Online Language Resources and Technology Tools provided by the Institute for Bulgarian Language

Svetla Koeva

Institute for Bulgarian Language, Bulgarian Academy of Sciences

*Corresponding author: Prof. Svetla Koeva, PhD, Institute for Bulgarian Language, 52 Shipchenski prohod, Bldg. 17, Sofia, 1113, Tel.: +359 9792939, Mobile: + 359 888 950997, Fax: + 359 2 8722302, email: svetla@dcl.bas.bg

Running title: Online Language Resources and Technology Tools

Abstract
We describe our efforts to provide researchers, Bulgarian language teachers and students, as well as the wide public, with easy access to comprehensive services with respect to language resources and technology tools. This includes: a) developing and accumulating a wide range of language resources (monolingual and multilingual dictionaries, lexicographic and terminological databases, semantic networks) and b) developing and advancing tools for finding, analysing, annotating, exploring and using language data.

In particular, we present the Bulgarian Language Portal LexIt (http://ibl.bas.bg/dictionary_portal/), which integrates twelve dictionaries and lexicographic databases, allows simultaneous access to the integrated resources and retrieves information for lemmas, word forms, grammatical information, word senses, examples, synonyms, antonyms, new words and phraseology. The Bulgarian Language Portal offers cross-references between the integrated resources, the most prominent among them being: a) the multivolume Dictionary of Bulgarian Language (comprising about 120,000 headwords, their word senses and many examples); b) the Grammatical Dictionary of Bulgarian (consisting of about 75,000 lemmas linked with their word forms and grammatical characteristics); c) the Infolex database (combining more than 54,000 synonyms, antonyms, idiomatic expressions and neologisms). In addition, we will present the Bulgarian WordNet, a lexical-semantic network that integrates about 120,000 sets of synonyms linked with different semantic, derivational, and extra-linguistic relations, as well as with their translation equivalents in 21 languages.

Practical applications
Nearly all major dictionaries developed at the Institute for Bulgarian Language have their online versions: either as electronic books, part of the Electronic library of the Institute for Bulgarian Language, or as stand-alone or integrated online dictionaries. The users can easily obtain the information they are searching for and the results can be compared online. The Bulgarian WordNet offers information for word senses and semantic relations between concepts. Further, through the inter-language relations of equivalence with the Princeton WordNet, the lexical and semantic information is accessible online in WordNets for 22 languages.

Key words: language resources, online dictionaries, wordnet, language technologies, dictionary portal
Introduction

In recent years a large number of Bulgarian dictionaries and other language resources focused on Bulgarian have been made available online aiming at researchers, teachers and students in the Bulgarian language, as well as the wide public. The term language resources (data) refers to sets of language data and descriptions in machine readable form, used in building, improving or evaluating natural language and speech processing systems (Cucchiari et al., 2001) and for language studies, language education, and language consultations. A broad definition of language resources also includes large variety of language processing tools building up on language resources to enhance them with more information.

Depending on their content and structure, language resources can generally be divided into four types (Gavrilidou et al. 2012): corpora (including written, spoken, multimodal and multimedia corpora; monolingual, domain-oriented, multilingual and linguistically annotated corpora, etc.); lexical or conceptual resources (including dictionaries, terminological resources, word lists, thesauri, wordnets, ontologies, etc.), language description (including grammars, typological databases, etc.), tools or services (including processing tools: tagging, lemmatisation, named entity recognition, anaphora resolution, clustering and domain classification; applications; web services, etc.).

The language resources are used to enhance the quality and effectiveness of communication by means of spelling and grammar checking; machine translation; categorisation and summarisation of documents, search and extraction of information, transformation of written text into speech and vice versa. Another important function of language resources is to improve language training in language learning, to support the language studies and to enhance linguistic culture of society.

In the present overview, we are focused on some of the online language resources of the Institute for Bulgarian Language (namely dictionaries, lexicographic databases and the lexical-semantic network WordNet) and on some of technology tools behind them and their public access.

