

# ОПИСАНИЕ РУССКИХ КОНСТРУКЦИЙ С ВНЕШНИМ ПОСЕССОРОМ В СИСТЕМЕ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

**Богданов А. В.** (bidon@inbox.ru),  
**Леонтьев А. П.** (taonick@yandex.ru)

АВВУУ, Москва, Россия

**Ключевые слова:** внешний посессор, автоматическая обработка, машинный перевод, семантико-синтаксический интерфейс

## DESCRIPTION OF THE RUSSIAN EXTERNAL POSSESSOR CONSTRUCTION IN A NATURAL LANGUAGE PROCESSING SYSTEM

**Bogdanov A. V.** (bidon@inbox.ru),  
**Leontyev A. P.** (taonick@yandex.ru)

ABBYU, Moscow, Russia

The paper shows how Russian external possessor constructions are treated in the АВВУУ Comprendo ® system. The specific tasks of our system require that sentences with external possessor constructions be considered as synonymous with those with internal possessors. Accordingly, the semantic structure is generated in such a way that the possessor, whether external or not, and the possessum form a single constituent. This is not the case with the syntactic structure because there is much evidence that the external possessor is not syntactically dependent on its possessum. The semantic and syntactic structures of external possessor constructions are not isomorphic so we have to apply a syntax-semantic interface to derive one from the other. We show that two different kinds of interface must be used. For constructions with strong lexical restrictions we use a special normalization module while leaving the syntactic description relatively simple. In contrast, constructions with fewer lexical restrictions require a more sophisticated syntactic description where movements are postulated.

**Key words:** external possessor, natural language processing, machine translation, syntax-semantics interface

*In memory of Aleksandr Yevgenyevich Kibrik*

Our research is a part of work on the natural language processing system ABBYY Compeno<sup>®</sup>. The main task of this system is to convert the input text into a semantic structure that is a tree where nodes are concepts and arcs are relations between these concepts. Further information on the project see in [Anisimovich et al. 2012].

Our system is not designed for processing only one specific language. We claim that the approaches we use in our system are applicable to any natural language. Hence the semantic trees we get must not contain any language specific phenomena and therefore the semantic structure for a certain sentence in Russian must be exactly the same as that of its full analogue in English. That can be illustrated by the example (1a) and its full analogue (1b)

- (1a) Мальчик съел хлеб. *Russian*  
 [[<sub>Subject</sub> Мальчик] есть [<sub>Object\_Direct</sub> хлеб]]
- (1b) The boy ate the bread.<sup>1</sup> *English*  
 [[<sub>Subject</sub> [<sub>Article</sub> The] boy] eat [<sub>Object\_Direct</sub> [<sub>Article</sub> the] bread]]

The syntactic structures of these sentences which are given below them are although slightly but different. English has articles while Russian does not. That is why in the semantic structure (1c), which corresponds to the both of the sentences, articles are not represented as separate nodes

- (1c) [[<sub>Agent</sub> BOY<sup>2</sup>] TO\_EAT [<sub>Object</sub> BREAD]]

Such a guideline may be good for making semantic structures truly universal. Nevertheless it can lead to some problems. Namely when the sentences we are inclined to consider as synonymous have considerably different syntactic structures. This is often the case when some syntactic phenomenon is language specific and cannot be syntactically reduced to a more universal one. Here we have got no other way out but to make the syntactic and semantic structures not isomorphic and, since one cannot be derived from the other by regular procedures, a special syntax-semantics interface will be needed. That is exactly the approach we apply regarding Russian external possessor constructions

<sup>1</sup> We use <sub>lower index</sub> to mark up semantic or syntactic relations. Square brackets [] mark constituent borders.

