

CASE-MARKING OF PROPER NAMES IN CLOSE APPPOSITION IN RUSSIAN

1. INTRODUCTION

Close apposition is often defined as juxtaposition of two noun phrases with a shared referent and no intervening pause (Matushansky 2013: 1).

(1) *poët Blok*

‘the poet Blok’

- In Russian appositional constructions proper names (PN) can have the same case marker as the preceding **sortal** (Moltmann 2012) (2), no case marker (3) or have a genitive ending (4).

(2) Na **ostrov-e** **Krit-e** čti-l-i peščer-u gde vyros Zevs ...
 on island-LOC Crete-LOC honor-PST-PL cave-ACC where grow.up.PST Zeus
 (RNC: M. Gasparov. Zanimatel’naja Grecija (1998))

‘On the island Crete they honoured the cave, where Zeus grew up’.

(3) ...volna ot nego razori-l-a moguč-ee carstvo na **ostrov-e** **Krit.**
 wave from he.GEN ruin-PST-F mighty-ACC.N kingdom on island-LOC Crete-NOM
 (ibidem)

‘...the wave from it ruined the mighty kingdom on the Crete island’.

(4) ...mačta čerti-l-a kop’ec-om po **sozvezdi-u** **Kassiope-i**
 mast trace-PST-F spear-INS along constellation-DAT Cassiopeia-GEN
 (RNC: K. Badigin. Sekret gosudarstvennoj važnosti (1974))

‘The mast was drawing (a line) with its spear along the Cassiopeia constellation’.

- In many cases both options with and without case agreement ((2) and (3)) are possible.
- Although the problem has long been known (ПГ 1980: 58; Розенталь 1989: 265–267; Голуб 2010: 278–279; Суперанская 2013), no corpus-based study was conducted on the issue before (except for (Граудина 1976: 138–145)).

In this talk:

- some previously made suggestions revisited based on the corpus data;
 - some new parameters tested.
- ! Under the term “proper name” mostly human and animal names, as well as geographical and astronomical terms are included (Суперанская 1973: 178). Although these boundaries seem to be restrictive, more work should be done to define the features of properhood in Russian.

1.1 DATA

- Data mainly comes from the General Internet-Corpus of Russian (GICR, <http://www.webcorpora.ru>).
 + examples are numerous and diverse (cf. number of entries for *Jadri*: RNC — 28, GICR — 1023);
 – great number of unwanted examples present (Internet advertising, duplicated examples).
- Samples are formed according to the following principles:
 - if it is possible to find enough data on certain names → statistics on selected names:
 • ([lemma="пека" pos="Ncfs"]-[word="пека"])[lemma="X"],
 where X = Oka/Neva/Don etc.
 - if it is difficult to find data on certain names → random names considered, number of entries < 1 000
 • ([lemma="деревня" pos="Ncfs"]-[word="деревня"])[pos="Np.s"].

1.2 PROBLEMS

- Although there seems to be a lot of different PNs of various kinds, it is sometimes difficult to find good examples (cf. names of stars are often not frequent and complex, names of lakes are exotic, names of bays, seas and some other geographical terms are mostly adjectival) → samples are limited in certain cases.

2. INVESTIGATION OF THE FACTORS POSSIBLY INFLUENCING VARIATION

2.1 ϕ -feature¹ congruency (Matushansky 2013)

1. **Gender congruency** appears to be of great relevance in the majority of cases.

Table 1. The case agreement of masculine and feminine PNs with the sortal *reka*(f) 'river'

Name (F)	agreement	no agreement	% of agreement cases	Name (M)	agreement	no agreement	% of agreement cases
<i>Oka</i>	1201	587	67 %	<i>Amur</i>	58	391	13 %
<i>Volga</i>	2480	1847	57 %	<i>Dnepr</i>	25	208	10 %
<i>Angara</i>	319	356	47 %	<i>Don</i>	38	417	8 %
<i>Pečora</i>	157	171	47 %	<i>Ural</i>	18	204	8 %
<i>Kama</i>	456	539	45 %	<i>Voronež</i>	9	100	8 %
<i>Lena</i>	589	790	42 %	<i>Enisej</i>	16	246	6 %
<i>Om'</i>	32	94	25 %	Total	164	1566	10 %
<i>Ob'</i>	311	1406	18 %				
<i>Lovat'</i>	17	156	9 %				
Total	5562	5946	48 %				

Except for the ACC forms, because of the homonymy of ACC and NOM in case of masculine names

- But there are exceptions:
 - sortals *respublika* 'republic' and *štat* 'state':

Table 2. Agreement of feminine and masculine PN with the sortals *štat* 'state' and *respublika* 'republic'²

Gender	<i>štat</i>			<i>respublika</i>		
	agreement	no agreement	% of agreement cases	agreement	no agreement	% of agreement cases
f	184	11330	1,5 %	122	34 427	0,4 %
m	68	8807	0,8%	599	20 728	2,8 %

2. **Number congruency** is as well significant for the agreement.

- The agreement of pluralia tantum names is avoided with singular sortals.

Table 3. The case agreement of inherently plural PNs with the sortal *strana*(f) 'country' in singular

Name	agreement	no agreement	% of agreement cases Total
<i>Niderlandy</i> 'Netherlands'	1	30	4%
<i>Filipiny</i> 'Philippines'	1	16	
Total	2	46	

- This is not the case when the sortal is plural as well:

¹ The term is mostly used in generative grammar to refer to the features, involved in a predicate-argument agreement, definiteness and honorification (Adger 2008: 2). Mainly under the term "phi-features" such properties as gender, number and person are understood.

