorist form of imperfective verbs does not appear next to existential expressions (existential quantifier) and universal expressions (universal quantifier):

   * (*Понякога ∥ винаги*) *спа у тях.* 'She (sometimes ∥ always) slept at their place.'

2. Aorist of imperfective verbs.

   Appears only next to iota-operator expressions (iota-operator):

   Тази нощ спа у тях. 'She slept at their place that night'

3. Imperfectum of imperfective verbs.

   Appears most often next to existential and universal expressions (existential and universal quantifiers):

   (*Понякога ∥ винаги*) *спеше у тях.* 'She (sometimes ∥ always) slept at their place.'

4. Imperfectum of imperfective verbs

   Appears next to iota-operator expressions (iota-operator):

   По това време тя спеше у тях. 'At that time she was sleeping at their place.'

   As we can see, Bulgarian aorist of imperfective verbs appears only next to the iota-operator, similarly as aorist of perfective verbs (Koseska, 2006).

6.2 Let us also consider the following verb forms:

избръмч|á, -иш *vp. event, intransitive*; clink,
избръмчáва|м, -ш *vi. state, intransitive*; start clinking,
забръщолéв|я, -иш *vp.pot. event, transitive*; babble,
забръщолéвя|м, -иш *vi. pot. state, transitive*; start, babbling
забарабáн|я, -ш *vp. event, intransitive*; start drumming, drum,
забарабáня|м, -ш *vi. state, intransitive*; start drumming.

In the above cases, the Bulgarian verb form has an inchoative meaning. However, this does not change the fact that both verb forms are perfective as well as imperfective, and have different aspectual meanings, see Bulg. Пчелата кацна за момент върху розата (event), **избръмча няколко пъти** (sequence of events

and states ended with an event) и полетя (event). 'A bee sat down on a rose for a moment, buzzed a few times, and flew away.'

From the above examples we can immediately see that "aspect of a verb" and the "kind of action" are two different things. Combining Aktionsart and aspect theories in Slavic languages, one erroneously identifies two mutually contradictory linguistic theories. This is a phenomenon resembling the erroneous combining of reference theory and scope-based quantification theory concerning the semantic definiteness / indefiniteness category, see e.g. the frequently encountered expression "referential quantification", from which we cannot tell whether we have to do with quantification or reference, and hence it is the proverbial "mumbo jumbo". What we do know, however, is that both theories are mutually contradictory and are related to different approaches to the semantics of a natural language. Identifying the meaning of the verb aspect and the type of the verb action is an inadmissible phenomenon in constructing the interlanguage in contrastive studies, because when developing tertium comparationis we try to formulate notions and terms that are unambiguous and do not contradict each other.

6.3 When describing the semantic category of time selected by way of example, we adopt STATES AND EVENTS as fundamental units of time and aspect description. The basic characteristic distinguishing these notions is the *temporal spread of states* and the *momentary character of events*. In other words, states "last", while events can only "happen". An abstract counterpart of this distinction is the difference between a section of the real axis (a state) and a point lying on that axis (an event). The adopted postulate of model finiteness implies that in constructing our description we cannot limited ourselves to events only, and in consequence treat states as sets of events, as e.g. done by Reichenbach (1967). This is because when describing a state as a set of events we have to answer the question: "Set of what events? All or only some of them? And if some, then how to choose them?" Omitting events in the model and limiting it solely to states deprives us of the capability to consider such phenomena as "collision", "opening", "uncovering", "awakening" and the like. A characteristic feature of events is that we cannot speak of them in the present tense: this is because an event does not last — it has no time spread. Referring to the quoted analogy with points and sections, we can say that events correspond to points, states — to sections, and the mutual relationship of events and states is like the relation between points and sections: each point is either the beginning or the end of some section (or in a special case of as half-line); each event is either the beginning or the end of some state (e.g. the state preceding the occurrence of the event or the state following the event). The analogy can be continued: each section, similarly as each state, has at most one beginning and one end, while each point (each event) can begin or end many sections that are interesting for us (many states). In other words, an event need not be the beginning or an end of only one state, and hence it cannot be treated as just an ordinary transition — "transition of a state into another state".

