COGNITIVE STUDIES | ÉTUDES COGNITIVES, 9
SOW Publishing House, Warsaw 2009

LUDMILA DIMITROVA[1]
VIOLETTA KOSESKA-TOSZEWA[2]
[1]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2]Institute of Slavic Studies, Polish Academy of Sciences, Warsaw, Poland

CLASSIFIERS AND DIGITAL DICTIONARIES

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.

1 Introduction: Basic advantages of the digital vs paper dictionary

Information technologies offer great possibilities to linguists and lexicographers for the development of various dictionaries, especially for bi- and multilingual digital dictionaries. The following remarks are based on our experience from the development of the first Bulgarian-Polish Digital Dictionaries.

First let us mention briefly the basic advantages of the digital vs paper dictionary. The preparation of the paper dictionary is a continuous process (it takes several months or even years) and the dictionary remains unchangeable after publication, i.e. the paper dictionary is a static collection of dictionary entries. The creation of a digital dictionary is also a continuous process in time, but the collection of words

can be continuously expanded. New dictionary entries can be added or their content can be enriched by addition of supplementary information about the headword (grammatical, etymological), of examples (for clarification of usage), of phrases and combinations, etc. The digital dictionary is a dynamic collection of dictionary entries, which provides a dynamical structure of the dictionary entry per se. This characteristic allows:

- a relatively easy adaptation of the lexical database, which the collection of words in a dictionary actually is, to a new (improved) model of dictionary entry and its enrichment with new information, for example the addition of the word-forming group of the headword, etc.
- a refinement of the system of classifiers, used for structuring the dictionary entry in order to describe optimally the headword.
- use of the digitally-presented information for the creation of a new (or different type of) digital dictionary, for example two monolingual digital dictionaries (explanatory or terminological) in two different languages can be used to produce a new bilingual dictionary (although in practice that is non-trivial);
- when necessary – last but not least – correction of various mistakes.

2 Problems and challenges

One of the main problems of the development of digital dictionaries is the *choice of classifiers* of the dictionary entry. Whenever the development of a system of bilingual digital dictionaries, serving as a basis for a system of multi-linigual dictionaries in perspective, is concerned, there arises an issue of *unification of the classifiers* in the dictionary entry. This is an *issue of harmonisation of the classifiers for various languages*, whose solution has to present a *unified selection of classifiers and a standard form of their presentation.* In a broader sense the issue of unification of classifiers in the dictionary entry *approaches the issue of a new part-of-speech classification* keeping in mind the specifications of a digital dictionary.

3 Classifiers

It is accepted that classifiers carry different morphosyntactic and/or semantic characteristics of the words (in particular, the dictionary entry). They split the set of words according to properties. Most often the classifier connects the word with its respective part of speech, depending on the class, to which the word belongs. But the classifier can show specific features of the word, such as gender, number, tense, etc. Tense is a meaning of the form, but has not been fully defined, see the examples about aorist (*аорист* in Bulgarian) and imperfectum (*имперфект*).

At the current stage of research the part-of-speech classification in a natural language continues to be under discussion because it is not consecutive. It is based on different criteria (morphological, syntactic or "narrow" semantic) which are reduced only to the separation of grammatical categories. Thus the part-of-speech classification is different not only depending on language but is also significantly

different in certain languages. This fact made us consider the unification of the part-of-speech classification at least in the two Slavic languages in our study, see [15]. In order to accept a common for these languages, i.e. a standard type of part-of-speech classification we start a discussion on these issues in this article. At the same time we offer new arguments on this issue on Bulgarian and Polish material using F. Slawski's *Bulgarian-Polish Dictionary* [17] as well as examples from machine translation from English to Polish and from English to Bulgarian.

So far the meaning of the forms has been the Achilles' heel of the description, dictionaries and corpora, both mono- and bilingual. That is why we shall focus our attention on some entries in the Bulgarian-Polish Dictionary depending on the form's meaning and its differentiation from a given meaning.

Examples

Let us have a look at the following examples of dictionary entries which do not explain anything in the dictionary. It is not clear whether they concern form or meaning. Neither is it clear what the meaning of this form is.

