КОМПЬЮТЕРНЫЙ АНАЛИЗ КОНСТРУКЦИЙ С ГЛАГОЛОМ ПОДДЕРЖКИ В РУССКОМ И ИТАЛЬЯНСКОМ ЯЗЫКАХ

Ключевые слова: глаголы поддержки, оценка компьютерных систем для анализа лингвистических корпусов, семантическая разметка

COMPUTATIONAL TREATMENT OF SUPPORT VERB CONSTRUCTIONS IN ITALIAN AND IN RUSSIAN[1]

Benigni V. (valentina.benigni@uniroma3.it)
Università "Roma Tre", Roma, Italia

Cotta Ramusino P. (paola.cottaramusino@unimi.it)
Università degli Studi, Milano, Italia

We aim at comparing some corpora-based computational resources that enable us to analyse the collocational profiles of the SVCs in both languages. The resources include SketchEngine, which works for both languages, Lexit for Italian and NKRJA for Russian.

The case study focuses on the Italian verb *mettere* followed by a prepositional phrase with the prepositions *in* and *a*, and the corresponding Russian verb *stavit'/postavit'* followed by a prepositional phrase with the prepositions *v* and *na*.

We discuss the options offered by the tools at the syntax-semantic interface. A closer comparison of the three tools shows that they provide different data. We have explored some aspects of the semantic tagging of Lexit and NKRJA and propose an integration of the two tools. It seems that further development of semantic tagging could be helpful in creating Italian-Russian lexicographic resources.

Keywords: Support Verb Constructions, evaluation of corpus-based computer resources for linguistic research, semantic tagging

---

[1] The article is the result of close collaboration between the two authors, whose names are listed in alphabetical order. For academic purposes only, Valentina Benigni is responsible for sections 1, 2.2.1, 2.2.3, 4.1, 4.2 and Paola Cotta Ramusino for sections 1.1, 2.1, 2.2.2, 3.1, 3.2.

1. Introduction

In this paper, devoted to Support Verb Constructions (henceforth SVCs) in Italian and Russian (cf. also Benigni & Cotta Ramusino 2011), we aim at comparing some corpora-based computational resources, that enable us to analyse the collocational profiles of SVCs in both languages. These resources are: SketchEngine, which works for both languages, Lexit for Italian only and NKRJA for Russian only.

The subject of the present case study is the Italian verb *mettere* "to put" followed by a prepositional phrase with the prepositions *in* "in, into" and *a* "to", and the corresponding Russian verb *stavit'/postavit'* followed by a prepositional phrase with the prepositions *v* and *na*.

A closer comparison of the three tools shows that they provide different data. We suggest that the integration of this data could be of great help in creating Italian-Russian lexicographic resources.

1.1. Support Verb Constructions: a case study

In Benigni & Cotta Ramusino 2011 we identified a number of morpho-syntactic criteria (tests) on the basis of which we could distinguish SVCs from free constructions. We then attempted at categorizing SVCs with the Italian verb *fare* "to make" by the following steps:

- first of all, we divided all SVCs into semantic classes within which the SV has the same meaning and combines with a N object that has similar semantic characteristics;
- then we grouped these classes into larger actional classes, according to Vendler's classification.

The data, which needed further processing, was obtained using a CQL query in SketchEngine. One of the most significant results of that classification was to propose relevant parameters for identifying SVCs, i.e., first of all the semantic class of the direct object and, secondly, when dealing with particularly opaque constructions, the semantic and actional class of the whole construction, which could be better treated as a single lexical item (*fare mente locale* "to try to remember", lit. "to make local mind").

2. Computational tools for the Italian language

2.1. SketchEngine

SketchEngine is "a corpus query system incorporating **word sketches**, one-page, automatic, corpus-derived summary of a word's grammatical and collocational behaviour" (http://www.sketchengine.co.uk). The Italian corpus available on SketchEngine is the *itTenTen Corpus* (3,076,908,415 tokens).

The query has been carried out by using the "WordSketch" function (henceforth WS). The WS function presents a list of the grammatical relations the word participates in and provides a list of collocates for each grammatical relation (subject, object, prepositional objects, modifying adverbs…).