It should be emphasized that imperfective verb forms express not only 1. state, but 2. states, that is, sequences of states and events, finally ended witha state,

while perfective verb forms express both 1. events and 2. sequences of events and states, finally ended with an event.

7 Semantic annotation

We will need the above research results for manual semantic annotation, which we want to present on the material of parallel corpora in four languages under the "Clarin"[1] project. CLARIN is a pan-European scientific infrastructure that will allow scholars from the area of humanities and social sciences to work conveniently with very large text sets. The aim of CLARIN is to overcome a number of barriers in access to the so-called language technology by scholars who have no specialist knowledge in the area of IT or natural language engineering. CLARIN-PL is the Polish part of the great European network, tightly integrated with it, but directed specially at the Polish language, both written and oral.

The semantic annotation applied in our corpora concerns also quantification of names and predicates rather than only meanings of aspectually-temporal verb forms. We are the first to use such annotation for the first time in the literature of the subject. It requires knowledge of deep semantic structure of natural language sentences and a reliable theoretical apparatus, thanks to which the meanings of language forms can be written down in a precise way. Though contemporary software used for parallel corpora covers the problems of the semantics, it concerns mainly lexical semantics — though not always and not only, see e.g. the works of M. Piasecki and his team on plWordNet, Polish "Słowosieć". "Słowosieć" is a unique dictionary providing a formalized description of the meanings of a huge number of Polish words. Each meaning is described through its semantic relations with many other meanings of other words (Piasecki, 2013).

7.1 The semantic annotation we propose can be divided into several levels:

1. annotation concerning time and aspect in Polish, Bulgarian, Russian and Lithuanian.

2. annotation concerning scope-based quantification of nomen at selected syntactical positions and predicates in the sentence.

3. semantic annotation combining annotation concerning the meaning of aspect and time contained in the predicate and quantification of the predicate in the sentence.

---

[1](*Common Language Resources and Technology Infrastructure*) — a scientific project, which in February 2012 obtained the European Commission legal status of ERIC (European Research Infrastructure Consortium). The founders of Clarin ERIC are Austria, Bulgaria, the Czech Republic, Denmark, Estonia, Germany, Holland and Poland. CLARIN is a project from the so-called ESFRI road map (European Roadmap for Research Infrastructures, European Strategy Forum on Research Infrastructures). The main goal of the project is to combine *language resources and tools* for European languages into a single, common and uniform network, which is to become an important working tool for scientists from the humanities in the broad sense of the term.
The Polish participants of Clarin include the following scientific entities: Wrocław University of Technology, Polish-Japanese Institute of Information Technology, Institute of Computer Science PAS, Institute of Slavistics PAS, University of Łódź, University of Wrocław. The main coordinator of Clarin-PL is dr Maciej Piasecki from Wrocław University of Technology.

Polish and Bulgarian examples:

Ad1. Pol. *Maria oczekuje kogoś* (state) *i kiedy od czasu do czasu słyszy samochód* (sequence of events and states, ended with a state), *szybko biegnie* (sequence of events and states, ended with an event) *do okna.*

Bulg. *Мария някого чака* (state) *и когато от време на време чува кола*, (sequence of events and states, ended with an event), *тя се втурва към прозореца* (sequence of events and states, ended with an event).

Ad2. *Motocykl* $?(ix)P(x)$ *ruszył* (incomplete unique quantification).

Моторът тръгна $(ix)P(x)$ — unique quantification.

(Incompleteness of any quantification is marked with a question mark preceding the formal notation.)

Ad3. *Ta służąca* $(ix)P(x)$ *czasami spała* $(\exists X)P(X)$ (sequence of states and events, finally ended with state) u nich.

Тази прислужница $(ix)P(x)$ *понякога спеше* $(\exists X)P(X)$ (sequence of states and events, finally ended with state) *у тях*

7.2 Our notation of quantificational expressions in Polish, Bulgarian and Russian sentences:

$(ix)P(x)$ — iota-operator preceding an Attribute expressed with a noun, (iX)P(X) — jota-operator preceding a predicate expressed with a verb form.