There have already been more than 20 years (in 2018) since three media for dictionaries and corpora are used—paper, desktop applications and the Internet. Until recently the online dictionaries for Bulgarian were not published by lexicographers, linguists or major publishers and they either were not particularly reliable or were illegal copies of already printed dictionaries and books. However, in the past few years a large number of Bulgarian dictionaries and other language resources focused on Bulgarian have been made available online by the Institute for Bulgarian Language, aiming at researchers, teachers and students in the Bulgarian language, as well as the wide public.

As a result, nearly all major dictionaries developed at the Institute for Bulgarian Language now have their online versions: either as electronic books, part of the Electronic library¹ of the Institute for Bulgarian Language, or as stand-alone or integrated online dictionaries. In the following section the most valuable dictionaries, part of the Electronic library, will be presented briefly; more attention will be paid to the Bulgarian Language Portal LexIt², integrating twelve dictionaries for simultaneous lookups; at the end, the Bulgarian WordNet will be presented, emphasising the different types of information it integrates.

Electronic library of the Institute for Bulgarian language

Some of the distinguish features of electronic books are that they are dynamic, reactive, can be made available in different formats and are easily accessible (Landoni et al., 2000). The Electronic library of the Institute for Bulgarian Language is maintained by the Internet archive BookReader³ – an open software providing online access to scanned books. So far the electronic books of the Institute for Bulgarian Language are not content searchable. The pages can be moved forward and backward, the text can be zoomed in and zoomed out, the reader can choose the page number on the screen.

The Electronic library of the Institute for Bulgarian language does not contain many titles so far; however, it provides an access to some of the most meaningful dictionaries, a subject of constant interest from

¹ http://ibl.bas.bg/informatsiya/uslugi/elektronna-biblioteka/
² http://ibl.bas.bg/dictionary_portal/
³ https://openlibrary.org/dev/docs/bookreader
scientists, teachers, students and the wide public. Among them, the following dictionaries can be highlighted:

**The multi-volume Bulgarian Etymological Dictionary**
The Bulgarian etymological dictionary explains the origin of the words in the Bulgarian language and clarifies the links between the Bulgarian language and other Balkan, Slavic and Indo-European languages (Deykova, 2017). From 1971 to 2017, 8 volumes of the Dictionary were issued (Georgiev, 1971, 1979, 1986; Georgiev & Duridanov, 1995; Dimitrova-Todotova & Selimski, 2017; Racheva & Todorov, 1999, 2002; Todorov, 2013). The Dictionary's articles have a complex structure and represent the meanings of the headword, its phonetic and morphological variants in the Bulgarian dialects and other Slavic languages, and explain the word's origin. The first seven volumes of the Bulgarian etymology Dictionary have been part of the Electronic library of the Institute for Bulgarian Language since 2017.

**The two-volume Old Bulgarian Dictionary**
The Old Bulgarian Dictionary represents the vocabulary encountered in classical Glagolitic and Cyrillic monuments (Ivanova-Mircheva, 1999, 2009). The dictionary articles present the senses of the headwords, examples, grammatical characteristics and the translation of the Old Bulgarian words with their contemporary equivalents – literary or dialectic. The Dictionary has been part of the Electronic library of the Institute for Bulgarian Language since 2015.

**The four-volume Bulgarian Dialect Atlas**
The Bulgarian dialect atlas contains 513 dialect maps representing the phonetic, accentual, lexical and morphological system of the Bulgarian dialects based on sources from over 2,300 settlements (Tetovska-Troeva, 2001, 2016). The Dictionary has been part of the Electronic library of the Institute for Bulgarian Language since 2017.

