Towards a Unification of the Classifiers in Dictionary Entries ⋆

Ludmila Dimitrova[1], Violetta Koseska-Toszewa[2], Joanna Satoła-Staśkowiak[2]

[1] Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences, Sofia

[2] Institute of Slavic Studies,
Polish Academy of Sciences, Warsaw

Abstract. In this paper we continue the discussion of the important problems related to the unification of the classifiers in the electronic dictionary entries, started in [2]. We focus our attention especially to dictionary entries with Bulgarian verbs as headwords. We analyze some examples from ongoing experimental version of the Bulgarian–Polish online dictionary.

1 Introduction

The first Bulgarian–Polish electronic dictionary is being developed in the framework of the cooperation between the Polish and the Bulgarian Academies of Sciences – the joint research project "Semantics and Contrastive linguistics with a focus on a bilingual electronic dictionary". 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 will be an entry point to the relational database (RDB) of the Bulgarian-Polish online dictionary. The proposed 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.

2 Classifiers of the Dictionary Entry

As we already wrote [2], [3], one of the main problems of the development of digital dictionaries is the *choice of classifiers* in the dictionary entries. The development of a system of multilingual dictionaries on a basis of bilingual ones requires at first a *unification of the classifiers* in the dictionary entries. The problem turns to the *harmonisation of the classifiers* for various languages, and its solution has to present a *unified selection of classifiers and a standard form of their presentation*.

The comparison of the Bulgarian and Polish material requires an explanation, which is important for the part-of-speech classifiers in the dictionary entries of the cited bilingual electronic dictionary. In the current paper we will mainly analyze the verb entries in both languages.

2.1 Headword in the verb entry

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.

⋆ The study and preparation of these results have received funding from the EC's Seventh Framework Programme [FP7/2007-2013] under grant agreement 211938 MONDILEX

2.2 The phenomenon "transitivity-intransitivity"

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, makes us confirm the necessity of a classifier reflecting transitivity or intransitivity of the verb [2]. 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).

The authors of "Słownik gramatyczny języka polskiego" [12] propose to exclude the voice category from the explanation of the phenomenon "transitivity-intransitivity". They suggest transitivity and intransitivity to be treated as a syntactic phenomenon. They do not introduce the "voice" category in the description of Polish morphology. Without starting a discussion with them, we must stress that this verbal phenomenon is related to the well-known linguistic fact about the existence of passive participles such as the Polish "chwalony", Bulgarian "хвален" which are frequently used in Polish in nominal constructions, for example *Dziecko często chwalone ma dobre samopoczucie* (an example from the cited "Słownik gramatyczny języka polskiego"). In Bulgarian we have a similar phenomenon, for instance: *Често хваленото дете има добро самочуствие*. The paraphrases of both sentences look alike:

„*Dziecko często chwalone ma dobre samopoczucie*" // *Дете, което е често хвалено, има добро самочувствие*".

In Polish and Bulgarian the verbs which form such passive participles are called transitive. They stand in contrast to the intransitive verbs which do not form such participles, for example in Bulgarian one can say "*Майка му спи*", but there exists no participial **спана*, in Polish "*Matka śpi.*", yet a participial like **spana* is missing.

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 case system, while Polish is a typical synthetic language. It is interesting to note that there exists a third type of classification related to this phenomenon. The above-mentioned authors propose a new classifier (quasi-transitivity). This concerns verbs which are weakly connected to their participle, for instance, *uśmiechnąć się - uśmięchnięty* (in Bulgarian *усмихнат*). In Polish such participles can be formed also from intransitive verbs. That is why this group is called "quasi", for example *Dziewczynka uśmiechnęła się. Uśmiechnęta dziewczynka.* Quasi-transitive verbs exhibit a tendency of exceptions in the classification of transitive and intransitive verbs. If a criterion is introduced such as "in Polish a transitive verb is followed by nouns in accusative case without a preposition", it will verify and clear exceptions from the classification of transitive and intransitive verbs. After *uśmiechnęła się* in Polish there follows no accusative case without preposition. One can not say for example **Dziewczynka uśmiechnęła się kogoś, coś...*, the right sentence is: *Dziewczynka uśmiechnęła się do kogoś, z powodu czegoś...* For this purpose it suffices to place the transitive verbs into a group containing only those which are followed by nouns in accusative case without preposition, such as: *Anna chwali Jasia – Jaś jest chwalony przez Annę.(*Chwali kogo, co?) – Jasia – accusative, animate object, singular. The transitivity of the Polish verb shows that it is always followed by nouns in the accusative case without preposition [12]: 109.

