БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ . BULGARIAN ACADEMY OF SCIENCES

ПРОБЛЕМИ НА ТЕХНИЧЕСКАТА КИБЕРНЕТИКА И РОБОТИКАТА, 48
PROBLEMS OF ENGINEERING CYBERNETICS AND ROBOTICS, 48

София . 1999 . Sofia

Analyzing Bulgarian and English Collocations*

Lydia Sinapova, Danail Dochev

Institute of Information Technologies, 1113 Sofia

1. Introduction

In the recent years a number of systems have been developed for automatic extraction and manipulation of collocations - fixed, identifiable, non-idiomatic phrases and constructions. No such systems exist for collocations in Bulgarian language. The presented work can be considered as an initial attempt to fill this gap.

A collocation can be defined as "a sequence of two or more consecutive words, that has characteristics of a syntactic and semantic unit, and whose exact and unambiguous meaning or connotation cannot be derived directly from the meaning or connotation of its components" (Choueka [3]). Some examples of collocational expressions as given in (Choueka [3]) are the following:

– personal nouns (names of specific, individual, unique entities): e.g. *Ronald Reagan*, *President Reagan*, *United Nations*;

– common nouns: e.g. *ice cream*, *high school*, *artificial intelligence*;

– idiomatic expressions: e.g. *change mind*, *sit down*, *hit and run*, *research and development* and idioms like *once upon a time*.

Thus, it generally assumed that a collocation consists of two or more words which have a strong tendency to be used together. For example, in English you say *turn off the light*, not *close the light*; *strong coffee*, not *powerful coffee*; *She was attacked in broad daylight*, not *She was attacked in bright day light*. In Bulgarian examples of collocations are *горещ спор*, *сърдечни поздрави*, *силен дъжд*.

Most studies on collocations stress the difference between truly fixed multi-word units (idioms) and collocations proper such as *pay attention*, *make a mistake* etc. Compared to idioms, collocations allow certain flexibility in their usage, as illustrated in the next examples:

John paid attention to the problem.
John paid little attention to the problem.
Little attention was paid to the problem.

* This research has been partially supported by project NI-I-608-96 "Modelling of structural text characteristics by knowledge-based schemata" of the National Science Foundation and project 010037 "Methods for knowledge representation and processing in the modern information technologies" of IIT – BAS.

Collocations are essential elements in all kinds of text. They can be extracted automatically by means of statistical processing of large volumes of texts – text corpora. Corpus data is most useful for the following:

– statistical methods make it possible to identify collocations associated with a lemma;

– variations in collocations can be shown (mainly syntactic and lexical variations – which traditional dictionaries usually do not indicate explicitly);

– canonical statements of the collocations in dictionaries for human users can be exemplified.

It has been observed that collocations occur much more frequently in formal/ informative texts than in informal/ imaginative ones. This difference would be explained by the more stereotyped aspect of formal texts.

2. Types of Collocations in English and Bulgarian

Each language has its own set of collocations. This chapter describes the types of English collocations as presented in [1], and their correspondence to Bulgarian collocations. Collocations fall into two major groups - grammatical collocations and lexical collocations. According to [1], a grammatical collocation is a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or clause. They are distinguished from the free combinations, which consist of elements joined together in accordance with the general syntactic rules of the language in consideration.

2.1. Grammatical collocations

The following types of grammatical collocations are discussed in [BBI91]:

G1. noun + preposition

English examples:	Bulgarian examples:	
apathy towards	*представа за*	*обвинение в*
admiration for	*режим на*	*тенденция към*
talent for	*талант за*	*отношение към*

The authors of [1] do not consider *noun + of* and *noun + by* combinations, as such combinations are used with a large number of nouns meaning "direct object", "subject" or "possession".

G2. nouns + to + infinitive

Examples are:

It was a pleasure to work there.

They made an attempt to do it.

Most of the nouns allowing G2 collocations can be used also with a verb form in -ing: *It was a pleasure working there*.