**Bulgarian Language Portal LexIt**
At the Institute for Bulgarian Language, 7 dictionaries have been published online in the last 8 years: (1) the Grammatical Dictionary of Bulgarian (the dictionary provides information about spelling, the grammatical characteristics of a word form, its lemma and the grammatical characteristics of the lemma; it has been available on the Internet since 2010); (2) the Dictionary of Bulgarian Language (the dictionary includes headwords, explanatory definitions, examples, grammatical and stylistic information; it has been available on the Internet since 2014; the online version of the dictionary is extended with the third volume of the Dictionary of Modern Standard Bulgarian); and (3) the Infolex database (it consists of four dictionaries: dictionary of synonyms in Bulgarian, dictionary of antonyms in Bulgarian, dictionary of neologisms in Bulgarian and idiomatic expressions in Bulgarian, including both multwort expressions and idioms; the database has been available on the Internet since 2014).

With the increase in the number of online dictionaries, the need has naturally arisen to create a single environment in which all dictionaries are integrated, so that the users can obtain all available information within a single search. To answer this need, the Bulgarian Language Portal LexIt was created. At the time of its creation in 2016 the Portal contained the Grammatical Dictionary of Bulgarian, the Dictionary of Bulgarian Language, the third volume of the Dictionary of Modern Standard Bulgarian and the four dictionaries of the the Infolex database. Recently three Bulgarian-English terminological dictionaries and two volumes of the Dictionary of New Words in Bulgarian have also been made available through the Bulgarian Language Portal.

In this paper we present the Bulgarian Language Portal following the classification of Engelberg and Mueller-Spitzer (2014) which outlines the main criteria for distinguishing the types of dictionary portals as follows: (1) types of access provided, (2) the implemented cross-reference structures between resources, (3) the ownership relation between the portal and the contributing resources and (4) the layout of the portal. Additional properties according to which dictionary portals can be described are also pointed out: (5) the presentation of the search results, (6) the languages described, (7) the number of dictionaries integrated. More information for the features of the online dictionaries is presented by Lew (2012) and Müller-Spitzer, Koplenig, and Töpel (2014) and dictionary portals for different languages are described by Ahačič, Ledinek, and Perdh (2015) and Müller-Spitzer, Michaelis, and Koplenig (2012).
Types of access
According to the Engelberg and Mueller-Spitzer classification (2014) three main types of access structures within dictionary portals can be distinguished: (1) External access structures which provide access to the surface of the dictionaries (such structures contain links to the starting pages of the dictionaries and all searches for particular lexicographic information are carried out within the dictionaries themselves, i.e., outside the portal); (2) Outer access structures (such structures access dictionary entries by searching through all the lemmas of the integrated dictionaries); (3) Inner access structures (such structures allow to call up information from specific positions within the microstructure of the entries of the participating dictionaries). Further, in some dictionary portals, entries (lemmas) of different integrated dictionaries are explicitly cross-referenced.

Following the above classification, the Bulgarian Language Portal can be described as follows: it provides outer access to all dictionaries integrated so far in its structure and inner access to the Grammatical Dictionary of Bulgarian, and the Infolex database (containing four dictionaries).

By default, the search is conducted in all dictionaries. It is possible to select a dictionary or a set of dictionaries in which the search should be performed. The Bulgarian Language Portal LexIt allows access to the integrated resources simultaneously: one, several or all resources may be selected for a particular search. The retrieved information includes lemmas, word forms, grammatical information related with lemmas and word forms, spelling, word senses of polysematic words, definitions of senses, stylistic labels, examples, homonyms, synonyms, antonyms, new words, phraseology and bilingual terminology. All the information related with a lemma can be accessed simultaneously. As the lemmas are associated with their word forms in the Grammatical Dictionary of Bulgarian, the information related with a lemma can be searched either by a lemma or by a word form. For example, a query for a word form детето (the child) returns the lemma дете (child) with its grammatical characteristics (common noun, neuter) and all four forms of the lemma and their grammatical characteristics (singular or plural, definite or indefinite) – Figure 1.