$(\exists x)P(x)$ — existential quantifier preceding an Attribute expressed with a noun,

$(\exists X)P(X)$ — existential quantifier preceding the Predicate expressed with a verb form,

$(\forall x)P(x)$ — universal quantifier preceding an Attribute expressed with a noun form,

$(\forall X)P(X)$ — quantifier preceding the Predicate expressed with a verb form.

Incompleteness of any quantification is marked with a question mark preceding the formal notation.

7.3 We believe that though our annotation is manual (and hence cannot cover a too large language material), it will help scholars analyze universal semantic categories in selected languages, and will surely facilitate translations in the languages we are interested in. This is because we do not translate forms, but their meanings, and not only lexical meanings either.

7.4 Examples from the Polish-Bulgarian-Russian parallel corpus:
Fragment from "Lord Jim" by Joseph Conrad[2]

---

[2]Lord Jim — This reading, like thousands others, is available on the website wolnelektury.pl. A version of the book elaborated substantially and critically (annotations and motives) is available on the website `http://wolnelektury.pl/katalog/lektura/lord-jim`. The work has been elaborated under the project titled Wolne Lektury (Free Readings) by the Nowoczesna Polska (Modern Poland) foundation.
This works is not protected by copyright and belongs to the public domain, which means it can be freely used, published and disseminated.

От устата на капитана на "Патна" *се изтръгна* (**event**) дрезгава ругатня и думата "Shwein" *запърха* (**sequence of events and states, finally ended with event**) като капризно перце, подхванато от лек ветрец.	Z grubego gardła kapitana Patny *wypłynął* (**event**) basowy pomruk, w którym dźwięk słowa "Schwein" *unosił się* (**sequence of events and states, finally ended with state**) to wysoko, to nisko, jak kapryśne piórko przy lekkim podmuchu wiatru.	С уст командира "Патны"*сорвалась* (**event**) хриплая ругань, и слово "schwein"[свинья (нем.)] *запорхало* (**sequence of events and states, finally ended with event**), как капризное перышко, подхваченное ветерком.
Той и старшият механик (ix)P(x) се познаваха от много години — заедно служиха при един весел, хитър стар китаец, който носеше очила с рогови рамки и слагаше червени копринени лентички в почтената си плитка от бели коси.	*On z tym pierwszym maszynistą (ix)P(x)* już od wielu lat byli kamratami, służąc jowialnemu, sprytnemu, staremu Chińczykowi z okularami w rogowej oprawie i kawałkami czerwonego jedwabiu wplecionymi w siwe włosy warkocza.	*Он и старший механик* ?(x)P(x) были знакомы много лет — вместе служили веселому, хитрому старику китайцу, носившему очки в роговой оправе и вплетавшему красные шелковые тесемочки в свою почтенную седую косу.
В базата на "Патна" *хората (∀x)P(x) споделяха* (**sequence of states and events, finally ended with state**) *мнението (ix)P(x)*, че тези двамата — капитанът и механикът — са си лика-прилика по безогледните злоупотреби.	W porcie, gdzie zwykle *stawała (∀X)P(X)* Patna, *krążyła* (**sequence of states and events, finally ended with state**) *pogłoska* ?(x)P(x), że co się tyczy bezczelnego okradania skarbu państwa, "to ci dwaj kompani zrobili wszystko, co się tylko dało".	В родном порту "Патны" *жители ?(∀x)P(x)* побережья *придерживались* (**sequence of states and events, finally ended with state**) того мнения, что эти двое — шкипер и механик — по части наглых хищений друг другу не уступают.

8 Cognitive approach and translations

In view of the different meanings given to the term "cognitive", let me return here to its understanding in our studies. Our contrastive studies remove strict divisions into the "grammar" and the "lexis". We have chosen here universal language categories of material importance for description of the language, which have not been described exhaustively in academic grammars od Polish and Bulgarian up to now — namely, such semantic categories, as time, modality or definiteness/indefiniteness, quantity, communicant, etc. The strict separation of the morphological, syntactic and lexical levels in traditional grammars of those languages did not allow for a comprehensive presentation of semantic language phenomena. In the BPCG we consciously resigned from a strict division into the grammatical and lexical levels, since the fact that selected "content" in a given language is expressed with gram-

matical means in one language, and lexically in another, does not imply that the content is absent in that other language. We refer to this new approach as "cognitive". Further, we understand cognitive studies on the one hand as theoretical semantic studies which allow for taking into consideration language means from different levels: lexical and grammatical ones, but perceived as a single whole. On the other hand, if necessary, we make use of broader language situations, where the phenomena we are interested in are unequivocally understood by language users. Those situations always take also in consideration the language user's state and attitude to the conveyed contents, especially modal ones.

