Description of russian external possessor consruction in a natural language processing system

Alexey Bogdanov (bidon@inbox.ru)
Alexey Leontyev (taonick@yandex.ru)
ABBYY, Moscow

- Abbyy Compreno® is a multitask natural language processing System.

- Abbyy Compreno® is a multitask natural language processing System.
- It is multilingual, not Russian or English oriented.

- Abbyy Compreno® is a multitask natural language processing System.
- It is multilingual, not Russian or English oriented.
- The system derives a semantic structure from the input text. The nodes of the structure are semantic concepts and the arches between them are relations

- Abbyy Compreno® is a multitask natural language processing System.
- It is multilingual, not Russian or English oriented.
- The system derives a semantic structure from the input text. The nodes of the structure are semantic concepts and the arches between them are relations.
- Synonymous sentences (whether in a single or different languages) must correspond to one and the same semantic structure.

- Abbyy Compreno® is a multitask natural language processing System.
- It is multilingual, not Russian or English oriented.
- The system derives a semantic structure from the input text. The nodes of the structure are semantic concepts and the arches between them are relations.
- Synonymous sentences (whether in a single or different languages) must correspond to one and the same semantic structure.
- Semantic structures must not contain language-specific phenomena.

Russian external possessor constructions.

possessor   possessum

(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.
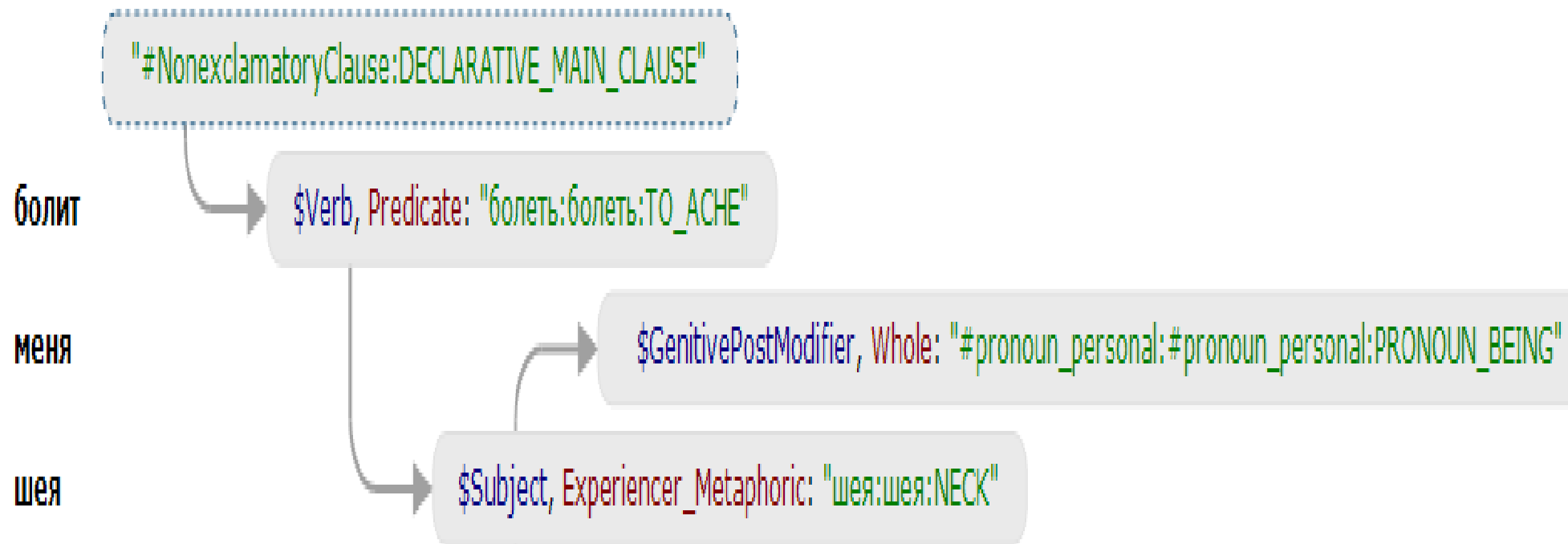
Where is the problem?