The lemma of the queried word links to any of the other resources it appears in: the Dictionary of Bulgarian, the database Infolex, the terminological dictionaries and the Dictionaries of new words. The query куче (a dog) returns synonyms of the word, 114 multiword and idiomatic expressions, containing the word куче and the entry for the headword куче from the Dictionary of Bulgarian language (two word senses, many examples, many idiomatic expressions and other lexicographic information).

Cross-reference links
The entries in integrated dictionaries can be explicitly cross-referenced by means of their lemmas. The word forms and lemmas are linked in the Grammatical Dictionary of Bulgarian and in the Infolex database. Synonyms and antonyms are linked in the Infolex database. Thus, links interconnecting particular words can be followed (namely between synonyms and antonyms) and each link initiates a new query. For example, Figure 2 shows the synonyms of word чудно representing six of its word senses: good-looking, aesthetic, appropriate, nice (for weather), scenic (for landscape), legible (for handwriting), and four antonyms: bad, ugly, not good, and bad (colloquial). Each synonym and antonym is a link initiating a new query retrieving available information from the twelve integrated dictionaries.

The ownership of the information
The interrelated resources are developed and maintained by one institution (the Institute for Bulgarian Language at the Bulgarian Academy of Sciences) which facilitates the structuring of the content according to the needs of the users and the installation of various cross-references between the different resources.

The layout of the Portal
The navigation and search facilities are equivalent across all dictionaries integrated within the Bulgarian Language Portal LexIt. The user is allowed to select one or several resources or to search globally across all dictionaries.

The presentation of the results
The Bulgarian Language Portal lexIt displays entire entries from all resources that contain the query word. The results of a search are shown directly without redirection to other online resources. In certain cases the information related to a query is large in volume (from the Dictionary of Bulgarian Language and the Infolex database) – then, only the first meaning or the first few synonymous sets (or pair of antonyms or phraseological units) are displayed; the rest of the information can be viewed following
the link “See more” and the results are displayed at the same page (without redirecting).

Languages
The Bulgarian Language Portal is devoted to the Bulgarian language and contains monolingual or bilingual dictionaries representing Bulgarian. The queries can be made in all languages present in the dictionary collection (so far Bulgarian and English).

Number and type of resources
The portal integrates newly developed dictionaries with dictionaries primarily published in print. The Bulgarian Language Portal Lexrix contains twelve dictionaries: one spelling and grammatical dictionary; two explanatory dictionaries combined in one online lexicographic resource; three Bulgarian-English terminological dictionaries, a dictionary of synonyms, a dictionary of antonyms, a dictionary of neologisms and a dictionary of idiomatic expressions; two dictionaries of new words.

Grammatical Dictionary of Bulgarian
The Grammatical Dictionary of Bulgarian (Koeva, 1998) contains approx. 75,000 lemmas and more than 1,100,000 word forms. Its structure follows the structure of the DELA type finite state dictionaries (Silberztein, 2005) although the formalisation differs. The lemmas are associated with an inflectional paradigm, common for all words that have equivalent word formation and express equivalent grammatical categories. The Grammatical Dictionary of Bulgarian was manually crafted, the inflection paradigms were defined and then words were classified to a particular class. The inflectional paradigms provide information for words formation, including the position of the stress and vowel alternations. The word form, the morphological alternations and the particular grammatical characteristics associated with each word form are provided with a formal description ensuring the processing with finite state transducers.

The Grammatical dictionary has been available on the Internet since 2010.

Multi-volume Dictionary of Bulgarian Language

The dictionary includes headwords, word senses and examples attested in the literature of the Bulgarian National Revival and of the period following the Liberation, in fiction, science and popular science, journalistic publications and periodicals of the 20th and the 21st century, as well in the spoken language. In addition to the general vocabulary, following a certain selection, the dictionary includes some obsolete words, words gradually moving to the passive vocabulary, foreign words which are widely used in modern Bulgarian, terminological expressions, dialect words, personal names from history, mythology or fiction which have become bywords or have acquired a figurative sense, some slang as an exception. The multi-volume Dictionary of Bulgarian Language records not only the modern state of the Bulgarian vocabulary but also its development from the period of the formation of the new standard Bulgarian language to modern times, presenting as thoroughly as possible the semantic properties of the lexical units, their grammatical peculiarities, their collocational preferences and phonetic variations (Krumova, 2013).

