

The Order of Objects in Russian:
A Corpus Study

The order of objects

Objects can be in different order in Russian:

Maša dala Pete_{IO} yabloko_{DO} ‘Masha gave Petya an apple’

No my dolžny garantirovat' [IO našim nalogoplatel'sikam] [DO kačestvo vysšego obrazovaniya].
‘But we must guarantee the quality of higher education to our taxpayers.’

Maša dala yabloko Pete ‘Masha gave an apple to Petya’

*Odnako v takom slučae bylo by spravedljivo ustanovit' pravila, garantiruûšie [DO vozmožnost'] [IO licam],
ne vladeušim estonskim yazykom, vystupat' na svoem rodnom yazyke.*

‘However, in such a case, it would be fair to establish rules guaranteeing the possibility for persons who do not speak Estonian to speak in their own language.’

This has been noted in many studies, various factors have been highlighted, but the issue has **never been investigated on corpus data.**

Relevant factors

- information structure

Slioussar 2007, Dyakonova 2009, Baylin 2012

- animacy

Tomlin (1986): Animated-first principle, Heine&König 2010, hierarchies of all sorts

- proper names

Heine&König 2010, hierarchies

- pronominality

Slioussar 2007, Dyakonova 2009, Wasow&Arnold 2011

- length (and depth)

Ross 1967: heavy NP-shift, Faghiri&Samvelian 2015, Wasow&Arnold 2011, Letuchiy 2018; Heine&König 2010

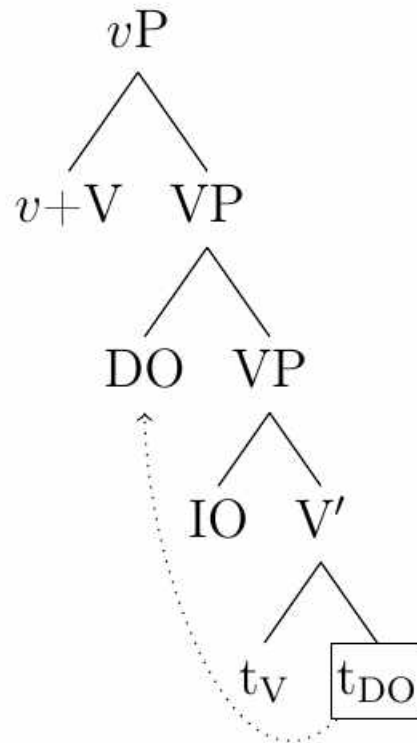
Previous studies

Slioussar 2007

- accessibility and salience as the basis
- IO DO as basic order
- other orders derived via movement

Vasya dal yabloko Maše

‘Vasya gave an apple to
Masha’



Dyakonova 2007

- argued that V+ACC (verb + Theme) makes a constituent
- proves that by examples of russian idioms like *peremyvat' kostočki* 'to gossip', *stroit' glazki* 'to flirt'
- and by examples of contrastive topic

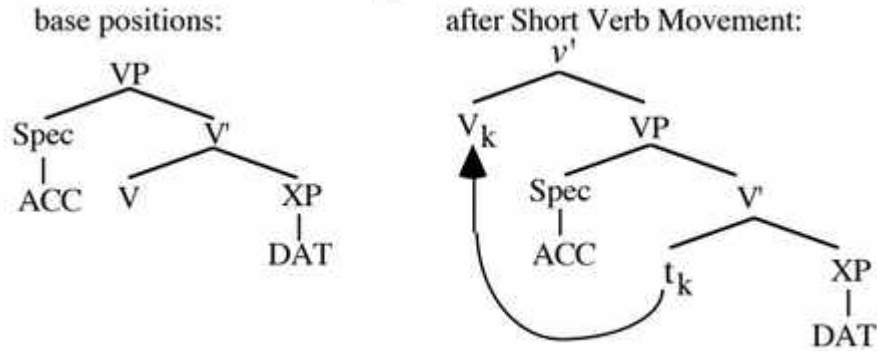
a. [TOP *Čitat' skazki*]_i *roditeli detyam očen' lûbyat t_i*
read_{INF} tales_{ACC} parents_{NOM} kids_{DAT} very love

b. ??/*[TOP *Čitat' detyam*]_i *roditeli skazki očen' lûbyat t_i*
read_{INF} kids_{DAT} parents_{NOM} tales_{ACC} very love