Example 1. Entry with headword "aorist"

аòрист, -и *m gram. aoryst m*

This entry with headword the verbal form "aorist" does not make clear what kind of aorist is meant. In Bulgarian aorist can be formed from perfective and imperfective verbs, for instance, *написа* and *писа*. In the sentence *Той написа интересна книга.* the form *написа* is a perfective aorist. But the form *писа* in *Той писа тази книга 5 години.* is an imperfective aorist.

Perfective aorist determines an event that has happened before the state of speaking and reserves a place for a unique quantifier in the sentence's semantic structure [11], [13].

Imperfective aorist means a configuration of states and events that have happened before the state of speaking and reserves a place only for a unique quantifier in in the sentence's semantic structure [11], [13], [15].

In order to describe the two different meanings of aorist we suggest the following two new dictionary entries:

аòрист от свършен вид, -и *m gram.* – единично събитие настъпило преди състоянието на изказването. (A unique event that has happened before the state of speaking.)

This meaning is conveyed by Polish perfective praeteritum [11]. For example:

Той боледува от грип.
On chorował na grypę.

аòрист от несвършен вид, -и *m gram.* – единично квантифицирана конфигурация от състояния и събития, извършваща се преди състоянието на изказването. (A unique-quantified configuration of states and events that have happened before the state of speaking.)

This Bulgarian meaning is conveyed by Polish imperfective praeteritum [11]. For example:

В четвъртък ходих пеша до центъра на града.
W czwartek chodziłam pieszo do centrum miasta.

Example 2. Entry with headword "imperfect":
Имперфект *m gram. Imperfectum n*

Just as in the case of aorist, we have no information that in Bulgarian this form (if form is meant here) is formed from imperfective as well as perfective verbs. We have no information about the difference in the meaning of the two. The imperfective imperfect serves to determine configurations of states and events that have happened and lasted before the state of speaking. The form here in contrast to the imperfective aorist (which is connected with a unique quantifier), reserves a place for all quantifiers (existential, universal, although rare, unique) [16]. In this case our suggestion about the new entry with headword "imperfective imperfect" is the following:

Имперфект от несвършен вид, -и, *m gram.* Многозначно квантифицирана конфигурация от състояния и събития, настъпили и траещи преди състоянието на изказването — по значение съответства полската форма praeteritum от несвършен вид. (Multiply-quantified configuration of states and events that have happened and lasted before the state of speaking – by meaning corresponds to Polish imperfective praeterium.)

Той понякога намираше време за разходка.
On od czasu do czasu znajdował czas na spacer.
Той понякога боледуваше от грип.
On czasem chorował na grypę. (See [15])

Concerning the alternative "имперфект от свършен вид" (perfective imperfect) we must note that it occurs very rarely and only in special modal, conditional contexts, such as: *Пийнеше ли* (perfective imperfect), *вдигаше* (imperfective imperfect) *много шум около себе си.*

Example 3.
Let us consider the entry:

минал *part. adi* przeszły, zeszły, ubiegły; **миналата година** dwa lata temu; **-о време** *gram.* Czas przeszły.

Here we have another type of problems. There are three Polish forms *przeszły*, *zeszły*, *ubiegły* that correspond to the Bulgarian form *минал* ('past'). As in the case of aorist and imperfect it is not clear what is meant — meaning or form of past tense.

If a meaning is meant, it is not clear what past tense is meant. If however a form is meant, it must be mentioned that this is a form with multiple meanings.

We already mentioned ([5], [6]) that a single form can have multiple meanings and they naturally vary in number across the various languages. This is a problem whose solution would allow the creation of a new $\mathbf{L}_2$-$\mathbf{L}_1$ dictionary from a $\mathbf{L}_1$-$\mathbf{L}_2$ dictionary. How do we invert a Bulgarian-Polish dictionary entry so that it represents a Polish-Bulgarian dictionary entry? It is obvious that the elimination

of shortcomings among the entries of a given $\mathbf{L}_1$- $\mathbf{L}_2$ bilingual dictionary, eliminating the impossibility of a new ordering of information with the scope of obtaining an inverted $\mathbf{L}_2$-$\mathbf{L}_1$ bilingual dictionary, requires a reconsideration of the representation of the relation "form-meaning" in the dictionary.