Authors do not include nouns if they are followed by infinitives normally associated with the whole sentence rather than with the noun, as in:

They sold the house to cut down on expenses.

For Bulgarian, the pattern consists of a noun followed by infinite clause, wherein the verb is in the so called *да* form. The following examples illustrate this pattern:

За него е удоволствие да работи там.

Те направиха усилие да свършат работата в срок.

Те положиха много усилия да възстановят разрушената сграда.

G3. nouns + that-clause

Example:

We reached an agreement that she would represent us in court.

These collocations have to be distinguished from nouns followed by a relative clause, as in: *We reached an agreement that would go in effect in a month*.

(in this example *that* can be replaced by *which*:

We reached an agreement which would go in effect in a month)

The correspondent collocational pattern in Bulgarian does not include a conjunctional or prepositional word. The noun is followed by infinite clause, wherein the verb is in the so called *да* form:

Постигнахме споразумение тя да ни представя в съда.

G4. preposition + noun

English examples:	Bulgarian examples:
by accident	*във всеки случай*
in advance	*по всяка вероятност*
	за разлика от

In some cases the noun can be preceded by a modifier.

G5. adjective + preposition

This pattern concerns combinations that occur in the predicate or as verbal clauses.

English examples:	Bulgarian examples:
angry at everyone	*сърдит на всички*
fond of children	*жаден за новини*
hungry for news	*готов за работа*

However, not all English phrases that fall into this pattern, are translated according to the correspondent Bulgarian pattern. For example, the predicate adjective *to be fond of* is translated by *обичам* (without a preposition) rather than *привързан съм към* which is the literal translation (English: *fond of children*; Bulgarian: *обичам децата*).

G6. predicate adjectives + to + infinitive

Examples are:

It was necessary to work.

It was stupid of them to go.

She was ready to go.

She was supposed to work today.

It should be noted that a large number of adjectives when used with adverbs such as *too* and *enough*, can be followed by an infinitive, as in:

He was too absorbed to notice.

She was alert enough to see it.

This kind of combinations are not included in [1]. Other cases that are not included are past participles followed by *to + infinitive* in the meaning of "in order to", as in:

The book was proofread (in order) to eliminate the errors.

There are some G6 adjectives that can be followed by a verb form in -ing:

It was nice working there.

In Bulgarian, the above pattern is translated by *да* clause:

Беше приятно да се работи там.

Беше глупаво от тяхна страна да тръгнат.

G7. adjectives + that-clause

Examples are:
She was afraid that she would fail the examination.
It was nice that he was able to come.

In Bulgarian, the adjectives within the pattern are translated as verbs, and the *that* clause is translated as subordinate *че* clause. The verbs form a collocational pattern, that corresponds to the English pattern G8(l) (see below):

Тя се страхуваше, че няма да издържи изпита.
Беше хубаво, че той можа да дойде.

G8. Verb patterns

This group of collocations consists of nineteen English verb patterns, briefly described below. The following notation is used:

S – subject, V – verb, O – object, C – complement, A – adverbial (when obligatory), V-ing – verb form in *-ing*.

G8-((a)-(c)) The first three groups of collocational patterns concern verbs that can have dative complement.

a. SVO to O / SVOO
Example:
She gave the book to him.
Verbs from this group allow dative movement transformation: *She gave him the book.*

b. SVO to O
An example of this pattern is:
She described the book to the students.
Verbs from this group do not allow dative movement transformation, thus the sentence
**She described the students the book* is incorrect.

c. SVO for O / SVOO
Verbs from this group used with the preposition *for* allow the dative movement transformation:
She bought a book for her daughter.
She bought her daughter a book.