Bailyn 2010, 2012

What inside VP?, Syntax of Russian

- VP structure in Russian looks like following, neutral order is DO-IO

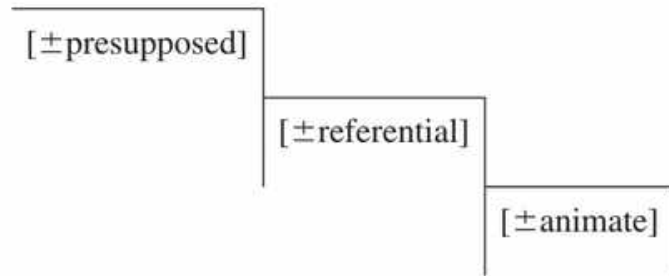


- and Baylin proves it by reciprocal binding and secondary predication

Titov 2017

- supports Baylin on neutral DO-IO order and provide Argument Prominence Hierarchy on which word depends order:

Argument prominence hierarchy



When none of the APH traits work - for example, in the context of a wide focus - the order in which objects are combined is governed by the thematic hierarchy, resulting in the canonical order of Theme-Goal, i.e. DO-IO.

LF rules are postulated, the consequences of which are 3/4 signature (Bobaljik & Wurmbrand 2011)

[+presupposed] > [-presupposed]

- $O_{[+presupposed]} > IO_{[-presupposed]}$
- $O_{[+presupposed]} > IO_{[+presupposed]}$
- $O_{[-presupposed]} > IO_{[-presupposed]}$
- $*O_{[-presupposed]} > IO_{[+presupposed]}$

Boneh, Nash 2017

- claim that the basic word order depends on individual verbs and divide verbs into three groups based on the results of several tests: reciprocal binding, frozen scope, depictives

core ditransitives (ACC-DAT and DAT-ACC are possible) like verbs of giving, manner of speech, future having or dispossessing, communication, sending, causation of motion

verbs of comparison and proximity (ACC-DAT) like *успокоить* 'calm down', *подчинить* 'subject'

verbs with the role of male/beneficiary (DAT-ACC) like *создать*, *успокоить*, *заколдовать*

Corpus sample and methodology

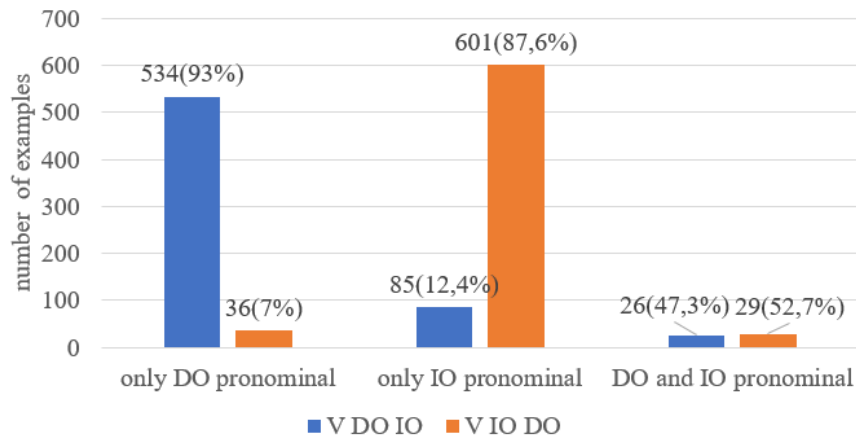
Data

SinTagRus Corpora which is manually marked. DFS by word forms marked as '1-КОМПЛ' и '2-КОМПЛ' with '1-КОМПЛ' in ACC.

Whole sample consists of dative and prepositional subsamples. We excluded some data:

- orders with objects before the verb
- IO in instrumental and in genitive and nominative, as in the verbs *lishit* "deprive" and *nazyvat* "name"
- all pronouns

word order	number of examples	percentage
DO V IO	1087	17%
IO V DO	616	9,6%
DO IO V	208	3,2%
IO DO V	125	2%
V DO IO	2591	40,5%
V IO DO	1771	27,7%



Final sample

For the remaining data, we considered these factors:

- animacy
- proper vs. common names (manual markup)
- length / depth (excl. appositives)
depth being the maximum distance in the tree from the head to the word in the constituent

Word order	V DO IO	V IO DO
Dative subset	313 (44,2%)	394 (55,8%)
Prepositional subset	1598 (69,5%)	701 (30,5%)
Total	1911 (63,5%)	1095 (36,5%)

Length, depth and animacy

- For prepositional subset IO is **longer** than DO
- For dative it is the other way around: IO is **shorter** than DO