8.1 A comprehensive description of cognitive language phenomena plays a very important role in translations from one language to another. Thanks to such an approach, a translator of Bulgarian will learn that, for example, in Polish there is a semantic category of definiteness/indefiniteness, which traditional grammar does not mention. He/she will also learn what lexical means in Polish convey the various meanings expressed via morphological means of that semantic category in Bulgarian. In other words, the translator will know how to translate the Bulgarian article. However, he/she will not learn that Polish lacks the modality known as imperceptivity, and that this is not a modality typical for Bulgarian only, as stressed by traditional grammars. Imperceptivity in Bulgarian grammar — known as преизказност and преизказно наклонение — is described as morphological means for expressing the state of speaker's lack of knowledge. In Polish similar contents are expressed lexically with help of expressions: *jakoby*, *ponoć*, as well as using the praesens and praeteritum forms of the verb *mieć* appearing with an infinitive, see:

Bulg. *Той бил добър лекар.* and Pol. *On jest ponoć dobrym lekarzem.* 'He is allegedly a good doctor.'
Bulg. *Той бил завършил филология.* and Pol. *On miał ukończyć filologię.* 'He has allegedly completed philological studies.'

A translation based on traditional grammar is often done by "groping in the dark", and its quality depends solely on the translator's intuition and experience. The comprehensive treatment of semantic phenomena in natural languages proposed here can have a large impact on translations, also machine ones. Regarding the latter, we refer the reader to our works in the EU grant "Mondilex" (Koseska & Mazurkiewicz, 2009; Dimitrowa & Koseska, 2008, 2009), as well as in the project titled "Semantics and contrastive linguistics with special emphasis on bi- and multilingual electronic dictionaries" implemented through cooperation between IS PAN and IMI BAN, as well as an emerging Bulgarian-Polish electronic dictionary, (Dimitrova, Koseska-Toszewa & Satoła-Staśkowiak, 2009).

9   International and interdisciplinary project "Clarin" (see footnote 1 above) and parallel corpora
Development of a trilingual "Polish-Bulgarian-Russian" corpus, together with a "Polish-Lithuanian parallel and topical corpus" is the scientific task of the IS PAN Corpus Linguistics and Semantics Group in the "Clarin" project[3].

---

[3]The Coordinator of Clarin-PL on behalf of IS PAN is Violetta Koseska-Toszewa. The Manager

9.1 A parallel corpus is known to be a corpus where each text has its counterpart in at least one other different language. In a parallel corpus, texts are aligned at the paragraph and sentence levels. In the two corpora being developed by the IS PAN Corpus Linguistics and Semantics Group, the Polish-Bulgarian-Russian and the Polish-Lithuanian one, the above four languages are aligned on the sentence level. At the present advancement level of work on our corpora, we already have literature covering about 4 million words. By the assumptions of Clarin, each of our corpora is to have 12 million words parallelized down to as sentence. Parallel corpora will provide a larger and richer list of possible translations of the word searched for, and will modernize semantic classifiers in contemporary dictionaries, which will make translator's work radically easier. As databases where translators can find the analogues of the words and expressions sought in different languages (together with the context, in which their given translation has been used), parallel corpora will also serve as an excellent tool during translation courses, which can inspire the students to look for a greater number of possible translations of the words and expressions sought. We should stress that parallel corpora are a reliable starting point for developing bi- and trilingual dictionaries — and as we known development of new bi- and multilingual dictionaries in paper and electronic forms is the task with the greatest importance for the process of integrating the communities in Europe, which was stressed on December 2, 2009 in Bratislava at the Conference of Directorate-General for Translations of the European Commission titled "Impact of European integration on Slavic languages. Trends in translation and multilingual communication". (See Appendix 1. Conference Programme).