In Bulgarian, the corresponding pattern for the above three groups is:
SVO prep O / S dative pronoun V O
Both versions are acceptable. The dative personal pronoun has to be in its short form.
Examples:
a. *Тя даде книгата на Иван. Тя му даде книгата.*
b. *Тя разказа приказка на децата. Тя им разказа приказка.*
c. *Тя купи книга на дъщеря си. Тя й купи книга.*

It has to be noted that the preposition *на* in **(a)** is ambiguous. It can denote either dative or possessive case.

d. SV prep O / SVO prep O
Verbs that form collocations with a specific preposition belong to this group. Most of them are normally not used without a prepositional phrase.
Examples:
adhere to
base a conclusion on
consist of
In Bulgarian there are a number of verbs which are used in a similar collocational

pattern. Some of them have direct correspondence in English, e.g.

придържам се към
състоя се от

G8-((**e**)-(**k**)) The next seven collocational patterns concern combinations of two verbs or a verb and an -ing form.

e. SV to + infinitive

Examples are:
begin to speak
decide to go
continue to write

These verbs have to be distinguished from the normal usage of verbs, followed by *to + infinitive*, meaning "in order to", as in:
He ran (in order) to catch the train.

This collocational pattern is used in Bulgarian too:
започвам да говоря
решавам да отида
продължавам да пиша

f. SV infinitive

There are some verbs that are followed by infinitive without *to*. They belong to this group. Examples are:
You must work hard.
They let him go.

In Bulgarian this case is not present – verbs take infinitive only with *да (to)*:
Трябва да работиш много.
Те му позволиха да тръгне.

g. SVV-ing

This pattern pertains to verbs are followed by a second verb in -ing:
enjoy reading
keep talking
need improving

In Bulgarian, the **V-ing** part of the pattern is usually translated by *да* infinitive form:
обичам да чета
продължавам да говоря

However, in some cases V-*ing* is translated as noun, preceded by a preposition:
нуждая се от подобрения

h. SVO to + infinitive

This pattern concerns transitive verbs followed by an object and *to + infinitive*:
She asked him to come.
I told him to go.

In Bulgarian, *да* form of the verb is used:
Тя го помоли да дойде.
Казах му да отиде.

i. SVO infinitive

Here transitive verbs are followed by an object and an infinitive without *to*:
He watched them unload the car.
He let the boy go to the park.

In Bulgarian, *да* form of the verb is used:
Той ги видя да разтоварват колата.
Той позволи на момчето да отиде в парка.

j. SVOV-ing

In this pattern, verbs are followed by an object and a verb form in *-ing*:

We found the children sleeping on the floor.

He watched them dancing.

Some of the verbs that are used according to this pattern, can be used also according to the *SVO infinitive* pattern, as in:

He watched them dance.

In Bulgarian, usually *да* form of the verb is used to translate the infinitive in the pattern:

Намерихме децата да спят на пода.

Той ги видя да танцуват.

However, in some cases it is also possible to use the participle of the infinitive:

Намерихме децата заспали на пода.

k. SV possessive V-ing

In this pattern, verbs can be followed by a possessive (pronoun or noun) and a gerund. Typical examples are:

Please excuse my waking you so early.

This fact justifies Bob's coming late.

There is no direct correspondence for this pattern in Bulgarian. Generally, two possible translations are used for the *possessive V-ing* component of the pattern:

a) The possessive is translated as subject and the V-*ing* is translated as verb in a subordinate *че* clause:

Моля те да ме извиниш, че те събудих толкова рано.

b) The V-*ing* form is translated as noun and the possessive is preserved:

Този факт оправдава късното идване на Боб.

l. SV(O) that-clause

In this pattern verbs can be followed by a noun clause beginning with the conjunction *that*. Examples are:

He admitted that they were wrong.

We hoped that the weather would be nice.

Some verbs always take a noun or pronoun object before the *that* clause.

He convinced us that he should go there.

In Bulgarian, the *that* clause is translated as subordinate *че* clause.

Той призна, че не е бил прав.

Надявахме се, че времето ще е хубаво.

Той ни убеди, че трябва да замине.

Той убеди приятелите си, че трябва да замине.

In the Bulgarian pattern the pronoun object usually stands before the verb.

m. SVO *to be* C

This pattern describes transitive verbs followed by a direct object, the infinitive *to be*, and either an adjective, or a past participle, or a noun/pronoun. Usually the same verb can be followed by any of these three forms. Examples of this construction are:

We considered her to be very capable.

