

# THE CASE OF RUSSIAN SUBJECT PRO IN MACHINE TRANSLATION SYSTEM

**Alexey Bogdanov**

abogdanov@abbyy.com

ABBYY, Moscow, Russia

**Irina Gorbunova**

igorbunova@abbyy.com

ABBYY / Russian State University for the Humanities, Moscow, Russia

# Floating quantifiers

According to Babby, floating quantifiers are “adjectives that adjoin to VP and agree in case, gender and number with the subject of the minimal clause containing them”. This can be illustrated by examples (1), (2), (3) and (4) (floating quantifier is marked nominative in (1), (3) and (4), but dative in (2); singular in (1-3), but plural in (4); masculine in (1-2), but feminine in (3), all in accordance to the features of the subject).

- (1) *Я пришел сам.*
- (2) *Мне прийти самому?*
- (3) *Она пришла сама.*
- (4) *Они пришли сами.*

## The case of Russian subject PRO

The subject of infinitive clause in Russian is normally dative, as in (2). Indeed, in dependent infinitive clause with a PRO subject a dative floating quantifier can also be found, as in (5). However, the nominative case is sometimes the only option for a floating quantifier in infinitive clause, as in (6).

(5) *Он приказал нам прийти самим.*

(6) *Он хочет прийти сам.*

## Local agreement hypothesis [Comrie 1974]

The idea is that the PRO of infinitive clause, whether lexically controlled or not, is assigned one of the cases – nominative or dative. Nominative is restricted to subject control PRO constructions and dative is a default case for the subject of infinitive clause. The syntactic structures for (5-6) are proposed as in (7-8).

(7) *[Он] [приказал нам<sub>i</sub> [PRO<sub>i,DAT</sub> прийти самим<sub>i,DAT</sub>]]*

(8) *[Он<sub>i</sub>] [хочет [PRO<sub>i,NOM</sub> прийти сам<sub>i,NOM</sub>]]*

## Long-distance agreement hypothesis [Franks 1990]

S. Franks claims that only a subject of a tensed CP infinitive clause can be assigned dative case, and PRO is essentially caseless. This claim is supported by the fact that most of the infinitives with an overt dative subject can take a tense auxiliary for future and past, as in (9) and its counterpart (10).

(9) *Куда нам поставить этот ящик?*

(10) *Куда нам было поставить этот ящик?*

## Direct predication hypothesis [Babby 1998]

Subject control PRO infinitive complement is viewed as a bare VP without a PRO, whereas other infinitives have a PRO which is assigned dative as a default case for the subject of infinitive. Floating quantifier thus receives the case form the nearest subject by agreement. The corresponding structures for (5-6) in this theory are (11-12).

(11) [*Он*] [*приказал нам<sub>i</sub>* [*PRO<sub>i,DAT</sub>* *прийти сам<sub>i,DAT</sub>*]]

(12) [*Он<sub>i</sub>*] [*хочет* [*прийти сам<sub>i,NOM</sub>*]]

## Problematic examples for the three hypothesis

(13) *Я купил машину, чтобы ездить на работу самому.*

There are also examples of a clear CP where the floating quantifier is nevertheless assigned nominative, as in (14), which is inconsistent with both long distance agreement hypothesis and direct predication hypothesis.

(14) *Андрей слишком труслив, чтобы прийти сам.*

None of the theories listed above can deal with ex. (15), where floating quantifier can take accusative.

(15) *Меня просят прийти самого.*

## PRO control in Compreno

Compreno transfigures an input text into a semantic-syntactic tree, where each node is a notion given a package of grammatical information and diathesis description. Therefore, if some node is a parent to an infinitive clause, given all the information about the model of the lexical item in this node we can predict, what kind of control will be applied in the particular construction. For instance, consider (16-19).

(16) *Я пришел починить трубу.*

(17) *Вы сказали мне починить трубу.*

(18) *Меня прислали вам починить трубу.*

(19) *Я был прислан вам починить трубу.*

(16) Я пришел починить трубу.