Database Infolex
As mentioned before, the database Infolex combines the following dictionaries: Dictionary of synonyms in Bulgarian, Dictionary of antonyms in Bulgarian, Dictionary of neologisms in Bulgarian and Dictionary of idiomatic expressions in Bulgarian (including both multiword expressions and idioms) (Blagoeva & Kolkovska, 2014). The Dictionary of Synonyms in Bulgarian contains about 28,000 words pertaining to

4 http://ibl.bas.bg/rbe/
four parts of speech, distributed into synonym sets (approx. 12,000 nouns, 10,000 verbs, 4,000 adjectives and 2,000 adverbs). The Dictionary of Antonyms contains about 14,000 words (approx. 5,000 nouns, 3,000 verbs, 4,000 adjectives and 2,000 adverbs. The Dictionary of Neologisms in Bulgarian contains over 2,200 new words that have entered the Bulgarian language in the past 25 years. The Dictionary of idiomatic expressions in Bulgarian contains units found in the main Bulgarian phraseological dictionaries. The total number of the idiomatic expressions is over 10,000. The Infolex database has been available on the Internet since 2014.

Bilingual terminological dictionaries
Recently three Bulgarian-English terminological dictionaries have also been made available through the Bulgarian Language Portal LexIt. The resource combines three bilingual Bulgarian-English terminological dictionaries, representing respectively 729 terms from the Computer domain; 907 terms from the domain of Economy (business administration, finance and accounting, trade, Insurance) and 1,175 terms from the Law domain (civil law, criminal law, commercial law, public law, international law and human rights). The terminological resources have been compiled from the Bulgarian WordNet (Koeva, 2014) but also some other terminological monolingual dictionaries and domain specific corpora were consulted.

Dictionaries of New Words
The two dictionaries embrace new words and meanings that have been used in the Bulgarian social, political, economic, technical, sports and cultural life (Pernishka et al., 2001, 2010). These are both words or complex terms or phrases, adopted mainly from English, but also from German, Italian, Arabic, Japanese, etc.

Each of the two dictionaries reflects a 10-year periods: from 1990 to 2000 and from 2001 to 2010. The dictionary from 2010 conations about 4,300 new words, terms and phrases, and about 700 new meanings attached to already existing words.

Most of the borrowed words undergo a phonetic, morphological and semantic adaptation in the system of the Bulgarian language. At the same time, they are often used for the formation of new derivative words, new complex terms, etc. Such processes testify the active use of part of the new vocabulary and its integration into the language system.

The dictionary entries of the two dictionaries include headwords, stylistic and grammatical labels, partial grammatical information, word senses, examples and phraseology. The two Dictionaries of New Words have been made available through the Bulgarian Language Portal LexIt since 2018.

General characteristic of the Bulgarian Language Portal
Engelberg and Mueller-Spitzer (2014) distinguishes three types of dictionary portals based on the types of access, cross-reference links, ownership and layout of the portal, i.e. by the degree to which the portal integrates the dictionaries: dictionary nets, dictionary search engines and dictionary collections. The Bulgarian Language Portal lexIt can be classified as a dictionary net because it allows inner, outer and external access to twelve dictionaries, shows inter-dictionary cross-reference and provides an integrated layout for the portal and the dictionaries. Table 1 illustrates the distribution of some Bulgarian words in the twelve dictionaries constituting the Bulgarian Language Portal LexIt. Table 2 shows the coverage of the information provided by the twelve dictionaries integrated within the Bulgarian Language Portal LexIt.