We considered her to be well trained.

We considered her to be a competent engineer.

In Bulgarian, the corresponding pattern is **SVO за C**:

Смятаме я за много способна.

Смятаме я за добре обучена.

Смятаме я за добър инженер.

Note, that this pattern concerns verbs that normally take *to be* after the direct object. Verbs that combine freely with infinitives other than *to be* are described in pattern (**h**).

n. SVOC

Here transitive verbs can be followed by a direct object and an adjective or a past participle or a noun/pronoun. Examples are:

She dyed her hair red.
He made his meaning clear.
She had her tonsils removed.
Her friends called her Becky.

Many of these verbs are also used in patterns (**h**) and (**m**).

Some English verbs within this pattern preserve the same construction when translated in Bulgarian, as in:

Тя боядиса косата си червена.
Наричаха я Беки.

However, there are verbs, e.g. *to have, to make, to do*, that are translated using other syntactic patterns, as in the pair:

English: *She had her tonsils removed.*
Bulgarian: *Махнаха й сливиците.*

In this example the subject in the English sentence is not translated, and the past participle (the C component of the pattern) is translated as finite verb (3rd person plural).

o. SVOO

This pattern describes constructions, wherein transitive verbs can take two objects, neither of which can normally be used in a prepositional phrase with *to* or *for*. Examples:

The teacher asked the pupils a question.
The police fined him fifteen dollars.

Note the structural similarity with pattern (**a**), where one of the objects is dative:

She bought him a shirt.
She bought a shirt for him.

There is also structural similarity with pattern (**n**):

They called him a fool.

In the latter sentence *a fool* is a predicate compliment (not a verb’s object).

We have not found a clear-cut rule for the translation of this pattern in Bulgarian. In some cases the pattern is preserved, as in:

Глобиха го петнадесет долара.

In Bulgarian we can also say:

Полицията го глоби с петнадесет долара.

In other cases however a preposition is used :

Учителят зададе на учениците въпрос.

p. SV(O)A

This pattern refers to intransitive, reflexive and transitive verbs that must be followed by an adverbial – an adverb, a noun phrase, a prepositional phrase, or a clause. Examples are:

He carried himself well.
She weights 100 pounds.

Most of the corresponding verbs in Bulgarian also require an adverbial:

Той се държи добре.
Тя тежи 100 паунда.

q. SV(O) *wh*-word

Here verbs are followed by an interrogative word: *how, what, when, where, which, who, why, whether*. Examples are:

He asked how to do it.
She new when to keep quiet.

Some verbs must have an object before the *wh*-word:

She asked us why we had come.

There is a similar pattern in Bulgarian. The *wh*-words correspond to the so called *к*-words – the interrogative pronouns *какво, кой*, the adverbs *как, кога, къде, защо*, and the particle *дали*.

Той попита как да го направи.

Тя знаеше кога да мълчи.

Usually the *wh*-word is followed by *to* + infinitive, in Bulgarian - by *да* + infinitive.

r. S(it)VO to infinitive / S(it)VO that-clause

In this pattern the subject is dummy (*it*):

It surprised me to learn of her decision.

It puzzled me that they never answer the phone.

In the correspondent Bulgarian translation of the first version of the pattern, the object in the English sentence is the subject in the Bulgarian, and the English infinitive is translated as the verb of a subordinate clause:

Учудих се, като научих за нейното решение.

For the second version of the pattern, there is a similar construction in Bulgarian, most often used with a subordinate clause:

Смущава ме това, че те никога не вдигат телефона.

s. SVC (adjective or noun)

Here some transitive verbs are followed by a predicate noun or a predicate adjective.