(17) *Вы сказали мне починить трубу.*



(18) *Меня прислали вам починить трубу.*

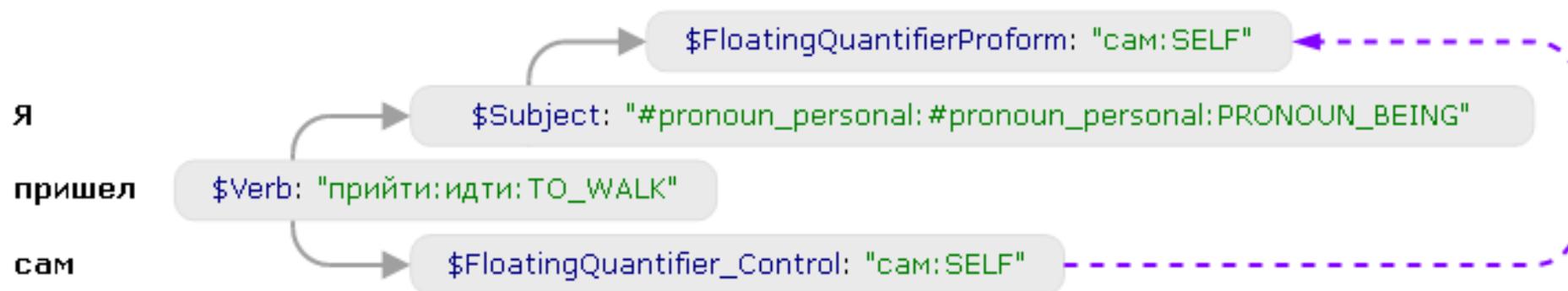




## Floating quantifiers in Compreno

Floating quantifiers are moved from within NP in Compreno syntactic structure and agree in case, number and gender with the parent node before movement.

Compreno syntactic structure for ex. (1) *Я пришел сам.*



So for the floating quantifier to be marked with case X the subject of the minimal clause containing it has to be assigned the same case X. This is consistent with the local agreement hypothesis rather than any other.

## PRO case: floating quantifier evidence

This implies that a subject of an infinitive clause can be nominative (as in (20)), dative (as in (21)) or accusative (as in (22)).

(20) *Я хочу починить трубу сам.*

(21) *Мне сказали починить трубу самому.*

(22) *Меня просят починить трубу самого.*

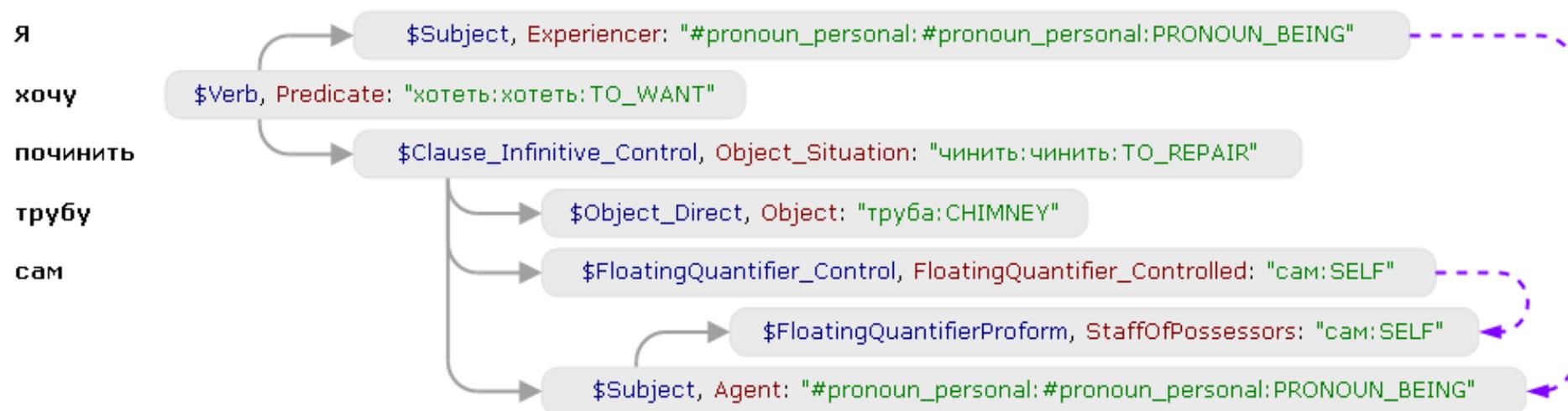
Moreover, there is a number of cases where the floating quantifier is unacceptable or at least dubious. Such are the instances where the PRO of the infinitive clause is coindexed with instrumental NP, as in (23).

(23) *Правительством планируется восстановить разрушенные территории \*само/\*самим.*

## PRO case choice in Compreno

Every infinitive node of the tree (after the tree is built and all the non-tree links are established) is assigned a special flag that encodes information for the type of control applied in the particular structure. Let us call it TypeOfPRO flag. Restricting the subject-predicate relation, which is represented as one arc in the tree, we assign a certain case to the subject according to the flag of the parent node. Consider semantic-syntactic tree for (20).