Appendix 1

The Programme of the Conference and Directorate-General for Translations of the European Commission titled "**Impact of European integration on Slavic languages. Trends in translation and multilingual communication**":

Dušan Čaplovič, wiceminister ds. społeczeństwa opartego na wiedzy, spraw europejskich, praw człowieka i mniejszości

Karl-Johan Lönnroth, dyrektor generalny w Dyrekcji Generalnej ds. Tłumaczeń Pisemnych (DGT) Komisji Europejskiej

Peter Zsapka, kierownik Wydziału Komunikacji Społecznej w Przedstawicielstwie Komisji Europejskiej w Republice Słowacji

Globalizacja i jej wpływ na dziedzictwo językowe i kulturowe. Języki słowiańskie i Unia Europejska

Wielojęzyczność w UE: wojna, konkurencja czy współistnienie?
Prof. dr Marko Stabej, Instytut Slawistyki, Uniwersytet w Lublanie

Jak rozwija się język "euroczeski" w UE?
Doc. dr Jana Hoffmannová, Dr Josef Šimandl, Instytut Języka Czeskiego, Czeska Akademia Nauk

---

of Module 8 of the Project is Wojciech Sosnowski. The Manager of Module 9 of the Project is Roman Roszko. Person working on Modules 8. and 9. include (in alphabetical order) Anna Kisiel, Violetta Koseska-Toszewa, Danuta Roszko, Roman Roszko, Joanna Satoła-Staśkowiak, Wojciech Sosnowski.

Pułapki językowe między bliskimi systemami leksykalnymi w językach słowiańskich
Prof. dr Marta Pančíková, CSc., Instytut Filologii Słowiańskich, Uniwersytet
Komeńskiego w Bratysławie

Neologizmy leksykalne w języku bułgarskim i słowackim w kontekście integracji
europejskiej
Doc. dr Veličko Panajotov, Uniwersytet im. St. Kliment Ohridski w Sofii

Język polski w publicznym dyskursie o UE: dobrze czy źle, zawsze gorąco. Jak
Polacy mówią o Unii Europejskiej?
Mgr Tomasz Łysakowski, koordynator projektu EU Kids Online I

Języki słowiańskie i technologie informacyjno-komunikacyjne

Semantyka i lingwistyka konfrontacyjna (na podstawie języka polskiego i bułgar-
skiego)
Prof. dr hab. Violetta Koseska, Instytut Slawistyki, Polska Akademia Nauk

Tłumaczenia automatyczne w zakresie języków europejskich
Prof. dr Jan Hajič, Instytut Lingwistyki Formalnej i Stosowanej, Uniwersytet Ka-
rola w Pradze

Zasoby cyfrowe w języku bułgarskim w kontekście wielojęzyczności
Prof. dr Ludmila Dimitrova, Instytut Matematyki i Informatyki, Bułgarska Aka-
demia Nauk

Technologie informacyjno-komunikacyjne w nauczaniu języków na uniwersytecie
Dr Martin Mačura, Wydział Studiów Brytyjskich i Amerykańskich, Uniwersytet
Konstantego Filozofa w Nitrze

Konkluzje i zamknięcie konferencji

Carole Ory, kierownik Wydziału ds. Koordynacji Biur Terenowych DGT i Współ-
pracy z Przedstawicielstwami, Dyrekcja Generalna ds. Tłumaczeń Pisemnych
Filip Majcen, dyrektor Dyrekcji D w Dyrekcji Generalnej ds. Tłumaczeń Pisem-
nych.

References

Ajdukiewicz K. (1974). *Logika pragmatyczna.* Warszawa: Państwowe Wydawnictwo
Naukowe.

Barwise I, & Cooper R. (1981). Generalized Quantifiers and Natural Language. *Linguis-
tics and Philosophy*, *4*(2), 159–219. doi: 10.1007/BF00350139.

Barwise J., & Perry J. (1983). *Situations and Attitudes.* Cambridge, Mass.: Bradford
Books, MIT.

Bellert I. (1971). On the Use of Linguistic Quantifying Operators. *Poetics*, *2*, 71–86.
doi: 10.1016/0304-422X(71)90011-8.

BPCG (1988–2007). *Gramatyka konfrontatywna bułgarsko-polska.* Sofia – Warszawa.