<sup>2</sup> Words written in CAPITALS mark semantic concepts. The concepts in our system are mostly named in English. But that is so only for convenience and does not mean that the concepts can not correspond to words in other languages

## Problems in description of external possessor constructions

External possessor constructions<sup>3</sup> are generally postulated in sentences (2) to (4)

- (2) **У меня** болит **шея**.  
on me aches neck  
*My neck aches.*
- (3) Мальчик наступил **девочке** **на ногу**.  
boy stepped girl.Dat on foot  
*The boy stepped on the girl's foot.*
- (4) Мальчик поцеловал **девочку** **в губы**.  
boy kissed girl.Acc in lips  
*The boy kissed the girl's lips.*

These sentences share one thing in common, namely the semantic equivalence<sup>4</sup> with sentences like (5) to (7) where the nouns in bold type are replaced with a complex NP headed by one of these nouns (further possessum) and the other (further possessor) is its genitive modifier or possessive adjective.

- (5) **Моя** **шея** болит.  
my neck aches
- (6) Мальчик наступил **на ногу** **девочки**.  
boy stepped on foot girl.Gen
- (7) Мальчик поцеловал **губы** **девочки**.  
boy kissed lips girl.Gen

Thus the semantic structures for (2) to (4) have to be exactly the same as for (5) to (7). But this is not the case for their syntactic structures. The genitive NPs always form a single constituent, but in the external possessor constructions there is some evidence that the possessor is mostly a modifier of a verb and not that of the possessum. In (8a) it is shown that in external possessor constructions possessor and possessum can be on the different sides of the verb. While for the genitive NP such an order is impossible.

<sup>3</sup> More on external possessor see: [Kibrik 2003], [Kibrik et al. 2006], [Payne, Barshi 1999].

<sup>4</sup> There is a lot of evidence that the sentences with the external possessor are not truly semantically equivalent to the sentences with genitive NPs. For example see [Shibatani 1994], [Podlesskaya Rakhilina 1999], [Brykina 2005]. Still in our project we have to neglect it for the reason given in next paragraphs.

- (8a) **У Васи** вчера сломалась **машина**.  
 on Vasya yesterday broke car  
*Vasya's car broke down yeasteday.*
- (8b) **?Васина/Васи** вчера сломалась **машина**.  
 Vasya.Poss/Vasya.Gen yesterday broke car

In the external possessor constructions the possessum can be a pronoun whereas the corresponding genitives NP is impossible.

- (9a) Голова, спрашиваешь? **Она** **у него** круглая.  
 it on him round  
*(You asked about the head). He has it round.*
- (9b) **\*Его** **она** круглая.  
 his it round

Thus we have to admit that in the external possessor constructions the semantic structure is not isomorphic to the syntactic and cannot be derived from it automatically. Further we shall describe the approaches we used to solve this problem.

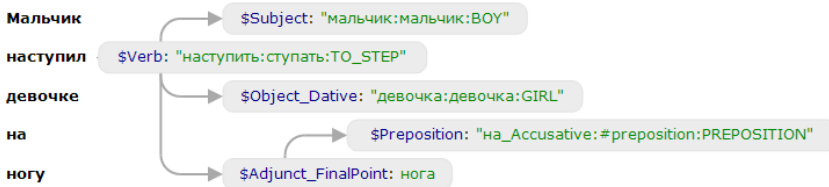
### The first approach: simpler syntax, more complicated semantics

The first approach used in our system may be described like this: we no longer assume that the sentences with external possessor are semantically equivalent to the sentences with genitive NPs and hence their semantic structures are not obliged to be the same. In that case nothing prevents us from making the semantics and the syntax isomorphic that is making the external possessor a modifier of verb in the sentences like (3).

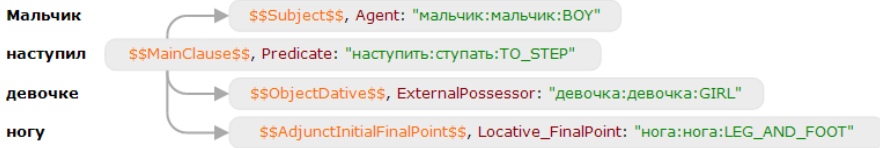
There is only one problem that remains: what semantic role must be assigned to the external possessor. This is a well known issue in theoretical linguistics discussed for example in [Podlesskaya, Rakhilina 1999] and [Shibatani 1994]. It is often proposed that a special extra-thematic role (like Affectee) must be postulated. This is the approach we used for the sentences like (10) where the external possessor is an NP in dative.