An automated inversion of the dictionary is possible and easy to implement only when the relation "form-form" is considered. But then the inverted dictionary is quite poor and its cognitive value quite weak.

In order to keep all different meanings we suggest for discussion the option where each meaning is shown with the same form but enumerated, for example:

минал (1) – przeszły
минал (2) – zeszły
минал (3) – ubiegły

In other words the form is indexed and appears in the list as many times as its different meanings.

Another example from the Bulgarian-Polish Dictionary – the dictionary entry for headword *май*:

май (1) *m* maj; първи май — pierwszy maja
май (2) *adv.* Chyba, prawie, zdaje się, prawdopodobnie

Maybe in this case it is necessary to list this form a third time so that its third Polish meaning *prawie* corresponding to Bulgarian *почти* ('almost') is listed as well.

май (3) *adv.* prawie

A short look at the Explanatory Dictionary of Bulgarian [2] shows us the following two ways to describe homonymy.

1. when the forms are different parts of speech, the difference in meaning is shown by indexing the different meanings
   малко^1 *нареч.* ...в ограничено или недостатъчно количество...
   малко^2 *ср.* Наскоро родено или излюпено същество...
   or it is implied by listing the respective part of speech.
   май *м.* Петият месец на годината...
   май *част.* За изразяване на предположение....
2. when the forms belong to the same class, the different meanings are indexed
   мина^1 *жс.* ... рудник
   мина^2 *жс.* ... снаряд
   мина^3 *жс.* ... израз на лицето

The usage of indexing for each meaning of a form (as in the above examples (2)) would allow the Bulgarian-Polish dictionary to be "inverted" and thus to obtain automatically a Polish-Bulgarian digital dictionary. Whenever a bilingual digital dictionary is being complied, in the beginning the most common words/forms (parts of speech) are selected in a given digital corpus of $\mathbf{L}_1$ language. Then this frequency dictionary is completed with the translated correspondences from $\mathbf{L}_2$ language. We must mention here that besides frequency the forms may be selected according to a certain topic which contains them and which they describe. In other words the dictionary may be compiled according to topics (something

like topic and frequency). 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 first Bulgarian-Polish digital dictionaries are being developed. The experimental version of the Bulgarian–Polish electronic dictionary is prepared in WORD-format and consist approximately 20 thousand dictionary entries. This 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 [9]. We remark here that the suggested LDB structure of Bulgarian-Polish dictionary entry is suitable for automated generation of a Polish-Bulgarian dictionary entry. 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.

4 Some comparative remarks on the classifiers of the verbs

The comparison of the Bulgarian and Polish material [7] requires an explanation, which is important for the part-of-speech classifiers in the dictionary entries of the cited bilingual electronic dictionary. It is a common practice to list as a headword in the dictionary entries the infinitive of the verb. In Bulgarian the infinitive has disappeared and has been functionally replaced by the *да*-construction, which connects the particle *да* to the present tense forms. In this respect Bulgarian is more similar to other Balkan languages (Modern Greek, for example), but differs from Polish where the infinitive is preserved. This is an important example for the requirement of distinguishing a form from its function and meaning. The present tense form in this case does not have "present tense"-meaning. In the Bulgarian verb entries it is accepted to list as headword the 1st person singular form of the present tense.