English examples:	Bulgarian examples:
She became an engineer.	*Тя стана инженер.*
She looks fine.	*Тя изглежда добре.*

The following table summarizes the correspondence between English and Bulgarian grammatical collocations:

G1.	noun + preposition	–	noun + preposition
G2.	noun + to + infinitive	–	noun + 'да' clause
G3.	Noun + that-clause	–	noun + 'да'-clause
G4.	Preposition + noun	–	preposition + noun
G5.	Adjective + preposition	–	adjective + preposition
G6.	Adjective + to + infinitive	–	adjective + 'да'-clause
G7.	Adjective + that-clause	–	verb/adjective + 'че'-clause
G8.	**a.** SVO to O/SVOO	–	SVO prep O/ S dative pronoun VO
	b. SVO to O	–	SVO prep O/ S dative pronoun VO
	c. SVO for O/SVOO	–	SVO prep O/ S dative pronoun VO
	d. SVO prep O/SVOO	–	SVO prep O/SVOO
	e. SV to + infinitive	–	SV + 'да'-clause
	f. SV + infinitive	–	SV + 'да'-clause
	g. SVV-ing	–	SV + 'да'-clause
	h. SVO to + infinitive	–	SVO + 'да'-clause
	j. SVOV-ing	–	SVO + 'да'-clause
	k. SV possessive V-ing	–	* no direct correspondence *
	l. SV(O) that-clause	–	SV(O) 'че'-clause
	m. SVO to be C	–	SVO 'за' C
	n. SVOC	–	SVO
	SVOO	–	SVOO/ SVO dative O (with preposition) / SV dative O (with preposition) O
	p. SV(O)A	–	SV(O)A
	q. SV(O) wh-word	–	SV(O) 'к'-word
	r. it VO to + infinitive	–	* no direct correspondence *
	it VO that-clause	–	VO 'че'-clause
	s. SVC	–	SVC

2.2. Lexical collocations

Lexical collocations, in contrast to grammatical collocations, normally do not contain prepositions, infinitives, or clauses. Typical lexical collocations consist of nouns, adjectives, verbs, and adverbs. An example of an adjective + noun collocation is *warmest regards.* Typical violations of lexical collocability are **hot regards* or **hearty regards.* The authors of BBI Dictionary list 7 types of lexical collocation, described below.

L1. creation and/or activation verbs (verb + noun/pronoun or a prepositional phrase)

Verbs used in this pattern usually denote creation and/or activation. This pattern can be found in Bulgarian too. Examples are:

come to an agreement	–	*постигам съгласие*
make an impression	–	*правя впечатление*
fly a kite	–	*пускам хвърчило*
launch a missile	–	*изстрелвам снаряд*
wind a watch	–	*навивам часовник*

L2. eradication and/or nullification verbs (verb + noun)

Typical examples for English and for Bulgarian are:

reject an appeal	–	*отхвърлям молба*
lift a blockade	–	*вдигам блокада*
break a code	–	*разшифровам код*
revoke a license	–	*отнемам лиценз*
annul a marriage	–	*анулирам брак*

L3. adjective + noun

Many L3 collocations in English correspond to L3 collocations in Bulgarian. Examples:

strong tea	–	*силен чай*
weak tea	–	*слаб чай*
a rough estimate	–	*груба оценка*

English nouns, used attributively, may enter into L3 collocations:

house arrest	–	*домашен арест*
land reform	–	*поземлена реформа*

L4. characteristic action (noun + verb)

Here the verb denotes an action characteristic of the person or thing designated by the noun. Examples (English and Bulgarian) are:

adjectives modify	–	*прилагателните поясняват*
alarms go off (ring, sound)	–	*алармите се изключват (звънят)*
bees buzz (sting)	–	*пчелите бръмчат (жилят)*
bombs explode	–	*бомбите избухват*

L5. ‘part-of’ relation (noun1 *of* noun2)

These collocations may indicate:

a) the larger unit to which a single member belongs. In Bulgarian, these collocations are translated by *noun noun* combination::

a colony of bees	–	*кошер пчели*
a herd of buffalo	–	*стадо бизони*
a bouquet of flowers	–	*букет цветя*

b) the specific, concrete, small unit of something larger, more general:

a bit (piece, word) of advice

an article of clothing
an act of violence
In Bulgarian we cannot give examples of this kind of collocations.