Bulgarian WordNet
The human knowledge about synonyms – words (and/or multiword expressions) denoting one and the same concept and semantic relations such as hyponymy, meronymy, antonymy, etc., is encoded in the lexical-semantic network WordNet (Miller, 1990: 235 – 312). WordNet is a semantic network that represents lexical knowledge in the form of interconnected nodes, which represent concepts by one or more words called synsets. Respectively, the arcs represent relations between nodes: lexical, semantic, morphosemantic, derivational and extralinguistic. For example, hyperonymy – hyponymy and hyernymy are transitive asymmetric inverse semantic relations which organises nouns and verbs in a tree graph, part of the semantic network.

Comparing conventional dictionaries and WordNet George Miller (1990: 245 – 264) points out four major differences between them. The prototypical dictionary definition of a noun consists of its immediate

---

5 http://ibl.bas.bg/infolex/
superordinate, followed by a differentia part that describes how this instance differs from all other instances. However, the definition does not specify which are the appropriate superordinates; it usually does not contain information about coordinate terms; it does not point downward to hyponyms. A dictionary definition draws some important distinctions and serves to remind the reader of something that is presumed to be already familiar; it is not intended as a catalogue of general knowledge. Instead, WordNet is a lexical-semantic network and ontology, and offers a large network of concepts and named entities along with an extensive multilingual lexical coverage. Table 3 and Table 4 depict the synsets, the gloss, the hyponyms and part of the hyponyms of the sense (company) in Bulgarian and English. The Bulgarian WordNet combines the qualities of well-known lexicographical resources – it contains definitions and examples as dictionaries, but also organises the words in synonymous sets, and the synonymous sets – in a semantic net that reflects the system of concepts in human thinking. In general, Bulgarian WordNet mirrors the Princeton WordNet structure in order to ensure interlingual correspondence with the other wordnets. Nonetheless, the Bulgarian WordNet structure includes language-specific concepts and relations, as well as notes that reflect the linguistic differences on the conceptual, lexical and grammatical level (Koeva, 2014). The Princeton WordNet relations are automatically transferred to Bulgarian WordNet and treated very conservatively in the following way (Piasecky & Koeva, 2017):

- a. The set of Princeton WordNet relations is preserved with a few exceptions.
- b. The Princeton WordNet relations are carefully validated; if a relation does not hold for Bulgarian, this is described in a note to the Bulgarian synset.
- c. Addition of new relations is restricted to derivational relations and to language specific concepts of Bulgarian.

The Bulgarian WordNet has been available online since 2016 (Rizov & Dimitrova, 2018). The WordNet browser Hydra for the Web allows access to 22 languages (Albanian, Arabic, Bulgarian, Catalan, Croatian, English, French, Galician, Greek, Hebrew, Dutch, Icelandic, Italian, Polish, Portuguese, Romanian, Serbian, Slovak, Slovenian, Spanish, Swedish, Finish) obtained either under an open license or by means of a bilateral contract. The search system provides results in all of the available languages selected by the user. The results show the information which the wordnets contain: synsets, definitions, literals, part of speech of the synset, semantic class of the synset, examples, relations between literals and synsets. The user can select a pair of wordnets to be displayed. In such cases, the translation equivalents are visualised simultaneously due to the synchronisation with the Princeton WordNet.

Conclusions

Twelve Bulgarian dictionaries of various types have been integrated and made accessible via the Bulgarian Language Portal LexIt providing access to an extensive database of resources. The portal shows a cross-reference between the resources and can be classified as a dictionary net. Further development of the portal will aim at providing more lexicographic resources describing the Bulgarian language and extending the cross-references between the dictionaries. In addition, the WordNet information can be interconnected with those coming from dictionaries.