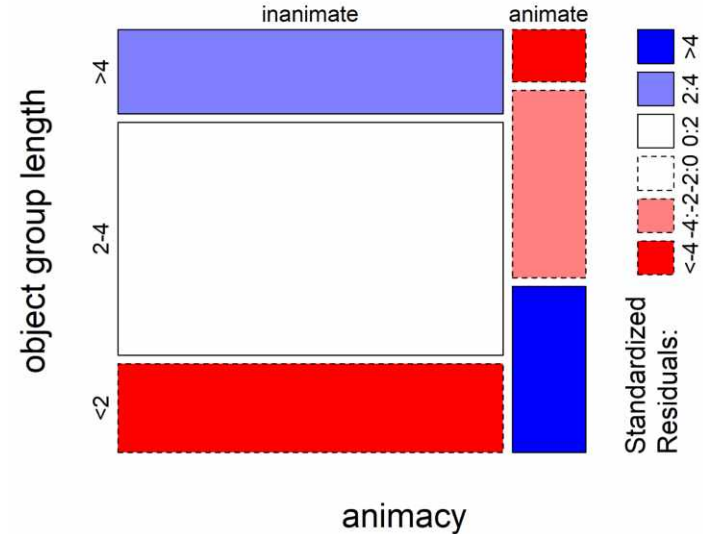
This can be related to the distribution of semantic roles.

		Average depth		Average length	
		Animate	Inanimate	Animate	Inanimate
Prepositional subset	IO	2,1	2,4	3,6	4,2
	DO	1,0	1,3	2,5	3,0
Dative subset	IO	0,9	1,3	2,3	2,9
	DO	1,3	2,1	3,3	4,2

Length, depth and animacy

Animate arguments tend to be more salient, and more salient arguments have shorter descriptions [Ariel 1990; Givón 1983; Kibrik 2011].

Length and animacy are indeed correlated
($\chi^2=55.51$, $p<0.001$, corrected Kramer's $V=0.196$)



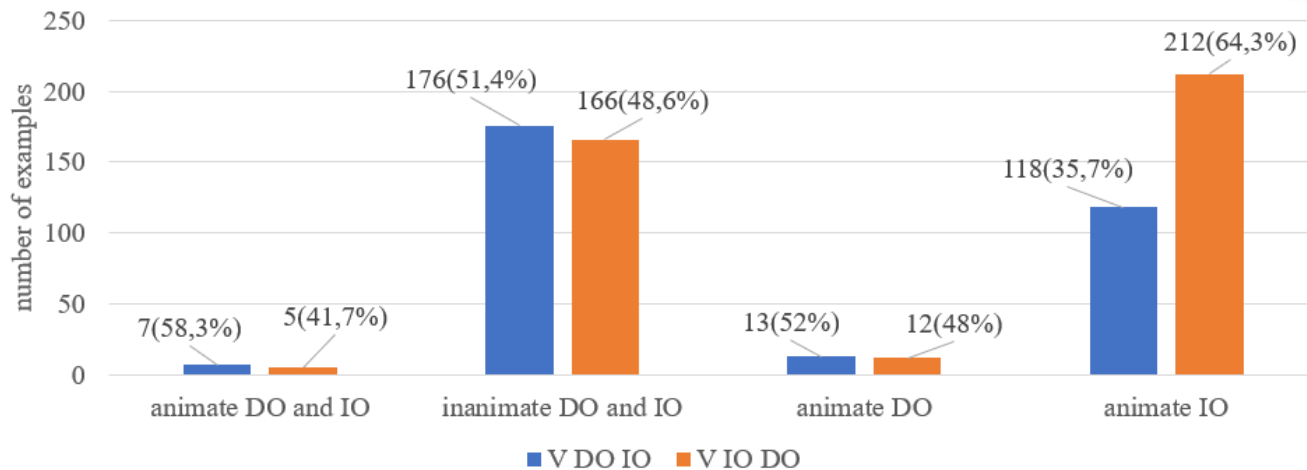
Correlation between animation and length.
Red - ratio is lower than expected under independency
Blue - ratio is higher than expected under independency

Animacy

Animate datives are more frequent than animate prepositional objects

Animate DOs are more common with the prepositionals IOs.