2.3 The "aspect" classifier

The classifier "aspect" of a verb is universally accepted. However we must stress also that the "aspect" classifier in the dictionary entry for a Slavic language is obligatory. The aspect in Slavic languages is a well-formed grammatical category whose meaning boils down to the expression of events – by the perfective aspect – and states – by the imperfective aspect, where we interpret "event" and "state" as described in the net description of temporality in a natural language at the MONDILEX forum [11], [10]. On aspect and the problems of its classification see [8] (in this volume), 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*" [6], according to others the tenses are meanings independent from the meaning of the "aspect" of the verbal form [1].

In languages such as Polish, Czech, Slovak, Ukrainian and Russian, in which "aspect" is a strongly developed semantic and grammatical category, there are few tense forms. This is not the case in South Slavic languages, in which, for example, in Bulgarian, has a high number of tense forms as well as a strongly developed semantic and grammatical category "aspect". As we know, the languages which lack the grammatical category "aspect", such as Latin, French, Italian or Spanish, has a high number of tense forms. As mentioned in [8], there are two distinct tendencies in the South Slavic languages – the first towards reduction of tense forms (Croatian/Serbian), the second one towards reduction or extinction of the aspect. So it should happen in Bulgarian and Macedonian, but does not! The example about the development of the category "aspect" in Bulgarian considered here shows that the development of category "aspect" does not lead to a reduction of the tense forms. Furthermore, as shown by Koseska and Gargov in the second volume of the Bulgarian-Polish Contrastive Grammar, all aspectual-temporal combinations of the verbal form in Bulgarian differ in meaning and are not redundant [9].

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 five Slavic languages in the MONDILEX project. It is an analytic language, and not synthetic (like the rest of the Slavic languages), has not cases (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 five Slavic languages from the MONDILEX project.

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.

2.4. A few short remarks

(1) Gender and number must be specified for the nouns and adjectives because in the two languages these classifiers may vary. For example, the Bulgarian noun "стая" /*room*/ is feminine, while the Polish "pokój" /*room*/ is masculine.

(2) The problem about adverb classification requires a separate study. In the literature on adverbs there are no clear-cut criteria about this part-of-speech.

3 Bulgarian-Polish dictionary entries analysis

Here we give an overview of some dictionary entries from the future Bulgarian-Polish online dictionary. The dictionary entries are divided in two groups, the first containing entries whose headwords belong to the open parts of speech - verbs (incl. verbal forms, esp. Bulgarian participles), nouns, adjectives, adverbs, and the second group comprises closed parts of speech (numerals, pronouns, conjunctions, prepositions, particles and interjections).

We plan to use the CONCEDE model [7] for dictionary encoding that respects the guidelines of the Text Encoding Initiative Dictionary Working Group (TEI-DWG) (TEI). The CONCEDE project (CONCEDE), supported by the European Commission under INCO-Copernicus program, developed a formal model for lexical databases (in the form of an SGML DTD). The lexical databases in accordance with the guidelines of the TEI-DWG for the six Central and East European languages: Bulgarian, Czech, Estonian, Hungarian,

Romanian, and Slovene were developed. In CONCEDE, all dictionaries use common tagset [5]. In the framework of the project the first LDB for Bulgarian, based on encoding standards established by the TEI, was developed [4].

3.1. Lexical database of the Bulgarian-Polish online dictionary

The tagset for LDB of the Bulgarian-Polish online dictionary contains 3 structural tags and a set of content tags. The full list of tags can be found in the Appendix.

(1) The structural tags are:

alt – a tag indicates alternation, though generally for use in quite different contexts,

entry - a tag, contains the dictionary entry,

struc- a tag indicates separate independent part in the dictionary entry:

<entry>
…
<alt>...</alt>
<struc type=”Sense” n=”1”>…</struc>
<struc type=”Sense” n=”2”> …</struc>
…
</entry>

(2) The set of content tags includes all other tags, among them:

The **hw** tag contains the headword and is used for alphabetization and indexing, access. The **pos** tag indicates the part of speech assigned to a dictionary headword (noun, verb, adjective, etc.):

<hw>**свобод|а’**</hw><pos>noun</pos>.

The **xr** tag uses to indicate a cross reference with the pointer:

<hw>**построя’ва|м**</hw> <xr>**постро|я’**<xr>.

The **orth** tag gives the orthographic form of words (part of word): <orth>**-и’**</orth>.