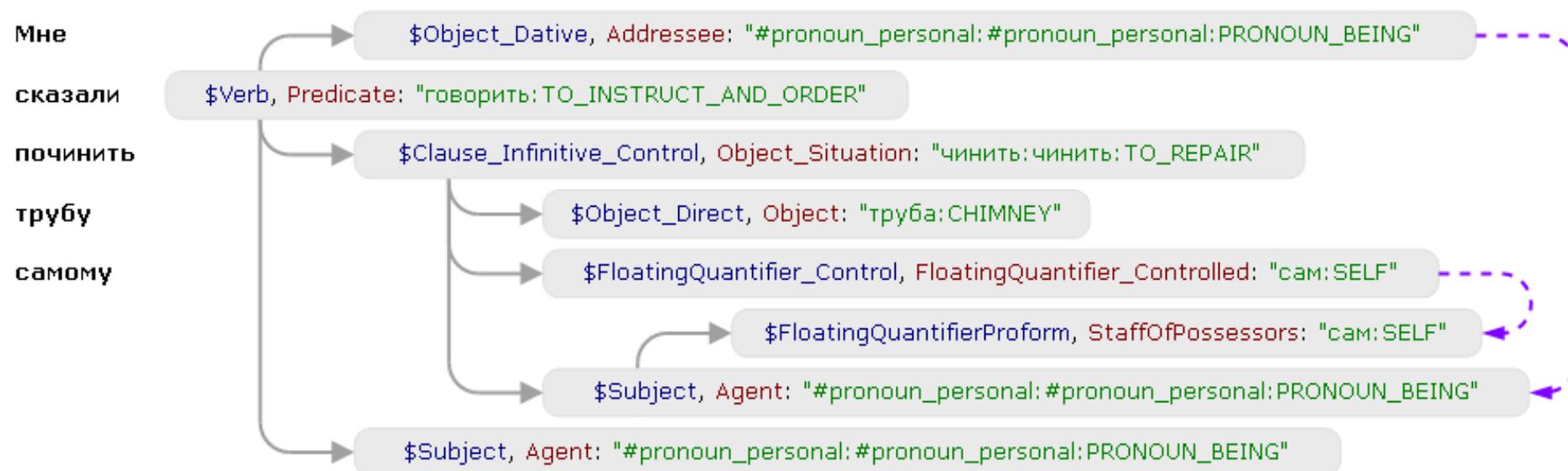
(20) *Я хочу починить трубу сам.*



In (20), figure above, the PRO of infinitive clause is controlled by the subject of the matrix predicate. Due to it, the infinitive node bares the NominativePRO flag and its PRO is assigned nominative case. It transmits nominative to the floating quantifier before movement, so that the moved quantifier is also marked nominative. This makes examples (24) with dative case invalid.

(24) \**Я хочу починить трубу самому*

(21) *Мне сказали починить трубу самому.*



In (21), figure above, the PRO of infinitive clause is controlled by the dative object of the matrix predicate, so the infinitive node bares DativePRO flag. Its PRO is assigned dative case and transmits dative to the floating quantifier by agreement. Thus (25) with nominative is also analyzed as invalid.

(25) \**Мне сказали починить трубу сам.*

## Further applications of the TypeOfPRO flag

The mechanism illustrated above has several other applications apart from the case choice for the floating quantifiers. It has been noticed before, that subject control PRO infinitive constructions can take a short form adjective as a complement (26), whereas object control PRO infinitive constructions cannot (27).

(26) *Я должен/хочу быть красив/красивым.*

(27) *Мне хочется быть \*красив/красивым.*

As there are lexical items without full form it is crucial for the machine translation to choose a synonymous lexical item for constructions with object control PRO infinitive, cf (28-29).

(28) *Я должен быть рад.*

(29) *\*Мне хочется быть рад.*

In Compreno it is simply done by applying the TypeOfPRO flag on the infinitive node to restrict form choice in the complement node. It is according to this test that the PRO in constructions such as (23) are assigned dative in Compreno: short form of adjective is unacceptable as a complement in such constructions, cf. (30).

(30) *Правительством планируется быть \*компетентно /  
компетентным.*

# Conclusion

Although a model-based approach to machine translation is known to be relatively labour-intensive, it looks more promising when interpreting and translating such complex structures as those with floating quantifiers in infinitive clauses. For the analysis of those constructions it seems more reasonable to follow the local agreement hypothesis and assign case to PROs, however intuitively dubious that may be.