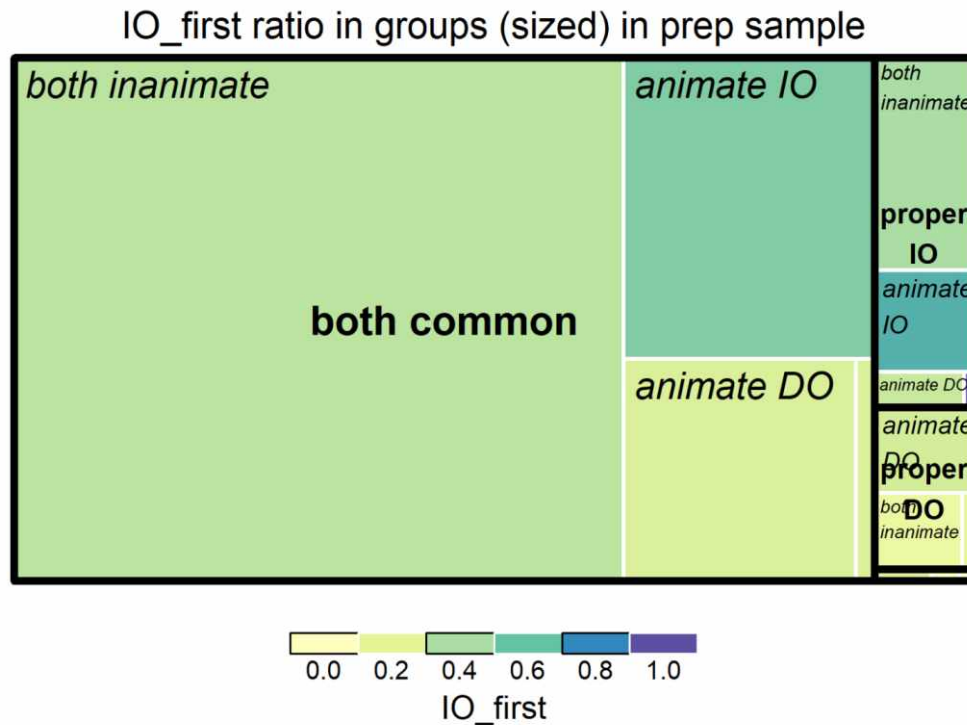
	animate DO	animate IO	both animate	both inanimate
Dative	25	329	12	341
Prepositional	356	184	21	1738
Total	381	513	33	2079



Proper names and animacy

- Proper names tend to be first
- Animate objects occur earlier than inanimate

The size of groups and subgroups is iconic — it indicates the number of examples, the color indicates the proportion of the order 'IO DO' in this subgroup.



Regression

Logistic regression: IO_DO ~ factors

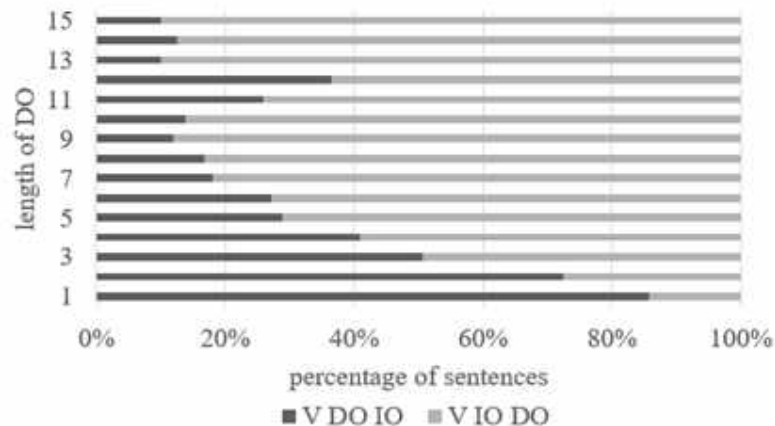
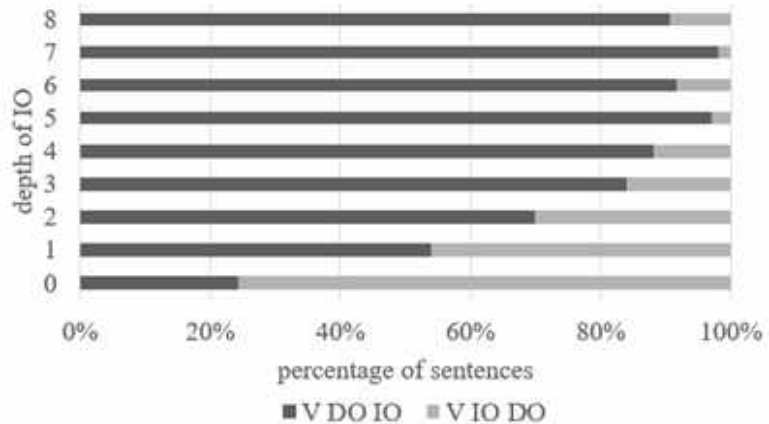
Length is the number of words

Depth is the maximum distance in the tree from the head to the word from the constituent (without appositive definitions)