Cooper, R. (1996). The Role of Situations and Generalized Quantifiers. In Shalom Lappin
(Ed.), *The Handbook of Contemporary Semantic Theory.* Oxford: Blackwell Refer-
ence.

Desclés J. P. (1999). Quantification, types, preuves et logique combinatoire. In *Studia
Kognitywne / Études cognitives* (Vol. 3: Quantification, Temps, Aspects, pp. 13–82).
Varsovie.

Dimitrova, L., & Koseska-Toszewa V. (2008). The Significance of Entry Classifiers in Digital Dictionaries. In *Lexicographic tools and techniques, Mondilex first open workshop, Moscow, Russia, 3–4 October, 2008, Proceedings* (pp. 89–97). Moscow.

Dimitrova, L. Koseska-Toszewa, V., Roszko, D., & Roszko, R. (2009). Bulgarian-Polish-Lithuanian Corpus — Problems of Development and Annotation. In T. Erjavec (Ed.), *Research Infrastructure for Digital Lexicography. MONDILEX Fifth Open Workshop. Ljubljana, Slovenia, October 14–15, 2009, Proceedings of the 12th International Multiconference Information Society 2009* (pp. 72–86). Ljubljana: Department of Knowledge Technologies, Jožef Stefan Institute.

Dimitrova, L., Koseska-Toszewa, V., & Satoła-Staśkowiak, J. (2009). Towards a Unification of the Classifiers in Dictionary Entries. In R. Garabík (Ed.), *Metalanguage and Encoding Scheme Design for Digital Lexicography. MONDILEX Third Open Workshop, Bratislava, Slovakia, 15–16 April, 2009. Proceedings* (pp. 48–58). Bratislava: L'. Štúr Institute of Linguistics. Slovak Academy of Sciences.

Frege, G (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik, NF 100*, 25–50.

Gramatyka (1984). *Gramatyka współczesnego języka polskiego* (Vol. 1–2).

Grzegorczyk, R. (1972). Wykładniki kwantyfikacji w języku polskim. In *Z polskich studiów slawistycznych*. Warszawa.

Grzegorczyk, R. (1976). Aktualizacja wypowiedzi poprzez kwantyfikację argumentów i predykatu. In *Otázky Slovanské syntaxe* (pp. 195–200). Brno.

Ivanchev, S. (1971). *Problemi na aspektualnostta v slavianskite ezitsi*. Sofiia: BAN.

Koseska, V., & Mazurkiewicz, A. (2009). Future and Possibility. In T. Erjavec (Ed.), *Research Infrastructure for Digital Lexicography. MONDILEX Fifth Open Workshop. Ljubljana, Slovenia, October 14–15, 2009, Proceedings of the 12th International Multiconference Information Society 2009* (pp. 87–98). Ljubljana: Department of Knowledge Technologies, Jožef Stefan Institute.

Koseska, V., & Mazurkiewicz, A. (2010). *Time flow and tenses*. Warsaw: SOW.

Koseska-Toszewa, V. (1977). *System temporalny gwar bułgarskich na tle języka literackiego*. Wrocław: Ossolineum.

Koseska-Toszewa, V. (1982). *Semantyczne aspekty kategorii określoności / nieokreśloności (na materiale z języka bułgarskiego, polskiego i rosyjskiego)*. Wrocław: Ossolineum.

Koseska-Toszewa, V. (1985). Jak rozgraniczyć użycie aorystu od niedokonanych od użycia imperfectum od niedokonanych? *Studia z Filologii Polskiej i Słowiańskiej, 23*, 195--201.

Koseska-Toszewa, V. (1995). O aspekcie i czasie w ujęciu kognitywnym. *Biuletyn PTJ, LI*, 25–31.

Koseska-Toszewa, V. (1997). Kvantifikaciya i vid. In *Semantika i struktura slavyanskogo vida* (Vol. II, pp. 141–147). Kraków.

Koseska-Toszewa, V. (2006). *Gramatyka konfrontatywna bułgarsko-polska* (Vol. 7: Semantyczna kategoria czasu). Warszawa: SOW.