L6. adverb + adjective

English examples:	Bulgarian examples:
deeply absorbed	*силно привързан*
strictly accurate	*дълбоко обиден*
closely acquainted	*силно накърнен*
hopelessly addicted	
sound asleep	

L7. verb + adverb

English examples:	Bulgarian examples:
affect deeply	*спя дълбоко*
amuse thoroughly	*обичам силно*
apologize humbly	*споря горещо*
appreciate sincerely	*уча задълбочено*
argue heatedly	

All types of lexical collocation in English, described in [BBI91], have correspondence in Bulgarian.

3. Statistical methods used to extract collocations

This section presents a brief outline of the statistical methods, underlying the extraction of collocations. The basic idea is to start from a 'node' (or the base of a collocation) and to count the words that appear to the left or to the right of the node. If some words appear more frequently together than chance would predict, they are considered collocations. There are several methods to compute the collocability of words widely employed in the developed systems: Mutual Information, T-score, Z-score, H-score and I-score.

3.1. Mutual Information and T-score

Mutual information (MI) statistic is based on the frequency of co-occurrence of words. With MI, words most closely associated with a given word (or word pattern) are found: the higher the mutual information, the stronger the association. The probability of observing two words together (the joint probability) is compared with the probability of observing the words independently. The joint probability *P(x,y)* can be obtained by counting the number of times that *x* is followed or preceded by *y* (*f(x,y)*) and dividing by *N*, the size of the corpus. The independent probabilities for *x* and *y, P(x)* and *P(y)* respectively, can be obtained by counting the number of observations of *x* and *y* in a corpus, *f(x)* and *f(y)*, and dividing by *N*. If words *x* and *y* are associated, then the joint probability *P(x,y)* will be much larger than *P(x) * P(y).*

The T-score statistic characterizes two words (or word patterns) by finding the words that are more likely to occur with the one than with the other. T-score is very useful for lexicographers to find differences in usage between nearly synonymous words.

3.2.2. Z-score. Z-score is defined generally as "the probability of the item x co-occurring with the items a, b, c, ... being greater than might be expected from pure chance". This method was first proposed by Berry-Rogghe, Codelieve in 1973 [2]. To find the Z-score of a collocate the following statistical data is needed:

N: total number of words in the text;

Fa: frequency of the node A in the text;

Fb: frequency of the collocate B in the text;

Fab: number of co-occurrences of B and A;

S: span size, i.e. the number of items on either side of the node considered as its environment;

P: the probability of B occurring at any place where A does not occur = Fb / (N–Fa);

E: the expected number of occurrences : = p*Fa*S.

The proposed formula is

$$z = (N - E) / \sqrt{(E*q)}\ ,\ q = 1 - p.$$

Z-score estimates the statistical significance of the difference between the observed and the expected frequencies. The formula has been modified and improved in different ways by researches in the field.

3.2.3. H-score and I-score. H-score refers to the Mutual Information score. It equals the probability *P(x,y)* that a word *x* occurs with a word *y* divided by the probability *P(x)* that *x* occurs times the probability *P(y)* that *y* occurs. The probability that *x* occurs equals the frequency of *x* divided by the total number *N* of words in the corpus. Thus we have

$$\text{H-Score}(x, y) = P(x, y) / (P(x)*P(y))$$

I-Score (Inclusion score) is a feature used to refine the standard MI score. I-Score is based on H-Score multiplied by the frequency of the collocation, which gives greater weighting to more frequent collocations.

4. Experiments

Some experiments have been made to determine the I-score and the H-score estimates for the collocability of the Bulgarian prepositions *на*, *от*, and *за*, and some of their collocates that have high I-scores, and for the English prepositions *of* and *on*. The results are organized in 16 tables, described in [5]. Tables 1 through 10 contain data for Bulgarian, tables 11 through 16 – for English. The computations are made for word-forms. The Bulgarian corpus for the experiments contains 87372 words, the English corpus (parallel to Bulgarian) contains 104875 words. Both texts ate tagged up to sentence level. The parallel corpus has been developed under the COPERNICUS project MULTEXT-EAST.