One of the important classifiers of the verbal form which must be included in the dictionary entry refers to the transitivity or intransitivity of the verb. In our opinion the tendency of including more classifiers in the dictionary entry which we consistently follow, leads us to confirm the necessity of a classifier reflecting transitivity or intransitivity of the verb [8]. It is a different question what this classifier should reflect. According to the tradition in the older Bulgarian and Polish grammars, transitivity and intransitivity used to be considered as a phenomenon related to the voice of the verb (active or passive). In Polish and Bulgarian the verbs which form the passive participles are called transitive. They stand in contrast to the intransitive verbs which do not form such participles. A fact which we must stress here is that the Polish transitive verbs are always followed by the accusative case of nouns or adjectives. This fact is important for the comparison of the dictionary entries in Polish and Bulgarian, because Bulgarian lacks a nominal declination, while Polish is a typical synthetic language. The classifier "aspect" of a verb is universally accepted. However we must stress also that the "aspect" classifier is obligatory in the dictionary entry for a Slavic language. The aspect in Slavic languages is a well-formed grammatical category whose meaning expresses events — perfective aspect, and states — imperfective aspect, where we interpret "event" and "state" as described in the net description of temporality in a natural language

[14], [16]. On aspect and the problems of its classification see [12] for an overview of the different interpretation of aspect in the linguistic schools and the treatment of this category as word-forming, morphological, lexico-grammatical, grammatical and semantical. We must stress that the connection of the "aspect" category to temporality depends on the interpretation of "aspect" category. If we assume that "aspect" is a semantic category, the question about its relation to the semantic category "temporality" is inevitable. According to some linguists, "aspect cannot be treated separately from tense" [10], according to others the tenses are meanings independent from the meaning of the "aspect" of the verbal form [1]. Based on Bulgarian language material we see how important are the aspectual-temporal relation in the language. This leads us to the conclusion that the forms and meanings of time, especially with respect to Bulgarian, are a key problem that must affect the dictionary entry in every bilingual dictionary, which contains Bulgarian. It must be stressed that the Bulgarian language differs typologically from the other Slavic languages. It is an analytic language, and not synthetic (like the rest of the Slavic languages), has no cases in its nominal system (except some vestiges of vocative), but has many tense forms as well as well-formed category "aspect". In this respect Bulgarian resembles a lot more English or Romance languages (French or Italian) than the other Slavic languages. In other words, the "aspect" problem opens the question about the "temporal" classifier in the dictionary entry: whether to include a "temporal" classifier and how to present it. This question must be answered in more detail later.

5 Suggestions

- Our suggestions can be grouped around the mode of form classification and the mode of writing the meanings of verb tense forms (two types with exact definition that can be "translated" in a formal language, for example, Petri nets). We take a step back so to say from the "form-meaning" principle and limit ourselves to the "form-form" principle in bilingual dictionaries.
- We suggest the headword form in the dictionary entry of the digital dictionary to be indexed according to the number of meanings, and each different meaning to be related unambiguously to the form. In this manner most meanings of the form can be encompassed. Such a description might require more classifiers but it is obvious that the greater number of classifiers provides a more adequate translation correspondence.
- We plan to use the CONCEDE model for dictionary encoding that respects the guidelines of the Text Encoding Initiative (TEI) Dictionary Working Group. The CONCEDE project [18], supported by the EC under INCO-Copernicus program, developed a formal model for LDBs. The LDBs using a common tagset for the six Central and East European languages: Bulgarian, Czech, Estonian, Hungarian, Romanian, and Slovene were developed in accordance with the guidelines of the TEI Dictionary Working Group. In the framework of the project the first LDB for Bulgarian, based on encoding standards established by the TEI, was developed.

6 Bulgarian experience

Traditional grammatical classifications for Bulgarian

Traditional Bulgarian grammar for instance recognizes three main grammatical classifications:

- Semantic-grammatical – depending on the most general common meaning and on the grammatical properties words are ordered in classes, called parts of speech:
  - Nouns (a general terminological meaning of objects with common grammatical categories – gender, number, definiteness/indefiniteness),
  - Adjectives (have something in common in their lexical meaning, which is "indication, property, quality" of an object,
  - Verbs (common lexical meaning is "action or state" of a person/objects with common grammatical categories "tense", "person", "number", "mood", "voice"),
  - Numeral,
  - Pronouns,
  - Adverbs
  - Prepositions,
  - Conjunctions,
  - Interjections,
  - Particles,
- Morphological classification – according to the criterion "Open-class words or closed-class words":
  - Open class words are nouns, adjectives, numerals, pronouns and verbs,
  - Closed class words are adverbs, prepositions, conjunctions, interjections and particles.
- Syntactic (functional) classification – depending on whether the word functions independently in the sentence or not:
  - Independent are nouns, adjectives, numerals, pronouns, verbs, and adverbs,
  - Dependent are prepositions, conjunctions, and particles. The interjections are excluded.