Our system must analyze the sentences with external possessor constructions as semantically equivalent to those where there is a single NP.

These sentences must correspond to one and the same semantic structure

(2) **У меня** болит **шея**.
    on me   aches   neck
    My neck aches.

(5) **Моя шея** болит.
    my   neck   aches
    My neck aches.

External possessor construction is a language specific phenomenon. It need not be represented in the semantic structure

But in the external possessor constructions the possessor and the possessum are evidently independent.

Lack of linear contact

External possessor - possible
(8a)   **У Васи** вчера   сломалась   **машина**.
        on Vasya yesterday   broke   car
        Vasya's car broke down yeasteday.

Not external possessor - impossible
(8b)   ? **Васина/Васи**   вчера   сломалась **машина**.
        Vasya.Poss/Vasya.Gen   yesterday   broke
        car

Pronominalization

External possessor - possible

(9a)   (Голова, спрашиваешь?)
       Она   у него   круглая.
       it   on him   round
       (You asked about the head). He has it round.

Not external possessor - impossible

(9b)   * **Его   она**   круглая.
       his   it   round

The syntactic strucuture is not isomorphic to the semantic structure.

We have no way out but to create a certain syntax-semantics interface

The first approach: simpler syntax, more complicated semantics

-First step: derive a semantic structure which is correspondent to the syntactic one. I. e. the possessor is a dependent of the verb, but not of the possessum

-Second step: apply a specific normalization rule that reassigns the possessor to the possessum
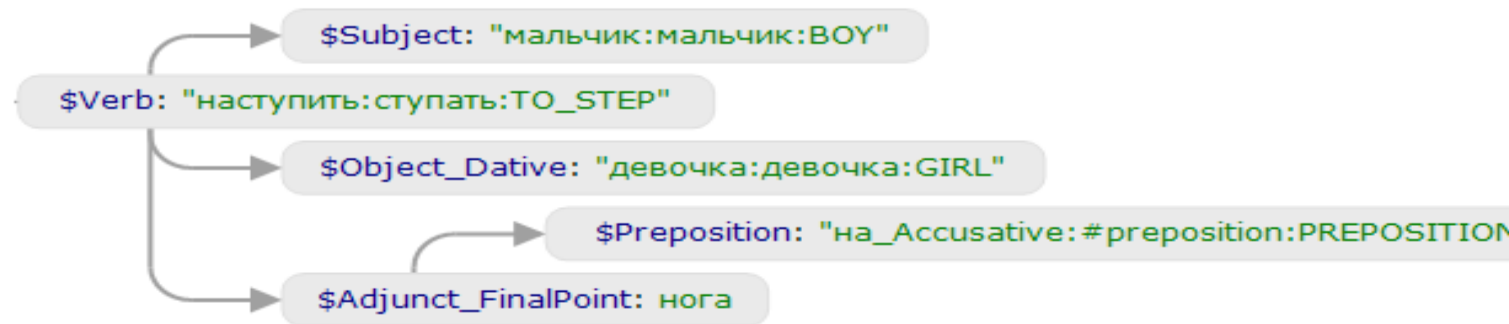
The dative external possessor construction

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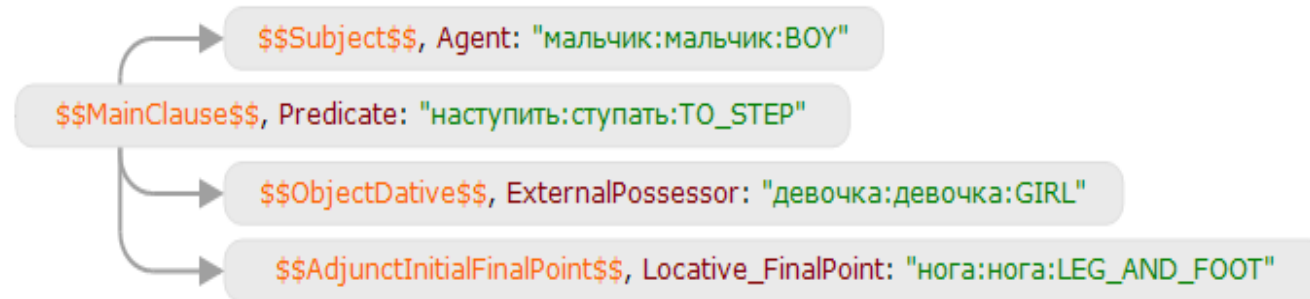   *The boy stepped on the girl's foot.*