The users can provide feedback and can ask questions about Bulgarian vocabulary, grammar, phraseology and spelling using the web question form of the Online Language Consultations Guide and on the Facebook Language Consultations page where answers are provided within an hour. The feedback concerns: most problematic areas of Bulgarian spelling, grammar and lexical system which need to be widely addressed and clarified; language phenomena which are widely discussed in the public, such as usage of foreign words, dialect expressions, new words, etc.; the specific needs of the education in Bulgarian and the ways it can be enhanced by electronic language resources and technologies.

References


---

6 http://dcl.bas.bg/bulnet/
7 http://compling.hss.ntu.edu.sg/omw/
8 http://ibl.bas.bg/ezikovi_spravki/


Georgiev, V. (ed.) Bulgarian etymological dictionary. V. 1, Sofia: Publishing House of the Bulgarian Academy of Sciences, 1971. [In Bulgarian]

Georgiev, V. (ed.) Bulgarian etymological dictionary. V. 2, Sofia: Publishing House of the Bulgarian Academy of Sciences, 1979. [In Bulgarian]


Dimitrova-Todotova, L. & Selimski, L.(eds.) Bulgarian etymological dictionary. V. 8, Sofia: Publishing House of Bulgarian Academy of Sciences "Prof. Marin Drinov", 2017. [In Bulgarian]


Koeva, S. Bulgarian Grammatical Dictionary. Description of the structure of language data. – In: Bulgarian Language, 1998, No 6, pp. 49 – 58. [In Bulgarian]


House of Bulgarian Academy of Sciences, 1959. [In Bulgarian]
Figure 1. Word form query retrieving the lemma, all word forms and grammatical information.

Table 1. Some examples of distribution across dictionaries

<table>
<thead>
<tr>
<th></th>
<th>Grammatical dictionary</th>
<th>Synonyms</th>
<th>Antonyms</th>
<th>Phraseology</th>
<th>Neologisms</th>
<th>Dictionary of Bulgarian</th>
<th>Terminology</th>
<th>New words</th>
</tr>
</thead>
<tbody>
<tr>
<td>пресагент (presagent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>премиера (premiere)</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>присвоявам (to appropriate)</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>присъда (judgment)</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>присъщ (inherent)</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>приятел (friend)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>про (for)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Coverage of the information in dictionaries – general overview.

<table>
<thead>
<tr>
<th></th>
<th>New words</th>
<th>New words</th>
<th>New words</th>
<th>New words</th>
<th>New words</th>
<th>New words</th>
<th>New words</th>
<th>New words</th>
</tr>
</thead>
<tbody>
<tr>
<td>head word</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>word forms</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>synonyms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>antonyms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>phraseology</td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>examples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>grammatical</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>translation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>equivalents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Retrieved information from the Dictionary of Bulgarian language.
**Table 3.** The synset, the gloss, the hyponyms and the hypernyms of the sense {company} in Bulgarian

<table>
<thead>
<tr>
<th>Synset: компания, дружество</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition: обединение, създадено за осъществяване на търговска или друга дейност, юридическо лице</td>
</tr>
<tr>
<td><strong>Hypernyms</strong></td>
</tr>
<tr>
<td>институция, учреждение</td>
</tr>
<tr>
<td>организация, групировка, група</td>
</tr>
<tr>
<td>социална група, обществена група</td>
</tr>
<tr>
<td>група, групировка</td>
</tr>
<tr>
<td>абстракция, абстаркно понятие</td>
</tr>
</tbody>
</table>

**Table 4.** The synset, the gloss, the hyponyms and the hypernyms of the sense {company} in English

<table>
<thead>
<tr>
<th>Synset: company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition: an institution created to conduct business</td>
</tr>
<tr>
<td><strong>Hypernyms</strong></td>
</tr>
<tr>
<td>institution, establishment</td>
</tr>
<tr>
<td>organisation</td>
</tr>
<tr>
<td>social group</td>
</tr>
<tr>
<td>group, grouping</td>
</tr>
<tr>
<td>abstraction, abstract entity</td>
</tr>
</tbody>
</table>