Lexical specifications for Bulgarian in MULTEXT-East

The semantic-grammatical classification of the Bulgarian wordforms was used during the development of lexical specifications for the Bulgarian language in the EC project MULTEXT-East [3], [4]. In the MULTEXT-East project multilingual parallel (Orwell's 1984) and comparable (fiction and newspapers) corpora for six East-European languages - Bulgarian, Czech, Estonian, Hungarian, Romanian, Slovene - were developed and a lexicon was complied for each corpus and language.

The lexicons have been prepared in the form of lexical lists where each line contains one entry in the following form:

word-form <tab> lemma <tab> morphosyntactic description

Morphosyntactic description (MSD) contains encoding lexical specifications of the corresponding word-form ("word-form" represents an inflected form of the lemma). When the the wordform (inflected form) coincides with its main form (lemma), then the entry "lemma" is replaced by "=".

The MULTEXT-East project has provided harmonised lexical specifications for the six East-European MTE languages and English. The specifications are presented as sets of attribute-values, with their corresponding codes used to mark them in the lexicons. The core features were determined (these features are shared by the most of the languages) and this provided the comparability of the information encoded in the lexicons across the MULTEXT-East languages. Except these "general properties" so-called language-specific features were defined, which describe language-specific morphosyntactic phenomena.

Bulgarian MSD

Here we shall briefly present the Bulgarian wordform MSD because these can provide useful information about digital bilingual Bulgarian-lang2 (digital bilingual dictionaries with Bulgarian language) as possible classifiers in the dictionary entry in regard to applications of digital dictionaries in machine translation systems, e-learning, etc.

MSD is defined as a linear string of symbols, representing the morphosyntactic descriptions, the positions of a string are numbered 0, 1, 2, etc. in the following way:

- the symbol at position 0 encodes part of speech;
- each symbol at position 1, 2, n, encodes the value of one attribute (person, gender, number, etc.);
- if an attribute does not apply, the position in the string contains a hyphen "-".

Some examples of Bulgarian MSDs:

барабан	=	Ncms-n	(Noun, common, masculine, singular, no-definit)
барабани	барабан	Ncmp-n	(Noun, common, masculine, plural, no-definit)
барабани	барабаня	Vmia2s	(Verb, main, indicative, aorist, 2^{nd} person, singular)
барабани	барабаня	Vmia3s	(Verb, main, indicative, aorist, 3^{rd} person, singular)
барабани	барабаня	Vmip3s	(Verb, main, indicative, present, 3^{rd} person, singul)
барабани	барабаня	Vmm-2s	(Verb, main, imperative, 2^{nd} person, singular)

май =	Ncms-n	(Noun, common, masculine, singular, no-definiteness)
май =	Qgs	(Particle, general, simple)
май мая	Vmm-2s	(Verb, main, imperative, 2^2 person, singular)

мина =	Ncfs-n	(Noun, common, feminine, singular, no-definiteness)
мина =	Ncft	(Noun, common, feminine, count)

малки	малко	Ncnp-n	(Noun, common, neutral, plural, no-definiteness)
малки	малък	A---p-n	(Adjective, plural, no-definiteness)
малките	малко	Ncnp-y	(Noun, common, neutral, plural, yes full_article)
малките	малък	A---p-y	(Adjective, plural, yes full_article)

7 Examples of machine translation

Let us have a look at some examples of machine translation, randomly picked from a web-page with an original text in the English language, which offers translation to Bulgarian, Polish and other languages. The lack of morphosyntactic descriptions that contain encoded (according to the standard) lexical specifications of the word-forms and the lack of adequate classifiers (or any classifiers) in the database (or in the digital dictionaries), used in the machine translation system, leads to the following translation mismatches:

First example

Original English text:
His play/direct partnership with the Scottish Chamber Orchestra has been particularly fruitful, and as well as touring extensively with the orchestra *he has recorded a disc featuring Mozart's G major and D minor piano concertos*.