The first step
Syntactic structure

Мальчик	→	$Subject: "мальчик:мальчик:BOY"
наступил	$Verb: "наступить:ступать:TO_STEP"	
девочке	→	$Object_Dative: "девочка:девочка:GIRL"
на	→	$Preposition: "на_Accusative:#preposition:PREPOSITION
ногу	→	$Adjunct_FinalPoint: нога

Semantic structure after the first step

Мальчик	→	$$Subject$$, Agent: "мальчик:мальчик:BOY"
наступил	$$MainClause$$, Predicate: "наступить:ступать:TO_STEP"	
девочке	→	$$ObjectDative$$, ExternalPossessor: "девочка:девочка:GIRL"
ногу	→	$$AdjunctInitialFinalPoint$$, Locative_FinalPoint: "нога:нога:LEG_AND_FOOT"

Normalization rule

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

Advantage: a simpler syntax

Disadvantage: we have to guess what node the possessum is and what relation we have to postulate between the possessor and the possessum.

Why we neglect it: the dative construction is only possible with restricted range of verbs that impose strong semantic restrictions on its arguments. Nothing but the external possessor construction is possible in the context described in the rule above.

The second approach: simpler semantics, more complicated syntax

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.

(12a) Мальчик   дал   девочке
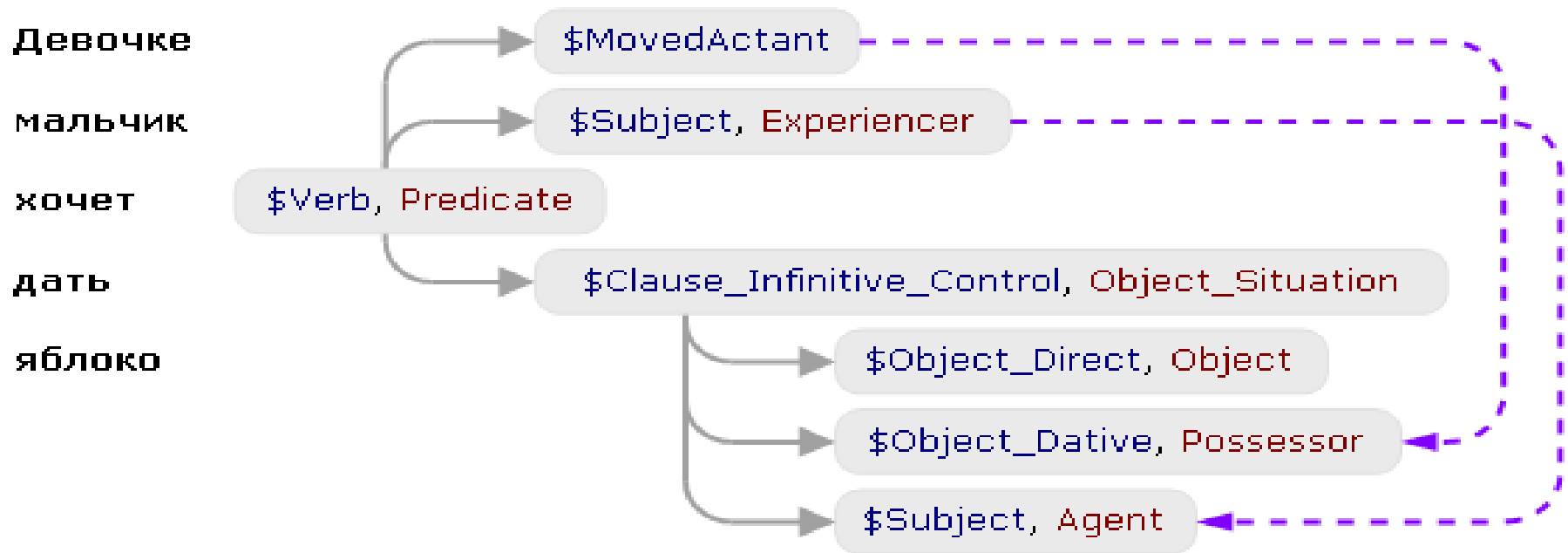   яблоко.