The **gram** tag contains grammatical information relating to a word other than gender, number, case, person, tense, mood, itype, as these all have their own element, for example, perfective aspect and imperfective (progressive) aspect: <gram>imperfective</gram>.

The **subc** tag contains sub-categorization information (transitive/intransitive for verbs, countable/non-count for nouns, etc.): <subc> transitive </subc>.

We suggest new tags, **conjugation** and **type,** to represent the conjugation of verbs -

conjugation: to represent the conjugation of verbs; its structure allows the sub tag **type** for the possible types of conjugations of Bulgarian verbs;

type: a tag in the frame of **conjugation** tag indicates explicitly one of the three types of conjugation of the Bulgarian verbs, for example:

<conjugation>
<orth>**-ш**</orth>
<type>**I**</type>
</conjugation>

The **trans** tag contains translation text and related information, everything under **trans** relates to the target language: <trans>**wolność**</trans>.

The **eg** tag forms a structure, contains an example, as given in a dictionary, and allows the tags **source** and **q;** the **q** tag contains a quotation or apparent quotation, the **source -** bibliographic source for a quotation: <eg><q>**-я на учи’лище**</q><trans**> chodzę do szkoły** </trans></eg>.

3.2. Examples

The examples contain the dictionary entry in WORD format and a comment on its classifiers. For verbs in particular we suggest a structure of dictionary entry in the LDB of the Bulgarian-Polish online dictionary.

(1) Verbs (*глаголи, czasowniki*):

(1.1) Entry in WORD-format:

постро|я́, -и́ш *vp.* zbudować; uszeregować, uszykować

Comment:

Verb: build/construct /*построя/;* **aspect: perfect** /*свършен вид*/, **transitive verb** /*преходен/,* **-и**ʾ**ш conjugation II type** /*II спрежение/*

LDB structure:

```
<entry>
<hw> постро|я’</hw>
        <pos>verb</pos>
        <gram> perfect </gram>
        <conjugation><orth>-иш</orth>
    <type>II</type>
        </conjugation>
<alt>
<orth>построя’ва|м </orth>
        <gram>imperfect</gram>
        <conjugation><orth>-ш</orth>
    <type>II</type>
        </conjugation>
</alt>
        <subc>transitive</subc>
<struc type="Sense" n="1">
        <trans>zbudować</trans>
</struc>
<struc type="Sense" n="2">
        <trans>uszeregować</trans>
        <alt><trans>uszykować</trans></alt>
</struc>
</entry>
```

(1.2) Entry in WORD-format:

построя́ва|м, -ш *vi*. v. **построя́**

Comment:

Verb: build/construct /*построявам* /**, aspect: imperfect (progressive)** /*несвършен вид*/**, transitive verb** /*преходен/,* **-ш conjugation III type** */III спрежение/*

LDB structure:

```
<entry>
<hw>построя’ва|м</hw>
        <xr>построя’</xr>
</entry>
```

(1.3) Entry in WORD-format:

ви́жда|м, -ш *vi*. widzieć; **~м се** widzieć się; **~ се** zdaje się, wydaje się; widać

Comment:

Verb: see /*виждам/*, **aspect: imperfect (progressive)** /*несвършен вид*/**, transitive verb** /*преходен*/, **-ш conjugation III type** */III спрежение/***,** *czas. ndk* **widzieć ~dzę, ~dzisz** *czas. ndk VIIa*; **~ м се** widzieć się; **~ се** zdaje się, wydaje się; widać

LDB structure:

```
<entry>
<hw>ви’ждам</hw>
        <pos>verb</pos>
        <gram>imperfect</gram>
        <conjugation><orth>-ш</orth>
    <type>III</type>
        </conjugation>
        <subc>transitive</subc>
<struc type="Sense" n="1">
        <trans> widzieć </trans>
```

```
</struc>
<struc type="Derivation" n="1">
        <orth>~м се</orth>
        <struc type="Sense" n="1">
        <trans> widzieć się</trans>
        </struc>
</struc>
<struc type="Derivation" n="2">
        <orth>~ се</orth>
        <struc type="Sense" n="1">
        <trans> zdaje się </trans>
        <alt><trans> wydaje się </trans></alt>
        </struc>
</struc>
<struc type="Sense" n="2">
        <trans> widać </trans>
</struc>
</entry>
```

(1.4) Entry in WORD-format:

сп|я́, -и́ш *vi*. spać; **~и́ ми се** chce mi się spać, ogarnia mnie senność

Comment:

Verb: sleep /*спя*/**, aspect: imperfect (progressive)** /*несвършен вид/,* **intransitive verb** /*непреходен*/, **conjugation II type** */II спрежение*/

LDB structure:

```
<entry>
<hw>сп|я’</hw>
        <pos>verb</pos>
        <gram>imperfect</gram>
        <conjugation><orth>-и’ш</orth>
    <type>II</type>
        </conjugation>
        <subc>intransitive</subc>
<struc type="Sense" n="1">
        <trans> spać </trans>
</struc>
<struc type="Derivation" n="1">
        <orth>~и́ ми се</orth>
<struc type="Sense" n="1">
        <trans> chce mi się spać </trans>
        <alt><trans> ogarnia mnie senność </trans></alt>
</struc>
</struc>
</entry>
```

(1.5) Entry in WORD-format:

хо́д|я, -иш *vi.* chodzić; kursować; **~и слу́х (мълва́)** *lud.* chodzą słuchy, pogłoski; **-я на учи́лище** chodzę do szkoły; **~я си** odchodzę, idę sobie; **~и ми се на ки́но** mam ochotę pójść do kina; **~я ерге́н** jestem kawalerem

Comment:

Verb: walk, go /*хо́дя*/**, aspect: imperfect (progressive)** /*несвършен вид/*, **intransitive verb** /*непреходен/,* **conjugation III type** */III спрежение*/

LDB structure:

```
<entry>
<hw> хо’д|я </hw>
        <pos>verb</pos>
        <gram>imperfect</gram>
        <conjugation><orth>-и’ш</orth>
    <type>III</type>
        </conjugation>
        <subc>intransitive </subc>
<struc type="Sense" n="1">
```

```
            <trans> chodzić </trans>
</struc>
<struc type="Sense" n="2">
            <trans> kursować </trans>
</struc>
<struc type="Phrases"><struc type="Phrase" n="1">
            <orth>~и  слух  (мълва́) </orth>
            <usg type="register"> lud.</usg>
            <trans> chodzą słuchy, pogłoski </trans>
</struc></struc>
<eg><q>-я на учи́лище</q><trans> chodzę do szkoły </trans></eg>
<eg><q>~я си </q><trans> odchodzę </trans>
<alt><trans> idę sobie </trans></alt></eg>
<eg><q>~и ми се на ки́но </q><trans> mam ochotę pójść do kina </trans></eg>
<eg><q>~я ерге́н </q><trans> jestem kawalerem </trans></eg>
</entry>
```

We remark here that the suggested LDB structure of Bulgarian-Polish dictionary entry is suitable for automated generation of a Polish-Bulgarian dictionary entry. For example, from this one in (1.5), a program could generate automatically the simple structures for the corresponding Polish verbs **chodzić** and **kursować:**

```
<entry>
<hw> chodzić </hw>
            <pos>verb</pos>
<struc type="Sense" n="1">
            <trans> хо'д|я </trans>
</struc>
</entry>

<entry>
<hw> kursować </hw>
            <pos>verb</pos>
<struc type="Sense" n="1">
            <trans> хо'д|я </trans>
</struc>
</entry>
```

All others classifiers for the Polish verbs in these entries, derivations, phrases, examples, etc., should be added additionally!

(1.6) **Participle (*причастие, imiesłow*)**

Entry in WORD-format:

следващ *imiesł. przym.* **1. studiujący** *imiesł. przym.*; **2. idący** *imiesł. przym.*, następujący za kimś, następny

Comment:

Participle: next / *следващ*/ *imiesł. przym.* **1. studiujący** *imiesł. przym.*; **2. idący** *imiesł. przym.*, następujący za kimś, następny.

(2) Nouns (*съществителни имена, rzeczowniki*):

(2.1) Entry in WORD-format:

хо́ра *pl* ludzie *pl*

Comment:

Noun: **people** /*xopa*/ *rzecz. l.mn* **(plural) ludzie** *rzecz. l.mn* **(plural)**

(2.2) Entry in WORD-format:

свобод|а́, -и́ *f* wolność *f*, swoboda *f*

Comment:

Noun: **freedom** /*свобода*/**, -и (plural)** *rzecz. ż* (**gender**) **1**.**wolność** *rzecz. ż*, **2. swoboda** *rzecz. ż*

(3) Adjectives (*прилагателни имена, przymiotniki*):

(3.1) Entry in WORD-format:

мек adi. miękki; łagodny; **~а дъждовна вода́** miękka deszczowa woda; **~а зима** łagodna zima; **~и съгла́сни** *gram*. spółgłoski miękkie; **~а ша́пка** kapelusz (męski)

Comment:

Adjective: **soft** /*мек/ przym.* **1. miękki** *przym.*; **2. łagodny** *przym.*; **~а дъждовна вода** miękka deszczowa woda; **~а зима** łagodna zima; **~и съгласни** *gram.* spółgłoski miękkie; **~а шапка** kapelusz (męski)

(3.2) Entry in WORD-format:

и́стински adi. prawdziwy; adv. naprawdę, prawdziwie

Comment:

Adjective**: true** /*истински/ przym.* **prawdziwy** *przym.*; *przysłów.* naprawdę, prawdziwie

(4) Adverbs (*наречия, przysłówki*):

(4.1) Entry in WORD-format:

рядко *adv.* rzadko

Comment:

Adverb: **seldom** /*рядко/ przysłów.* **rzadko** *przysłów.*

(4.2) Entry in WORD-format:

ско́ро *adv.* prędko, rychło, szybko**;** niedawno, wkrótce; **мно́го ~ свърших та́я ра́бота** bardzo prędko skończyłem tę pracę; **ще се върна ~** wkrótce wrócę; **ча́с по-~** czym prędzej

Comment:

Adverb: **soon** /*скоро/ przysłów.* **1. prędko** *przysłów.*, **2. rychło** *przysłów.*, **3. szybko** *przysłów.*; **4. niedawno** *przysłów.*, **5. wkrótce** *przysłów.*; **много ~ свърших тая работа** bardzo prędko skończyłem tę pracę; **ще се върна ~** wkrótce wrócę; **час по-~** czym prędzej

(5) Pronouns (*местоимения, zaimki*):

Entry in WORD-format:

не́гов *pron. poss*. jego

Comment:

Pronoun: **his, its** /*негов/ zaimek dzierż.* **jego** *zaimek dzierż. r. męski* (**gender**) *D. B.*

(6) Conjunctions (*съюзи, spójniki*):

Entry in WORD-format:

но *coni.* ale, lecz; **не са́мо то́й, ~ и а́з** nie tylko on, ale i ja; **и́скат, ~ не мо́гат** chcą, ale nie mogą

Comment:

Conjunctions**: but** /*но/ spójnik* **1. ale** *spójnik*, **2. lecz** *spójnik*; **не са́мо той, ~ и а́з** nie tylko on, ale i ja; **и́скат, ~ не мо́гат** chcą, ale nie mogą

(7) Prepositions (*предлози, przyimki*):

Entry in WORD-format:

пред *praep.* przed; wobec; ~ **университе́та** przed uniwersytetem; **явя́вам се ~ съда́** stoję przed sądem; **вино́вен съм ~ ва́с** czuję się wobec was winny; **вси́чки гра́ждани са ра́вни ~ зако́на** wszyscy obywatele są równi wobec prawa; **оста́на глу́х ~ молби́те му** pozostał głuchy na jego prośby; **и́мам ~ ви́д** mam na uwadze; ~ **ви́д на ...** z uwagi na…

ze względu na…

Comment:

Preposition: **in front of; before; at; to;** /*пред/ przyim.* **1. przed** *przyim.*; **2. wobec** *przyim.*;

(8) Particles (*частици, partykuły*):

Entry in WORD-format:

не *partyk*. *przecząca* nie

Comment:

Particle: **no** *не partyk. przecząca* **nie** *partyk. przecząca*

(9) Numerals (*числителни имена, liczebniki*):

Entry in WORD-format:

четири́ма *num.* czterej; czworo

Comment:

Numeral: **four persons** /*четирима*/ *liczeb.* 1st sense: **czterej;** 2nd sense**: czworo** *liczeb.*

(10) Interjections (*междуметия, wykrzykniki*):

Entry in WORD-format:

о́х! *interi.* o!, och! (na wyrażenie bólu, smutku, radości, zachwytu, zdziwienia itp.)

Comment:

Interjection: **oh** /*ox!/ wykrzyk.* **o!, och!** *wykrzyk.* (Explanation: na wyrażenie bólu, smutku, radości, zachwytu, zdziwienia itp.)

4. 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.

Bibliography

[1] Andrejchin, L. (1944). Основна българска граматика. София. (in Bulgarian)

[2] Dimitrova, L., Koseska-Toszewa, V. (2008). The Significance of Entry Classifiers in Digital Dictionaries. In Proceedings of the MONDILEX First Open Workshop, Moscow, Russia, 3–4 October 2008, pages 89–97, Russian Academy of Sciences, IITP.