Machine translation in Bulgarian:
Неговата игра/преки партньорство с шотландски камерен ансамбъл е било особено ползотворно, а както и още по обстойно с оркестър *той е записано диск, с участието на Моцарт* G *големи и малки* D *пиано concertos*.

Comment:
(For the sake of comparison — English translation (as far as it is possible) of the Bulgarian text:
His game/direct partnership with a Scottish Chamber Orchestra has been particularly beneficial, and as well as more extensively with an orchestra *he was recorded a disc with the participation of Mozart G major and minor D piano concertos*.)

The errors in the machine translation of the sentences in the examples can be grouped as follows:
first, wrong choice of lexical meaning for the translation:
play = изпълнение ↔ игра = гаме
direct = ръководи, дирижира ↔ пряк = direct, straight; immediate
fruitful = плодотворно ↔ ползотворно = beneficial
featuring = включвайки ↔ участието ı тхе партиципатион
second, **lack of concordance between pronoun (as subject) and the verb form in the sentence**:
he | той (pronoun, **masculine)**
recorded | записано (participle, **neutral**).

Machine translation in Polish:
Jego *grać / bezpośrednej współpracy* ze Scottish Chamber Orkiestra była szczególnie owocna, jak również szerokie tournee z orkiestrą *ma zapisane dysk* zawierający Mozarta G- dur i d – moll koncerty fortepianowe.

Comment:
The errors in this sentence are:
play is translated as a verb infinitive due to lack of classifiers, in this case the English *play* is a noun, not a verb.

ma zapisane — rodzaj nijaki is related to *dysk* — rodzaj męski the participle *zapisane* is neutrum and is not in accordance with the masculine noun *dysk*.

Second example

Original English text:
Piotr Anderszewski was born in Warsaw to *Polish-Hungarian parents*.

Machine translation in Bulgarian:
Пьотр Anderszewski е роден във Варшава с полския-унгарски родители.
(For the sake of comparison – English translation (as far as it is possible) of the Bulgarian text: Piotr Anderszewski was born in Warsaw with the Polish-Hungarian parents.)

Comment:
Lack of concordance between qualifier and *word that it qualify* in the translation of *Polish-Hungarian parents* | *с полския-унгарски родители*.

Machine translation in Polish:
Anderszewski urodził się *w Warszawa – Węgier do Polski rodziców*.

Comment:
The error here is triggered by the preposition *to*, to which only one meaning is given (from... Hungary to Poland). The English phrase *Polish-Hungarian parents* is not quite logical. Rather it should say "parts of Polish and Hungarian origin" or "Hungarian mother and Polish father".
Furthermore, *Warszawa* instead of *Warszawie* — lack of casus locativus form.
The errors in this sentence are:
"play" is translated as a verb infinitive due to lack of classifiers, in this case the English "play" is a noun, not a verb.

Third example

Original English text:
An exclusive artist with Virgin Classics since 2000, Anderszewski's first disc on the Virgin label was Beethoven's Diabeli Variations, a work which had already fascinated him for a decade.An exclusive artist with Virgin Classics since 2000, Anderszewski's first disc on the Virgin label was Beethoven's, a work which had already fascinated him for a decade.

Machine translation in Bugarian:
Един изключителен артист с Вирджински класика от 2000 г. насам, Anderszewski първия диск на Богородица етикет е на Бетовен Diabeli варианти за работа, която вече е очарован му за едно десетилетие.

Comment:
(For the sake of comparison – English translation (as far as it is possible) of the Bulgarian text: One exceptional artist with Virginia classic since 2000, Anderszewski first disc of Virgin Mary label is of Beethoven Diabeli work versions, which is already fascinated to him for a decade.) His strong identification with this work went

on to become the subject of a film by Bruno Monsaingeon (creator of documentaries on Sviatoslav Richter and Glenn Gould).