(12b) [[Subject мальчик]   дать   [Object_Dative девочка]

      [Object_Direct яблоко]]

(12c) [[Agent BOY]   TO_GIVE   [Possessor GIRL]

      [Object APPLE]]

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

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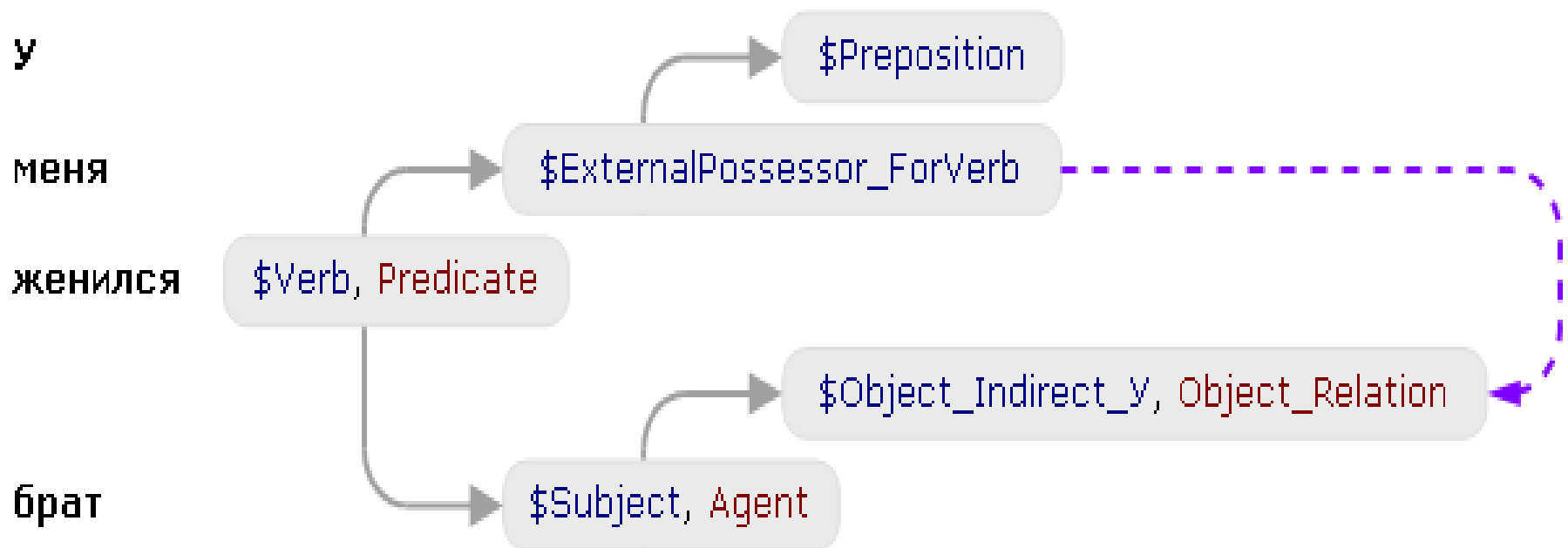
(13)   *Девочке мальчик хочет дать яблоко.*

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

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_У;

Picture shows that *у меня* is connected to the verb in the moved position without semantic role and the movement to the position Object_Indirect_У 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.