Koseska-Toszewa, V. (2013) About Certain Semantic Annotation in Parallel Corpora. *Cognitive Studies / Études cognitives, 13*, 67–76. `doi: 10.11649/cs.2013.004`.

Koseska-Tosheva, V., & Gargov, G. (1990). *Bŭlgarsko-polska sŭpostavitelna gramatika* (Vol. 2, Semantichnata kategoriia opredelenost / neopredelenost). Sofiia: BAN.

Koseska-Toszewa, V., Korytkowska, M., & Roszko, R. (2007). *Polsko-bułgarska gramatyka konfrontatywna*. Warszawa: Wydawnictwo Akademickie Dialog.

Koseska-Toszewa, V., Korytkowska M., & Roszko, R. (2009). Contrastive Studies and Semantic Interlanguage. *Cognitive Studies / Études Cognitives, 9*, 33–52.

Koseska-Toszewa, V., & Mazurkiewicz A. (1988). Net Representation of Sentences in Natural Languages. In *Lecture Notes in Computer Science 340. Advances in Petri Nets* (pp. 249–266). Berlin, Heidelberg: Sprinter–Verlag.

Koseska-Toszewa, V., & Mazurkiewicz, A. (2009). Net-based Description of Modality in Natural language (on the Example of Conditional Modality). In V. Shyrokov, L. Dimitrova (Eds.), *Organization and Development of Digital Lexical Resources. MONDILEX second Open Workshop, Kyiv, Ukraine, 2–4 February, 2009. Proceedings* (pp. 98–105). Kyiv: Ukrainian Lingua-Information Fund, National Academy of Sciences of Ukraine.

Laskowski, R. (1986). System temporalno-aspektowo-modalny języka polskiego a sieci Petriego (Uwagi do artykułu A. Mazurkiewicza). In *Studia gramatyczne bułgarsko--polskie*, vol. I, *Temporalność* (pp. 23–30), Wrocław.

Mazurkiewicz, A. (1986). Zdarzenia i stany: elementy temporalności. In *Studia gramatyczne bułgarsko-polskie* (Vol. I: Temporalność, pp. 7–21). Wrocław.

Petri, C. A. (1962). Fundamentals of the Theory of Asynchronious Information Flow. In *Proc. of IFIP'62 Congress*. Amsterdam: North Holland Publ. Comp.

Piasecki, M. (2013). Self-organising Logic of Structures as a Basis for a Dependency-based Dynamic Semantics Model, *Cognitive Studies | Études cognitives*, *13*, 25–55. doi: 10.11649/cs.2013.002.

Polański, K. (1999). *Encyklopedia języka polskiego*. Wrocław: Ossolineum.

Rasiowa, H. (1975). *Wstęp do matematyki współczesnej*. Warszawa: Państwowe Wydawnictwo Naukowe.

Reichenbach, H. (1947). *Elements of Symbolic Logic*. New York: Free Press.

Reichenbach, H. (1967). Analiza języka potocznego. In *Logika i język*. Warszawa: Państwowe Wydawnictwo Naukowe.

Roszko, D. (2014). *Zagadnienia kwantyfikacyjne i modalne w litewskiej gwarze puńskiej (na tle literackich języków polskiego i litewskiego)*. Warszawa: SOW.

Roszko, R. (2004). *Semantyczna kategoria określoności/nieokreśloności w języku litewskim (w zestawieniu z językiem polskim)*. Warszawa: SOW.

Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. London: Allen and Unwin.

Russell, B. (1967). Denotowanie, Deskrypcje. In *Logika i język* (pp. 259–293). Warszawa: Państwowe Wydawnictwo Naukowe.

Russell, B. (1970). *Mój rozwój filozoficzny*. Warszawa: PWN.

Selinker, L. (1972). Interlanguage, *Iral*, *10*(3), 209–213.

Studia (1984) Projekt gramatyki konfrontatywnej bułgarsko-polskiej i serbsko-chorwacko--polskiej. In K. Polański (Ed.), *Studia polsko-południowosłowiańskie*, Warszawa.

Zaimov, Ĭ. (1982). *Suprasŭlski ili Retkov sbornik. Uvod i komentar na starobŭlgarskiia tekst*. Sofiia: BAN.