As for indirect objects, always introduced by prepositions in Italian, WS allocates simple and compound prepositions to different patterns. We obtained the following profiles for each preposition: **pp_in-x, pp_nel-x, pp_nella-x, pp_nell'-x, pp_a-x, pp_al-x, pp_alla-x, pp_all'-x.**

This function does not account, however, for singular and plural, so that SVCs like *mettere nei guai* "to get into trouble", in which the filler is always plural, are assigned to the singular profile **pp_nel-x**.

Table 1. Prepositional complements of the Italian verb *mettere* according to SketchEngine

mettere in	**MI**	**mettere nel**	**MI**	**mettere nella**	**MI**	**mettere nell'**	**MI**
evidenza	10.68	panno	10.76	condizione	7.19	angolino	7.33
discussione	10.31	guaio	9.67	pentola	6.60	armadio	5.29
luce	9.74	mirino	8.81	ciotola	6.35	impossibilità	5.09
atto	9.53	cassetto	8.12	valigia	6.33	angolo	5.05
scena	9.49	calderone	8.07	mano	6.32	agenda	4.54
moto	9.26	carrello	7.93	padella	6.21	ottica	4.38
risalto	9.20	dimenticatoio	7.78	teglia	6.10	impasto	4.21
campo	9.16	culo	7.55	tasca	6.08	urna	3.76
guardia	9.12	sacco	7.18	bara	5.80	orecchio	2.74
pratica	9.08	frulla tore	7.14	culla	5.66	animo	2.40
mettere a	**MI**	**mettere al**	**MI**	**mettere alla**	**MI**	**mettere all'**	**MI**
disposizione	11.69	riparo	10.53	berlina	10.24	asta	9.21
nudo	9.52	bando	9.36	gogna	9.81	indice	7.00
punto	9.46	corrente	8.87	stretta	9.63	incanto	6.98
fuoco	9.45	volante	8.85	prova	9.55	inseguimento	6.53
segno	9.02	rogo	7.94	corda	8.20	angolo	6.20
rischio	8.97	centro	7.77	frusta	7.46	odg	5.99
confronto	8.89	tappeto	7.36	guida	7.25	occhiello	5.12
frutto	8.50	collo	7.15	spalla	7.15	opera	4.72
prova	7.97	posto	7.12	porta	6.90	ordine	4.30
agio	7.44	fornello	7.08	calcagno	6.90	incasso	4.30

So, the WS function applied to the Italian corpus allows us:

- to single out verb collocates; in particular WS extracts the prepositional collocates to the right of the verb and the noun collocates both left and right of the verb. The collected data contains a substantial amount of noise: for example, one of the most frequent left collocates is the word *santa*, which is part of the multiword noun *santa messa* ("Holy Mass"), where *messa* is a noun and not the past participle of the verb *mettere*;
- to view the contexts in which the token occurs;
- finally, to infer some information about noun gender starting from the compound preposition gender, except for hyphened prepositions like *nell'* and *all'*, which can be either masculine or feminine.

On the other hand, the WS function does not allow us:

- to extract all possible syntactic frames within which the target lemma occurs (for instance, when extracting a prepositional phrase, it does not provide the user with information about presence or absence of a direct object);
- to separate reflexive and non-reflexive verb forms.

2.2. Lexit

Lexit is a corpus-derived lexical resource for the analysis of Italian verbs, nouns and adjectives that extracts distributional profiles at the syntax-semantic interface.

The current version of Lexit contains information gathered from two different corpora: the *La Repubblica* corpus (Baroni *et al.* 2004, about 331 millions tokens) and the Italian section of *Wikipedia* (ca. 152 millions tokens) (Lenci *et al.* 2012: 3713)

The resource has been developed by Lenci at the University of Pisa (Computational lab, Department of Linguistics) and is available at the address http://sesia.humnet.unipi.it/lexit.

The resource allows us to:

- Extract the syntactic frames[2] of a target lemma, going beyond the traditional distinction between argument and adjunct;
- Extract all the fillers of a target syntactic slot;
- Get the semantic classes of the fillers.