- (10) Мальчик наступил **девочке** **на ногу**.  
 boy stepped girl.Dat on foot  
*The boy stepped on the girl's foot.*

The semantic (Picture 1) and the syntactic (Picture 2) structures we proposed are as follows



Picture 1



Picture 2

The NP in dative (*девочке*) is a modifier of verb in the syntactic structure. In the semantic structure it remains the same: the concept correspondent to this NP takes the role ExternalPossessor and is a modifier of the predicate. No complicated interface is needed and the syntax remains quite simple. Which is the most prominent argument for this approach.

Nevertheless we have to face another kind of problem. The extra thematic role we use must have a strange peculiarity: in some languages (e.g. English) this role is never overtly expressed. In that case we will get different semantic structures for the Russian sentence (10) and its English translation. That is not very good for machine translation, which is one of the tasks of our system.

However this problem can be easily solved by a transfer module that transforms the semantic structure in Picture 2 into the suitable structure for the English sentence, and vice versa. A part of this module is presented below (Picture 3).

```

«TO_WALK»
[
  ExternalPossessor: y,
  Locative_FinalPoint: loc «PART_OF_ORGANISM»
]
=>
[
  loc
  [Whole: y]
];
    
```

Picture 3

Capitalized text in quotation marks («TO\_WALK») marks a semantic concept. Sequence of symbols ended with a column (Whole:) marks a semantic role, loc and y are variables used to mark nodes, and the arrow => divides the input structure from what we get as a result. The sense of the rule is as follows: if a concept of the semantic

class TO\_WALK has a dependent with the role ExternalPossessor and another dependent of the semantic class PART\_OF\_ORGANISM with the role Locative\_FinalPoint, the first dependent must be reattached to the second with the role Whole.

This transformation can easily be described due to the special characteristics of the external possessor in dative. This construction is possible only for certain semantic classes of verbs and the semantic roles that are available for the possessor and the possessum have very precise restrictions. That helps us to define the semantic role of the possessor in the output structure. We know that the range of possible concepts in that context is very restricted and so is the range of the possible relations between them.

The range of possible syntactic positions in the input structure is also restricted. Thus we may be sure that our transformation will apply to all the sentences where it must and on the same time will never apply where it must not.

## The second approach: simpler semantics, more complicated syntax

The previous approach does not, however, fit with the most frequent construction with an external possessor, namely the construction with the preposition *y* (see (2)). Unlike a dative one, this construction is restricted with neither a class of a verb nor a syntactic position of possessum, nor a possible semantic relation. It makes impossible the application of semantic normalization rules like that in Picture 3.

For description of this construction we had to use a special semantic-syntactic interface. There are several candidates for this role in our system, but the movement was declared as most suitable. It's a mechanism which is typically used in our system for description of complex cases of communicative dislocation and other phenomena associated with violation of projectivity of syntactic structure.

Let's consider the architecture of the system and the mechanism of movement in detail.

In general case the analysis module works as follows: syntactic structure is being built; it means that every word form of the input text takes some syntactic position of some parent. Further the transition from syntactic structure to the semantic one follows as a result of which each arc between a parent and a child is interpreted — each child gets semantic role related to its parent. The process of a switch from syntactic positions to semantic roles is possible because each lexeme has a diathesis description — a list of correspondences between syntactic positions that can connect to it and their semantic roles. Let's consider an example.

(12a) Мальчик дал девочке яблоко.

(12b) [[<sub>Subject</sub> мальчик] дать [<sub>Object\_Dative</sub> девочка] [<sub>Object\_Direct</sub> яблоко]]

(12c) [[<sub>Agent</sub> BOY] TO\_GIVE [<sub>Possessor</sub> GIRL] [<sub>Object</sub> APPLE]]

(12d) *The boy gave the girl an apple.*

The input Russian text is represented in (12a). In (12b) one can see the syntactic structure of the input text. All the arcs in syntactic structure are marked with syntactic positions. Semantic structure is represented in (12c) where the arcs are marked with semantic roles and the lexemes are replaced with their semantic concepts. For example, the node GIRL receives its semantic role of Possessor because the lexeme *дать* has among others the diathesis <Object\_Dative — Possessor>.