Machine translation in Polish:
Artysta na wyłączność z Virgin Classics od 2000 roku, Anderszewski pierwszy dysk na etykiecie Dziewicy było Beethovena Diabellego wariacje na pracę, która fascynowała go już od dekady.

Comment:
1. Casus genetivus for Anderszewski in the sentence is missing
2. "było" is neutrum and is not in accordance with the masculine noun "dysk".
3. "work" is translated as a "pracę", the right translation is "tvorba"
4. in the phrase "wariacje na pracę, która fascynowała go" predicate is missing, correct: "jest to dzieło, albo jest to utwór, który go fascynował..."

Fourth example

Original English text:
The 2008-09 season will see Anderszewski *giving recitals at* (*as points at*) Carnegie Hall, *Chicago's Symphony Center* (*Chicago of the Symphony Center*), the Walt Disney Concert Hall in Los Angeles and the Royal Festival Hall, London.

Machine translation in Bugarian:
В сезон 2008-09 ще видите Anderszewski *като точки в* (*as points at*) Карнеги Хол, *Чикаго на Симфония център* (*Chicago of the Symphony Center*), Уолт Дисни Концертната зала в Лос Анджелис и Роял Фестивал Хол, Лондон.

Comment:
For the sake of comparison – English translation (as far as it is possible) of the Bulgarian text: During the 2008–09 season you will see Anderszewski as points at Carnegie Hall, Chicago of the Symphony Center, the Walt Disney Concert Hall in Los Angeles and the Royal Festival Hall, London.

Machine translation in Polish:
W sezonie 2008–09 *będzie zobaczyc* Anderszewski podając motywów w Carnegie Hall, Chicago of the Symphony Center, the Walt Disney Concert Hall in Los Angeles and the Royal Festival Hall, London.

Comment:
In the Polish translation "można" is missing, correct *można będzie zobaczyć*.
"Motyv" is not correctly translated, "recital" is meant instead.

Fifth example

Original English text:
Currently he lives in Paris and Lisbon.

Machine translation in Bugarian:
Currently he lives in Paris and Lisbon.
В момента той живее в Париж и Лисабон.
"Successful" translation correspondences.

Machine translation in Polish:
Aktualnie mieszka w Paryzu i *Lisbona* (correct *w Lizbonie*).

Comment:
Casus locativus for *Lisbon* in the sentence is also missing.

Briefly, in Polish we observe the following mistakes:

- wrong gender,
- lack of cases,
- incorrect translation of tenses – see above the lack of "można",
- incorrectly translated prepositions,
- incorrect translation of lexical meanings (motyv — recital).

There is not a single correctly translated sentence in the Polish text, in contrast to Bulgarian, but that is due to the analytical character of English and Bulgarian, whereas the Polish cases pose an additional difficulty to the translation software.

8 Conclusion

The dictionary entry classifiers must reflect the specifics of the compared languages, for example the transitivity/intransitivity classifier is important for the syntax of both languages, but is much more important on the morphologic-syntactic level for Polish, a synthetic language, in contrast to Bulgarian, an analytic language. As mentioned before, the Polish transitive verbs require an accusative case for their object.

We must also distinguish between forms and the meanings of the forms in the dictionary entries. In traditional grammatical descriptions this distinction is missing, which creates intolerable errors in the description of the respective language. This is especially important for the aspect characteristic of the verbs in Slavic languages, where the category "aspect" is not only semantic but also grammatical.

We must stress again that we should not fear the greater quantity of dictionary entry classifiers in the electronic dictionary. On the contrary, this is an advantage of the electronic over the printed dictionary. The increase of the number of classifiers of the headwords in the entry will make machine translation more adequate and enrich electronic dictionaries. A dictionary with more classifiers will be significantly more useful to the user. We believe that it is necessary to establish a possibility to obtain the inverse dictionary automatically. With traditional bilingual dictionaries this is impossible because of the polysemy of forms. Using the contemporary process theory (Petri nets theory) we suggest that dictionary entries related to time in a natural language render the content as well as the form. The content must reflect the main elements of time: the event, the state and the configuration of events and states (see above *Example 1 and 2*; [16]).