2.2.1. Syntactic frames

The syntactic frames of our target lemma *mettere* are extracted and ordered by decreasing values of LMI (Local Mutual Information, used to measure the association between verb and subcategorization frames, frame slots and their lexical fillers) (Lenci *et al.* 2012: 3713).

[2] Lenci refers to them as Subcategorization frames (SCF), i.e. "a pattern of syntactic dependencies headed by the target lemma" (Lenci 2012: 3713).

This is the list of frames individuated by Lexit:

subj#obj#comp-*in*
subj#obj#comp-*a*
subj#comp-*in*
subj#si#inf-*a*
subj#obj#comp-*su*
subj#obj#comp-*sotto*
subj#si#comp-*in*
subj#si#obj#comp-*in*
subj#comp-*sotto*
subj#si#comp-*a*

The extracted data needs further analysis, given that Lexit treats reflexive and non-reflexive forms as the same lemma (*mettersi/mettere*), while allocating them to different frames. Lexit also shows that the same prepositional phrase can occur within different frames, i.e. the **compl-in** is used both with and without a direct object.

By clicking on a specific frame we obtain general information about the lexical fillers in the different syntactic slots of the frame. It is not possible, however, to combine information about a specific argument filler with the data concerning the fillers of other complements.

2.2.2. Lexical fillers

The "slot function" allows extraction of all the lexical fillers of the prepositional phrases compl-*in* and compl-*a*, regardless of the syntactic frame in which they occur.

Table 2 shows the first 20 fillers of **compl-*in*** and **compl-*a*** by decreasing values of LMI.

Table 2. Prepositional complements of the Italian verb *mettere* according to Lexit

mettere in	**LMI**	**mettere a**	**LMI**
discussione	31,804	punto	49,883
moto	27,716	disposizione	45,121
scena	18,489	segno	21,385
evidenza	17,300	repentaglio	10,141
guardia	15,668	prova	9,488
luce	14,715	fuoco	8,927
piede	11,978	riparo	8,202
difficoltà	10,969	confronto	7,083
dubbio	10,125	posto	6,774
pericolo	9,307	bando	6,724
crisi	9,285	nudo	5,305

mettere in	**LMI**	**mettere a**	**LMI**
atto	8,712	asta	4,220
vendita	7,794	lavoro	4,112
ginocchio	7,550	rischio	4,084
campo	7,479	frutto	3,701
condizione	7,034	soqquadro	3,153
pratica	6,213	verbale	3,009
risalto	5,578	servizio	2,873
cantiere	5,372	portafoglio	2,411
conto	5,217	mondo	2,159

The "slot function" shows some shortcomings:

- It is possible to view the lexical set of the most prototypical fillers, but it is not yet possible to see the examples in context (the resource is being implemented);
- Lexit currently does not supply information about the preposition (whether simple or compound): *mettere in scena* vs *mettere nei guai* ("to put on stage" vs "to get into trouble").
- Lexit does not supply information about the number value of the noun, which can be singular or plural: *mettere in un guaio* vs *mettere nei guai* (lit. "to put into a trouble" vs "to put into troubles").

(These last two shortcomings limit the use of the tool in self-learning, because the data needs further checking to get rid of redundant results.)

- At the moment, Lexit does not supply information on the position of the prepositional object or the presence of other lexical items between verb and prepositional phrase (e.g. *mettere in un guaio* vs *mettere in un grosso guaio*; lit. "to put into a trouble" vs "to put into a big trouble"), as the frames are formed by unordered sets of slots representing the syntactic constituents.

2.2.3. Semantic classes

As previously specified, Lexit not only provides the lexical set for the most prototypical fillers for each syntactic slot; it also supplies semantic information on the semantic classes (ordered by LMI) to which the nouns belong.

Nouns are classified into 24 groups[3] corresponding to the 24 "top-nodes" dominating the semantic noun ontology in the Italian section of MultiWordNet (Pianta *et al*. 2002), a multilingual lexical database linked to WordNet and structured in hierarchically organized semantic classes.