The analysis module is organized such a way that syntactic structure must be projective. Therefore in all cases of real not-projectivity in an input text we have to use a special mechanism — movement. It works as follows: in syntactic structure the word form can connect to its parent into special moved position which does not have any diathesis (does not correspond to any semantic role), but has a rule of movement assigned to it. The rule of movement consists of a path via the syntactic tree from a parent of the moved position to the initial position of movement (pro). The paths can be arbitrarily deep but the moved position (the target of movement) always C-commands the pro. In case of successful search of position for a pro, corresponding to one of the paths, the target of movement takes the position of the pro in semantic structure. At the same time the position of a target erases.

Let's consider the example.

(13a) *Девочке мальчик хочет  
дать яблоко.*

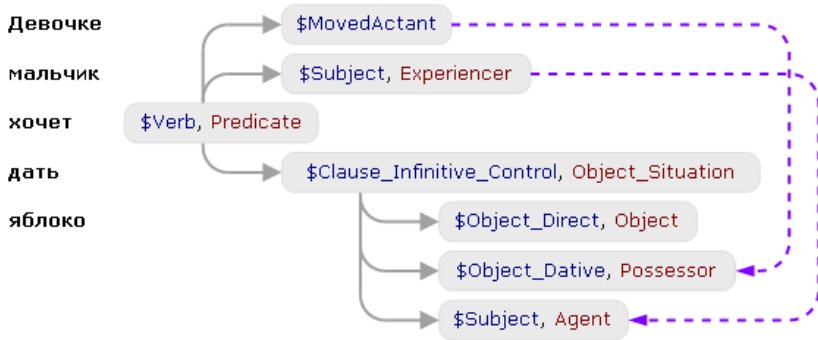
(13b) [[<sub>MovedActant</sub> девочка] [<sub>Subject</sub> мальчик] хотеть  
[<sub>Clause\_Infinitive\_Control</sub> дать [<sub>Object\_Direct</sub> яблоко]]]

(13c) [[<sub>Experiencer</sub> BOY] TO\_WANT [<sub>Object\_Situation</sub> TO\_GIVE  
[<sub>Object</sub> APPLE] [<sub>Possessor</sub> GIRL]]]

(13d) *The boy wants to give the girl an apple.*

An input text with non-projective structure is represented in (13a). Syntactic structure where *девочка* is connected to *хотеть* into moved syntactic position *MovedActant* is given in (13b). Semantic structure where GIRL is connected to its real parent TO\_GIVE and has a semantic role Possessor is shown in (13c).

This structure can be seen at Picture 4.



Picture 4

In Picture 4 one can see schematically shown structure of (13). The syntactic positions are shown with the symbol “\$” in blue color. Semantic roles are shown in red color. So-called non-tree links are shown with the dotted arrows, one of which shows movement of *девочке* from the MovedActant position to Object\_Dative position under *дать*. The second arrow shows a link of control between a subject of a matrix verb and a subject of infinitive.

For the shown movement to be possible in the movement rule there is a path like this:

**Clause\_Infiniteive\_Control.Object\_Dative;**

In accordance with this path the analysis module restores a pro in Object\_Dative position under a constituent in the position Clause\_Infiniteive\_Control which has the same parent as the moved position.

In case of external possessor the position with *y* is the target position of movement, which connects to a verb without diathesis (without semantic role). In movement rule there is a path like this:

**(Subject | Actants).Object\_Indirect\_Y;**

In accordance with this path the pro restores in the position Object\_Indirect\_Y under subject or some actants like in example below.

(14a) У меня женился брат.