[3] Dimitrova, L., Koseska-Toszewa, V. (2008). Some Problems in Multilingual Digital Dictionaries. SOW, 8, 237–254.

[4] Dimitrova, L., Pavlov, R., Simov, K. (2002). The Bulgarian Dictionary in Multilingual Data Bases. Cybernetics and Information Technologies, 2(2), 33–42.

[5] Erjavec, T., Evans, R., Ide, N., Kilgarriff, A. (2000). The Concede Model for Lexical Databases. Proceedings of the Second International Conference on Language Resources and Evaluation, LREC'00. 355-362, ELRA, Paris.

[6] Ivanchev, S. (1971). Проблеми на аспектуалността в славянските езици. София. (in Bulgarian)

[7] Kilgarriff, A. (1999). Generic encoding principles. CONCEDE Project Deliverable 2.1. University of Brighton, UK.

[8] Koseska – Toszewa, V. (2009). Form, its meaning, and dictionary entries (in this volume)

[9] Koseska, V., G. Gargov. (1990). Bulgarian-Polish Contrastive Grammar, vol. 2. Special Definiteness-Indefiniteness category, Sofia. (in Bulgarian)

[10] Koseska, V., Mazurkiewicz, A. (2009) Net-Based Description of Modality in Natural Language (on the Example of Conditional Modality). Proceedings of the MONDILEX Open Workshop, Kiev, 2–3 February 2009. (be appear)

[11] Mazurkiewicz, A. (2008) A Formal Description of Temporality (Petri net approach). Proceedings of the MONDILEX First Open Workshop, Moscow, Russia, 3–4 October 2008. pages 98–108.

[12] Saloni, Z., Gruszczyński, W., Woliński, M., Wołosz R. (2007) Słownik gramatyczny języka polskiego, Wiedza Powszechna, Warszawa.CONCEDE:

[13] http://www.itri.brighton.ac.uk/projects/concede/

[14] TEI: http://www.tei-c.org/index.xml

APPENDIX

The ***structural tags***, used in the LDB of the Polish-Bulgarian online dictionary, are three: **entry, struc, alt.**

alt: alternation, though generally for use in quite different contexts
entry: dictionary entry
struc: indicates separate independent part in the dictionary entry.

The set of ***content tags*** includes the elements:
case: contains grammatical case information given by a dictionary for a given form
conjugation: ***a new tag*** is added to represent the conjugation of verbs; its structure allows the sub tag **type** for the possible types of conjugations of Bulgarian verbs
def: directly contains the text of the definition
domain: domain
eg: a structure, contains an example, as given in a dictionary, and allows the tags **source** and **q**
etym: a structure, contains etymological information and allows the tags **lang** and **m**, as given in a dictionary
gen: identifies the morphological gender of a lexical item, as given in the dictionary
geo: geographic area
gram: contains grammatical information relating to a word other than gender, number, case, person, tense, mood, itype, as these all have their own element, for example, perfect aspect and progressive aspect
hw: the headword; used for alphabetization and indexing, access
itype: indicates the inflectional class associated with a lexical item, as given in a dictionary
lang: language; for use in etymologies (in **etym**)
m: indicates a grammatical morpheme in the context of etymology
mood: contains information about the grammatical mood of verbs, as given in a dictionary
number: indicates grammatical number associated with a form, as given in a dictionary
orth: gives the orthographic form of a dictionary headword
person: indicates grammatical person associated with a form, as given in a dictionary
pos: indicates the part of speech assigned to a dictionary headword (noun, verb, adjective, etc.)
q: contains a quotation or apparent quotation
register: register, for type attribute on **usg** tag
source: bibliographic source for a quotation
subc: contains sub-categorization information (transitive/intransitive, countable/non-count, etc.)
time: temporal, historical era, for example, “archaic”, “old”, etc.
type: ***a new*** subtag in the frame of **conjugation** tag indicates explicitly one of the three types of conjugation of the Bulgarian verbs
tns: indicates the grammatical tense associated with a given inflected form in a dictionary **trans**: contains translation text and related information, so may contain any of the content tags; the principle is that everything under **trans** relates to the target language
usg: contains usage information in a dictionary entry, other than **time**, **domain**, **register** (as these all have their own element), like “dialect”, “folk”, “colloquialism”, etc.
xr: uses to indicate a cross reference with the pointer.