Unfortunately, this semantic resource is not fully developed as yet, so that we cannot see which fillers correspond to each semantic class, and the link between filler and semantic class has to be reconstructed by manually checking on MultiWordNet the top-node corresponding to each filler.

---

[3] Animal, Artifact, Act, Attribute, Food, Communication, Knowledge, Body Part, Event, Natural Phenomenon, Shape, Group, Location, Motivation, Natural Object, Person, Plant, Possession, Process, Quantity, Feeling, Substance, State, Time.

This hitherto partially developed semantic classification could be of great help both in teaching and in NLP, as it provides an array of lexical fillers for a given slot.

Therefore, in terms of semantic classification of the fillers, which appears to be Lexit's potentially strong point, we would like to point out that:

- The link between semantic classes and fillers has been carried out automatically without disambiguation;
- The automatic processing does not account for regular polysemy, whereby the same word can be linked to different top-nodes, for instance the filler *posto* "place" is present in the same prepositional **compl-a** with three different syntactic profiles: *mettere* $X_{[+anim]}$ *a posto* "to put sb into place", *mettere* $X_{[-anim]}$ *a posto* "to put sth in order", *mettere* X *al posto* di Y "to put X in the place of Y". In the first and in the second case *posto* "place" is a State, because the entire construction *a posto* "in place" has this meaning, in the third case *posto* is a Location, albeit a metaphorical one.
- In the same way, the words *mano* "hand", *bocca* "mouth", *testa* "head", *piede* "foot", *ginocchio* "knee", are linked to the top-node BodyPart, nevertheless in SVCs such as *mettere in mano* "to put in the hand", *mettere in bocca* "to put in / into mouth", *mettere in testa* "to put in / into head", they acquire the meaning Location, whereas in SVCs as *mettere in piedi* "to set sthg up", lit. "to put on feet", *mettere in ginocchio* "bring to one's knees" they mean Position.
- Although there are different criteria for semantic classification, it seems more appropriate to classify fillers based on productive categories: for example there are many fillers of **compl-in** which could be classified as Location (*prigione* "prison", *galera* "jail", *pista* "track") or as "position" (*fila* "line", *linea* "row", *cerchio* "circle"), but there are no other SVCs, apart from those listed above, with a BodyPart acting as filler of **compl-in** (with the exception of idiomatic expression *mettere la pulce nell'orecchio* "plant a seed of doubt", lit. "to put a flea into the ear").
- Even if the primary meaning of words like *mano, bocca* is BodyPart, their coerced meaning within the SVC is Location or Position, so it would probably be more appropriate to link them to these semantic classes, or, at least, to add a semantic tag which could account for the semantic shift, something like BodyPart — Location, BodyPart — Position.
- Lastly, the semantic classification does not supply information about the degree of idiomaticity of a SVC, which is often due to the desemantization of the filler, because of metonymic or metaphorical processes (for instance *mettere in palio* "to raffle" lit. "to put as a flag". It would seem therefore more useful to tag these constructions as a single lexical item.

3. Computational resources for the Russian language

3.1. SketchEngine for the Russian language

The Russian corpus uploaded on SketchEngine is the *ruTenTen* Corpus (20,162,118,568 tokens).

We used the WS function to carry out a query on the lemma *stavit'* "to put". The tool extracts:

- MI-ordered collocational profiles of the lemma, in particular internal, external arguments and prepositional phrases. The first problem is that for Russian SketchEngine easily mixes up internal and external arguments as it selects by default the noun on the left of the verb as Subject and the one on the right as Object. On the other hand, it distinguishes different internal arguments as type 1 (acc), type 2 (gen) and type 3 (gen part);
- Allomorphs of the same preposition, specifically *v/vo*[4] "in";
- Word contexts;
- The overall frequency and the MI, i.e. information on the collocational nature of the contructions.