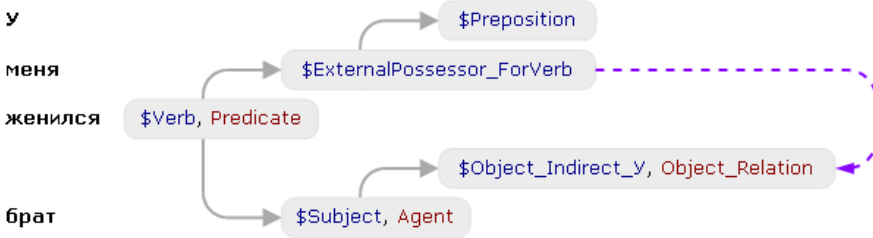
(14b) [[<sub>ExternalPossessor\_ForVerb</sub> я] жениться [<sub>Subject</sub> брат]]

(14c) [TO\_MARRY [<sub>Agent</sub> BROTHER [<sub>Object\_Relation</sub> I]]]

(14d) My brother has married.



In syntactic structure (14b) external possessor *у меня* is connected to the verb *жениться*, whereas in semantic structure (14c) it is connected to its possessum BROTHER having semantic role *Object\_Relation*. It can be seen at Picture 5.



Picture 5

Picture 5 shows that *у меня* is connected to the verb in the moved position without semantic role and the movement to the position *Object\_Indirect\_Y* under the subject can also be seen. Note that the preposition is connected to the noun and does not have a semantic role because it takes the so-called grammatical syntactic position.

Thus in the semantic structure we have an original external possessor under its possessum with necessary semantic role. Using of movement allows us to solve a problem of not-isomorfism of a syntax tree and a semantic tree.

Therefore, although the Russian external possessor constructions have some common characteristics, it may be plausible to apply different approaches while describing them in a system of natural language procession.

## References

1. *Anisimovich K. V., Druzhkin K. Ju., Minlos F. R., Petrova M. A., Selegey V. P., Zuev K. A.* (2012) Syntactic and semantic parser based on ABBYY Comprepro linguistic technologies, Computational linguistics and intellectual technologies vol. 11 (18).Izdatel'stvo RGGU, Moscow
2. *Brykina M. M.* (2008) The linguistic means of coding of possessivity (basing on a corpus research of Russian) [Yazykovye sposoby kodirovaniya posessivnosti (na materiale korpusnogo issledovaniya russkogo yazyka)], PhD thesis, Moscow.
3. *Kibrik A. E., Brykina M., Leontiev A., Leontieva A. L.* (2007) A single phenomenon-oriented corpus: benefits and problems, Second international conference "Grammar and Corpora". Abstracts, Ustav pro jazyk cesky AV CR, Liblice.
4. *Payne, D. L., Barshi, I.* (eds.) (1999) External Possessor. John Benjamins, Amsterdam/Philadelphia.
5. *Podlesskaya, Vera I., Rakhilina, Ekaterina V.* (1999) External Possession, Reflexivization and Body Parts in Russian, in External Possessor. John Benjamins, Amsterdam/Philadelphia:
6. *Shibatani, M.* (1994) An integrational approach to possessor raising, ethical datives and adversative passives, Proceedings of the 12th Annual meeting of the Berkeley Linguistic Society. Berkeley
7. *Apresyan Yu. D.* (2006) Types of correspondence between semantic and syntactic actants [Tipy sootvetstviya semanticheskikh i sintaksicheskikh aktantov], Problems of typology and general linguistics [Problemy tipologii i obshchey lingvistiki]. St. Petersburg., pp. 15–27
8. *Kibrik A. E.* (2000) External possessor as a result of valence splitting [Vneshniy posessor kak rezultat rasshepleniya valentnosti], A word in a text and dictionary. Papers presented to academician Yu. D. Apresyan at his 70th birthday [Slovo v tekste i slovare. Sbornik k semidesyatiletiju akademika Yu. D. Apresyana ]. Languages of Russian culture, Moscow.
9. *Kibrik A. E., Brykina M. M., Khitrov A. N., Leontyev A. P.* (2006) Russian possessive constructions in the light of corpus statistics research [Russkiye possessivnye konstrukcii v svete korpusno-statisticheskogo issledovaniya]. Problems of linguistics [Voprosy yazykoznaniya] 2006, vol. 1, Nauka, Moscow.