The developed program for the experiments implements the following relations:

I-score is equal to $(f(X, Y)*f(X, Y)*N)/(f(X)*f(Y))$,

H-score is equal to $(f(X, Y)*N)/(f(X)*f(Y))$,

where N is the total number of word-forms in the text, $f(Y)$ is the total number of occurrences of the word in the text, $f(X, Y)$ shows how many times the node (X) and the collocate (Y) appear together in the text.

The formula for the I-score is considered to be more adequate since it takes into greater account the number of the joint occurrences.

The following conclusions might be drawn from the obtained results:

A) Using a small span may lead to neglecting some flexible collocations. Thus in Table 1 the phrase *от другата* has higher I-score than the phrase *от страх*, though intuitively we would consider the latter as a collocational phrase rather than the former one.

This can be explained by the small span, not considering the noun that follows the adjective *другата*. However, using a greater span may be inaccurate, if the text is not parsed, since some words that belong to different constituents, may falsely be considered as used together.

B) Computing the scores with respect to lemmas may lead to better results. Thus Table 2 contains separate entries for the wordforms *върна, върнал, върнаха, върне, върнем*, which decreases the estimate for the verb *да върна*. However, there are cases where the different wordforms have different meaning, when used with a given word, as for example the phrases *от начало* and *от началото* (Table 3).

C) Finally it is obvious that some of the results are dependent on the particular text, and would not be obtained if the computations are carried over another text. Thus *от телекрана*, which has the highest score in Table 3 would not be encountered in another text. Also, while the collocates of the phrase *от болка* seem quite natural semantically, the collocates of *на мисълта* (Table 10) are text specific.

Ideally, very large corpora containing different in genre and style texts should be used if one wants to extract automatically collocations, that are actually used in the language.

References

1. Benson, M., E. Benson, R. Ilson. Introduction. The BBI Combinatory Dictionary of English: A Guide to Word Combinations. John Benjamins Publishing Co, 1991, IX–XXVIII.
2. Berry-Rogghe, Godelieve. The Computation of Collocations and their Relevance in lexical Studies The Computer and Literary Studies. – In: A. J. Aitkin, R.W. Bailey and N. Hamilton-Smith eds., Edinburgh, Edinburgh University Press, 1973, 103–112.
3. Choueka, Y. Looking for Needles in a Haystack or Locating Interesting Collocational Expressions in Large Textual Databases. – In: Proceedings of the RIAO Conference on User-oriented Context Based Text and Image Handling, Cambridge, MA, 1988, 609–623.
4. Mel'cuk, I., Zhlkovsky, A. The Explanatory Combinatorial Dictionary. – In: Relational Models of the Lexicon, M. Evens, ed., CUP, Cambridge, 1988.
5. Mel'cuk, Igor. Collocations and lexical functions. – In: International Symposium on Phraseology, University of Leeds, April 1994.
5. Sinapova, L., D. Dochev. Extracting collocations from english and bulgarian texts. – In: IIT Working Papers IIT/WP-55, Sofia, 1998, 35 p.
6. Smadja, F. Retrieving collocations from text: Xtract. – In: Computational Linguistics. Volume 19, Number 1, 1993, 143–177.

Анализ болгарских и английских колокаций

Лидия Синапова, Данаил Дочев

Институт информационных технологий, 1113 София

(Резюме)

Работа связана с автоматическим обнаружением лексико-семантических отношений в текстах. Исследования являются очень актуальными для специалистов, работающих в области определения характеристик конституентов реальных текстов в формах, которые непосредствено имеют компьютерные применения.

Основная часть исследований представляет граматические и лексические колокации. Демонстрированырезультаты экспериментов при обнаружении колокаций при помощи компьютера в паралельных английских и болгарских текстах.