Table 3. Prepositional complements of the Russian verb *ставить* according to SketchEngine

ставить в	MI	ставить во	MI	ставить на	MI
тупик	9.91	МХАТе	6.04	кон	8.34
известность	9.21	глава	5.57	огонь	6.79
духовка	8.85	всеуслышание	4.69	подоконник	6.39
упрек	7.46	фрунт	3.33	стол	5.91
холодильник	7.36	двор	3.03	повестка	5.84
вина	7.28	главу	2.88	полка	5.76
укор	6.88	Фронтеры	2.87	пауза	5.72
микроволновку	6.85	ГЛАВУ	2.84	колено	5.68
морозилка	6.32	флоп	2.81	ручник	5.6
кавычка	6.06	гла	2.62	подставка	5.54

3.2. NKRJA — Russian National Corpus

The NKRJA is a corpus of modern Russian incorporating 300 million words. Although it is not a computational resource for the extraction of statistical information on words and constructions, its rich morphological and semantic tagging makes it a useful tool for linguistic research, including identification of collocational profiles and systematic semantic patterns.

In particular, in this section we will discuss the functions supplied by the corpus' semantic tagging.

Semantic tagging is based on the classification system developed for the Lexicograph database from 1992 onwards under the leadership of Paducheva and Rakhilina (http://www.ruscorpora.ru/en/corpora-sem.html).

[4] It should be noted that among the results of *vo* complements quite a substantial amount of noise can be found, and both the overall frequency and MI are very low, so that the data is not fully reliable.

The set of semantic and lexical parameters is different for different parts of speech. Nouns, which are the POS we are dealing with here, are divided into three subclasses: concrete nouns, abstract nouns, and proper names, each one with its own hierarchy of tags.

Lexical and semantic tags are grouped as follows:

1. Taxonomy (a lexeme's thematic class) — for all nouns;
2. Mereology ("part — whole" and "element — aggregate" relationships) — for concrete and abstract nouns;
3. Topology — for concrete nouns;
4. Evaluation — for abstract and concrete nouns.

In the first place, in order to compare NKRJA with other tools, we carried out a search of the verb *stavit'* using morphological tagging. The query was as follows:

first slot
Word: ставить
Distance: from 1 to 1

second slot
Word: в /на
Distance: from 1 to 2
third slot
Gramm. features: noun, accusative

The result of this first query is the following: 1825 tokens with *stavit' na* "to put on" and 1271 with *stavit' v* "to put in/into"; unfortunately, this search does not supply any information regarding the MI of each filler.

We thus proceeded to the second step of the query by adding some semantic features, selecting them among those that showed the highest MI for the corresponding Italian SVC in Lexit. Results are as follows:

natural phenomena = 3
mental sphere = 224
space and places = 247
human body parts and organs = 10

A brief examination of the results reveals a substantial amount of noise: among abstract natural phenomena we find *sneg* "snow" (which is a natural phenomenon, but a concrete one) and *vedro* "bucket", which is neither a natural phenomenon nor an abstract noun, so it becomes clear that the results need further processing.

For this reason we decided to select more generic semantic features. In the second step we selected only abstract nouns and got 1067 results with *stavit' na* and 982 results with *stavit' v*. The choice of abstract nouns is associated with the peculiarities of SVCs, in which abstract nouns show a higher frequency.

The results have been processed manually and ordered by overall frequency.

Table 4 reports the first 20 results for each pattern:

Table 4. Prepositional complements of the Russian verb *ставить* according to NKRJA

ставить на + $N_{ACC[+ABSTRACT]}$	**FQ**		**ставить в+ $N_{ACC[+ABSTRACT]}$**	**FQ**
X[5] место / места	119		вину	84
(X) карту	85		пример	84
место	76		тупик	81
вид	52		положение / положения X	79
стол	44		известность	62
колени	31		X зависимость	46
голосование	29		связь с чем-то	40
(X) огонь / огонек =1	28		один ряд	38
X план	28		упрёк	30
пол	26		заслугу	19
кон	24		место / места	19
счет	15		соответствие	14
очередь	14		условия X	11
полку / полки	11		основу	9
сцену	11		ряд / ряды	8
повестку дня	10		затруднение	7
X почву	9		недоумение	7
обсуждение	9		необходимость	7
работу	9		строку	7
учёт	9		счет	7

We expected a clear-cut and fully reliable result, but on the contrary, we immediately spotted several concrete nouns (*ogon'* "fire", *stol* "table", *pol* "floor", *voda* "water", škol*a* "school", *kotël* "pot").