Bibliography

[1] Андрейчин, Л. (1944). *Основна българска граматика.* София. (In Bulgarian).

[2] Andreichin, L., Georgiev, L., Ilchev, St., Kostov, N., Lekov, I., Stoikov, St., Todorov, Tsv. (1997) Bulgarian Explanatory Dictionary. 4th revised edition, prepared by D. G. Popov. Nauka i Izkuvstvo Publishing House, Sofia. (In Bulgarian).

[3] Dimitrova, L. (1998). Lexical Resource Standards and Bulgarian Language. In: *International Journal Information Theories & Applications,* Vol. 5, No. 1, 27–34.

[4] Dimitrova, L., Erjavec, T., Ide, N., Kaalep, H.-J, Petkevic, V., Tufis, D. (1998). Multext-East: Parallel and Comparable Corpora and Lexicons for Six Central and Eastern European Languages. In: *Proceedings of COLING-ACL '98.* Montréal, Québec, Canada, 315–319.

[5] Dimitrova, L., Koseska-Toszewa, V. (2007). Digital Dictionaries — Problems and Features. In: *Proceedings of the Jubilee International Conference Mathematical and Computational Linguistics.* 6 July 2007, Sofia, Bulgaria, 25–34.

[6] Dimitrova, L., Koseska-Toszewa, V. (2008). Some Problems in Multilingual Digital Dictionaries. In: *International Journal Études Cognitives*, Vol. 8, SOW, Warszawa, 237–255.

[7] Dimitrova, L., Koseska-Toszewa, V. (2009). Bulgarian-Polish Corpus. In *International Journal Cognitive Studies – Études Cognitives*, Vol. 9, SOW, Warszawa, (In: this volume).

[8] 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.* Proceedings of the MONDILEX Third Open Workshop, Bratislava, Slovak Republic, 15–16 April 2009. 48–58.

[9] Dimitrova, L., Panova, R. Dutsova, R. (2009). Lexical Database of the Experimental Bulgarian-Polish online Dictionary. In: R. Garabík (Ed.), *Metalanguage and Encoding scheme Design for Digital Lexicography.* Proceedings of the MONDILEX Third Open Workshop, Bratislava, Slovak Republic, 15–16 April 2009. 2009, 36–47.

[10] Иванчев, С. (1971). *Проблеми на аспектуалността в славянските езици.* София. (In Bulgarian)

[11] Koseska-Toszewa, V. (2006). Bułgarsko–polska gramatyka konfrontatywna, t. VII. Semantyczna kategoria czasu. SOW, Warszawa.

[12] Koseska-Toszewa, V. (forthcoming). Form, its meaning, and dictionary entries.

[13] Koseska-Toszewa, V., Mazurkiewicz, A. (1988). *Net representation of sentences in natural languages,* Advances in Petri Nets, LNCS 340, Springer Verlag, 249–259.

[14] Koseska, 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.* Proceedings of the MONDILEX Second Open Workshop, Kiev, 2–3 February 2009. 98–105.

[15] Koseska-Toszewa, V., Roszko, R. (2008). Remarks on Classification of Parts of Speech and Classifiers in an Electronic Dictionary. In: L. Iomdin, L. Dimitrova (Eds.), *Lexicographic Tools and Techniques.* Proceedings of the MONDILEX First Open Workshop, Moscow, Russia, 3–4 October 2008, 80–88.

[16] Mazurkiewicz, A. (2008). A Formal Description of Temporality (Petri net approach). In: L. Iomdin, L. Dimitrova (Eds.). *Lexicographic Tools and Techniques*. Proceedings of the MONDILEX First Open Workshop, Moscow, Russia, 3–4 October 2008, 98–108.
[17] Sławski F., (1987). Podręczny słownik Bułgarsko-Polski z suplementem. Warszawa.
[18] CONCEDE: `http://www.itri.brighton.ac.uk/projects/concede/`