	β	SE	z value	p-value
IO type: dative	0,73	0,19	3,83	<0,001
IO length	-2,12	0,18	-12,05	<0,001
DO length	1,80	0,11	15,80	<0,001
IO animacy: animate	0,39	0,20	2,00	0,049
DO animacy: animate	-0,25	0,20	-1,27	0,203
IO length * animacy	0,32	0,34	0,95	0,341
DO length * animacy	0,65	0,39	1,68	0,092

Length and depth

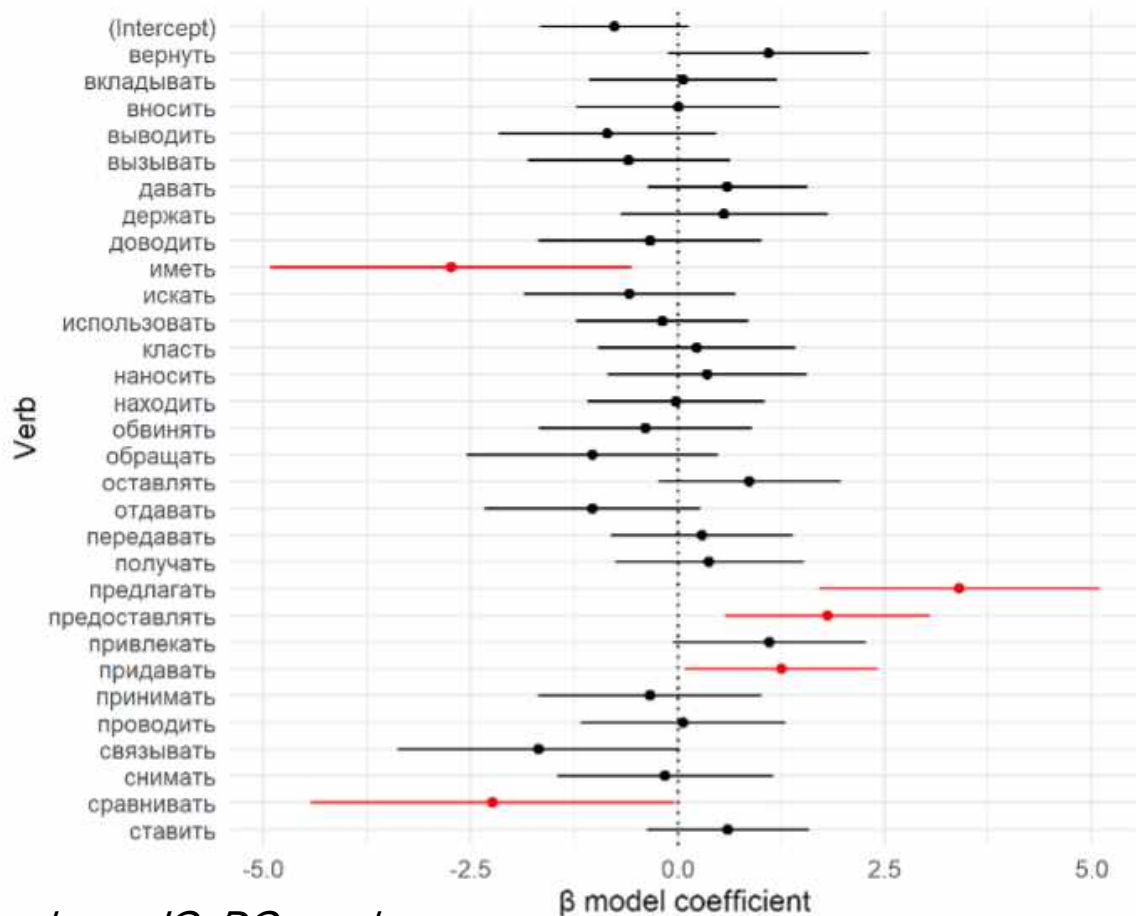
The longer and the more complex object is likelier last



Individual verbs

Contrary to Boneh and Nash (2017), no clear-cut tendencies for different verb groups.

- *imet* 'have' and *sraivivat* 'compare' prefer DO IO
- *poluchat* 'receive', *predlagat* 'offer' and *pridavat* 'lend' prefer IO DO



logreg: IO_DO ~ verb
verb frequency >= 20

Verb influence (p=0.05) ● insignificant ● significant

Conclusions

- When the indirect object is expressed by a prepositional group, the DO IO order is more neutral.
- When the indirect object is expressed by a dative, the IO DO order is more frequent. In the functional approach, it could be called more basic.
- In general, for each of the samples, the prevailing order can be explained by object properties