Moreover, polysemy represents a serious problem for semantic tagging: among the results, there are nouns which acquire an abstract sense if used metaphorically or metonymically (*stavit' na kartu* "to stake", lit. "to put on the card", *stavit' na koleni* "bring/force sb to his knees", *stavit' v tupik* "to lead into a dead end", *stavit' v rjady* "to set in the ranks of"). In particular, semantic tagging does not account for nouns like *mesto* "place", which appears in contexts like *stavit' na X*[6] *mesto* "to put in X place" */ stavit' na mesto kogo-to/***čego-to** "to put in sb's/sth's place", where the noun means *place* and the verb maintains its primary meaning of motion verb, but also in constructions like *stavit' kogo-to na mesto* "to put sb into place" and *stavit'* **čto**-*to na mesto* "to put sth in order", where the noun exhibits the more abstract meaning of Position and the verb undergoes desemantization and functions as support verb.

5 X indicates the presence of a modifier in that syntactic position.

6 Cf. footnote 5.

NKRJA pros: the query system allows the user:

a) to query the different syntactic frames in which the predicate appears, but does not provide a list of them;

b) to find, by means of the *distance* function, more lexical items (adjectives, adverbs, and so on) within the frame, so that it is possible to obtain collocates that occur far away from their prepositional head (*stavit' v polnuju/polnejšuju/prjamuju zavisimost'* lit. "to put in full/fullest/direct dependence) and multiword expressions (*stavit' na pervyj plan* "to put in the foreground").

c) to refine the search by inserting more semantic features, although this function is not totally reliable, as shown by the examples above.

NKRJA cons: the query system does not provide the user with:

a) a list of all the possible frames in which the target lemma may appear;

b) the overall frequency and MI data of the filler, so that we don't get any information about the association of the analyzed words;

c) a semantic classification of the filler (what Lexit is trying to do): in other words we can verify the absence/presence of fillers with certain semantic features, but the tool does not tell us which semantic classes occupy a given syntactic slot. We can obtain this information by manually processing the data.

4. Three different computational resources: final remarks

4.1. The collocational profile

In this paper we discussed how three different computational resources can provide information for linguistic research. We chose SVCs for testing purposes, since for this kind of structures both morpho-syntactic and semantic features are relevant (and should be identified by the tools). As far as morpho-syntactic features are concerned, both Lexit and SketchEngine, although to different degrees, extract the syntactic and collocational patterns of the target verb. The data, when subject to further manual processing, results in a selection of SVCs. In particular, Lexit extracts a more exhaustive list of the syntactic pattern of a target lemma. On the other hand, NKRJA allows a query of the syntactic frames within which the predicate occurs, but does not identify or list them all.

4.2. The semantic profile and semantic tagging

With respect to semantic classification, which could introduce significant advantages both for the creation of lexicographic resources and for self-learning and teaching activities, we should make some observations on hitherto unsolved problems.

In particular, with respect to Lexit we were confronted with:

- the absence of a link between semantic class and actual filler;
- a choice of semantic classes which is not always functional in creating a classification with a predicting character (a case in point being the BodyPart tag);
- the association of the filler with top-nodes that were too high in the hierarchy.

On the contrary, NKRJA provides the user with a very refined semantic tagging, although:

- it is not fully satisfactory yet, especially when the manual disambiguation of homonimy has not been carried out;
- it requires an "expert" use of the query system.

Moreover, at present, both tools fail to reflect the polysemy produced at the *parole* level, where lexemes may acquire new senses, by metonymical or metaphorical processes (a phenomenon usually called polysemy (Apresjan 1974), coercion (Pustejovsky 1993), deferred reference (Nunberg 1995)).

NKRJA seems to suggest a partial solution to this problem, as it allows, once you obtain the query results, to click on a given token and check all the assigned tags. Semantic tagging is organized in a two-tier system, which includes *main* and *secondary* semantic features, which accounts for both the connotation level and the semantic shift, as shown by the semantic tags for the lemma *tupik*:

Semantics main	r:concr, t:space
Semantic shifts	ev:neg, r:abstr, r:concr, t:space

This kind of two-tier semantic tagging suggests that it could be possible to reflect this hierarchy in the query system.

Finally, we observed that information on preposition semantics could be included in the semantic tagging. In Italian, for instance, *in* can have different meanings:

- it describes a movement towards sth (≃into) in SVCs like *mettere in testa* "to put in/into head",
- it can refer to the way sth is done (≃how) in SVCs like in *mettere in ginocchio* "bring to one's knees".

Both patterns are regular and productive: *in* (≃into) occurs also in other SVCs, like *mettere in tasca* "to put in pocket", *mettere in galera* "to put in jail", whereas *in* (≃how) occurs in SVCs like *mettere in difficoltà* "to hinder sb", lit. "to put in difficulty", *mettere in pericolo* "to put in danger".

At the same time, we observe the same regularity in Russian; *v* means:

- "towards" or "into" sth, in SVCs like *stavit' v tupik* "to lead sb into a dead end", *stavit' v kavyčki* lit. "to put into inverted commas";
- "in which way", "how" sth is done, in SVCs like *stavit' v rjad* lit. "to put in line", *stavit' v parallel'* "to put in parallel", *stavit' v zatrudnenie* "to hinder sb", lit. "to put in difficulty";

- “as, like” in SVCs like *stavit’ v primer* “to cite as an example”, *stavit’ v osnovu* “to assume sth as a basis”.

We thus suggest that further research and implementation of these computational resources should focus first of all on semantic tagging: the link between fillers and semantic classes, as in Lexit, and the semantic tagging by means of a two-tier system adopted by NKRJa, seem to be useful devices in clarifying polysemy. Moreover, as previously observed, prepositions show a high degree of semantic regularity and distributional similarity, even cross-linguistically, and we maintain that this kind of tagging which takes into account large lexical contexts should be implemented.

References

1. *Apresjan Ju. D.* (1971), Regular polysemy [O reguljarnoj mnogoznachnosti], Bulletin of the Academy of Sciences of the USSR: Section of. Language and Literature [Izvestija Akademii Nauk SSSR: Otdelenie literatury i jazyka], T. XXX, Nauka, Moskva, pp. 509–523.
2. *Baroni M., Bernardini S., Comastri F., Piccioni L., Volpi A., Aston G., Mazzoleni M.* (2004), Introducing the ”La Repubblica” corpus: A large, annotated, TEI (XML)-compliant corpus of newspaper Italian, Proceedings of LREC 2004, Lisboa, pp. 1771–1774.
3. *Benigni V., Cotta Ramusino P.* (2011), Italian constructions with support verb *fare* in comparison with russian [Italjanskie konstruktsii s glagolom podderzhki *fare* v sopostavlenii s russkin jazykom],
4. *Computational* Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2011” [Komp’iuternaia Lingvistika i Intellektual’nye Tehnologii: Trudy Mezhdunarodnoj Konferentsii “Dialog 2011”], Bekasovo, pp. 68–83.
5. *Lenci A., Lapesa G., Bonansinga G.* (2012), LexIt: A Computational Resource on Italian Argument Structure, available at: www.lrec-conf.org/proceedings/lrec2012/pdf/622_Paper.pdf.
6. *Nunberg G.* (1995), Transfers of meaning, Journal of Semantics, Vol. 12 (2), pp. 109–132.
7. *Pianta E., Bentivogli L., Girardi C.* (2002), Multiwordnet: developing an aligned multilingual database, Proceedings of the First International WordNet Conference, Mysore, pp. 293–302.
8. *Pustejovsky J.* (1993), Type Coercion and Lexical Selection, in Pustejovsky J. (ed.), Semantics and the Lexicon, Kluwer Academic Publishers, Dordrecht, pp. 73–94.