² In this table, the following PNs were considered: 1) *Vašington* 'Washington', *Kanzas* 'Kansas', *Massačusets* 'Massachusetts', *Vermont* 'Vermont', *Illinojs* 'Illinois', *Texas* 'Texas', *Arkanzas* 'Arkansas', *Aljaska* 'Alaska', *Arizona*, *Virdžinija* 'Virginia', *Kalifornija* 'Kalifornia', *Džordžija* 'Georgia', *Indiana* 'Indiana', *Nebraska* 'Nebraska', *Florida* 'Florida'; 2) *Xakasia*, *Adygeja*, *Burjatija*, *Karelija*, *Dagistan*, *Baškortostan*, *Tatarstan*, *Altaj*.

- (5) ...dobira-l-i-s' iz Xakas-ii v Tyv-u čerez pereval
get.to-PST-PL-REFL from Xakasia-GEN in Tyva-ACC through pass
gor **Sajany**
mountain.PL.GEN Sajany
'(They) were getting from Xakasia to Tyva over the pass of the Sajany mountains'. (GICR)
- (6) Èto uže čast' menja... otrog-i **gor** **Sajan**
this already part me.GEN spur-PL mountain.PL.GEN Sajany.GEN
'Now it is a part of me... the spurs of the Sajany mountains'. (GICR)

Table 4. Case agreement of inherently plural PNs with the sortal *gora(f)* 'mountain' in plural

Name	agreement	no agreement	% of agreement cases Total
<i>Andy</i> 'Ands'	23	6	75 %
<i>Sajany</i> 'Sajans'	14	8	
<i>Al'py</i> 'Alps'	20	5	
Total	57	19	

- When the sortal is plural and there is more than one PN in an appositional construction (combined PNs), the agreement is rather rare both in case of inherently singular and plural PNs.

- (7) Massov-ye piščev-ye otravleni-ja v **gorod-ax** **Lermontov** **i**
mass-PL.NOM food-NOM.PL poisoning-NOM.PL in city-LOC.PL Lermontov and
Kislovodsk...
Kislovokodsk
'Mass food poisoning in the cities of Lermontov and Kislovodsk...'. (GICR)
- (8) ...radiacionn-aja obstanovk-a v raboč-ej zon-e atomn-oj
radiological-NOM.F situation-SG.NOM in working-SG.LOC zone-SG.LOC nuclear-SG.GEN
elektrostanc-ii v **gorod-ax** **Balakov-e i Saratov-e** normal'n-aja
power.station-SG.GEN in city-PL.LOC Balakov and Saratov-LOC normal-NOM.F
'...the radiological situation in the working area of nuclear power station in the cities of Balakov and Saratov is normal'. (GICR)

Table 5. Case agreement of combined PNs with the sortal *gorod* 'city' in plural

	agreement	no agreement	% of agreement cases
<i>gorod(ov/am/ami/ax) X (i/,) Y</i> ³	196	582	25 %

- When the coordinated PNs are inherently plural and the sortal is contextually plural the agreement is rare:

- (9) ... na sklad-ax v podmoskovn-yx **gorod-ax** **Ximk-ax** **i**
on warehouse-PL.LOC in Moscow.suburb-PL.GEN city-PL.LOC Ximki-LOC(PL) and
Mytišč-ax konfiskova-l-i 200 kilogramm-ov narkotik-ov
Mytišči-LOC(PL) confiscate-PST-PL kilogram-PL.GEN drug-PL.GEN
'... 200 kilograms of drugs were confiscated in warehouses of Moscow suburban towns Ximki and Mytišči'. (GICR)

Table 6. Case agreement of combined inherently plural PNs with the sortal *gorod* 'city' in plural

	agreement	no agreement	% of agreement cases
<i>gorod(ov/am/ami/ax) X(PL) i Y(PL)</i>	7	401	1,7 %

- All in all:

Table 7. Summary of case agreement patterns in case of number congruency/ incongruency

	singular PN	inherently plural PN
singular sortal	√	—
inherently 'plural' sortal	— ⁴	√
contextually plural sortal	—	—

³ Search formula: ([lemma="город" pos="Ncmp"]-[word="города"])[pos="Nc.s"]([word="и" | word=","])

⁴ Very little data available: ... na ostrovaax_{LOC} Zanzibar_{NOM} bolee 90% - musulmane '... on the Zanzibar island more than 90% are muslims'.

2.2 Morphological factors

2.2.1 Internal complexity

- Internal complexity of some PNs has been always (Розенталь 1989: 266) claimed to be a factor lowering the probability of agreement.

Table 8. Case agreement of structurally complex PNs with the sortal *gorod* 'city'

Name	Frequency, IPM ⁵	agreement	no agreement	% of agreement cases	% of agreement cases with PNs with simplex structure having the comparable frequency	
<i>Nižnij Novgorod</i>	≈7.75	453	1018	30 %	<i>Jaroslavl'</i>	84 %
<i>Nižnij Tagil</i>	≈1.02	115	484	19 %	<i>Vitebsk</i>	79 %
<i>Sergeev Posad</i>	≈0.42	70	509	12 %	<i>Mogilev</i>	80 %
<i>Velikij Ustug</i>	≈0.26	23	302	7 %	<i>Bratsk</i>	84 %
<i>Staryj Oskol</i>	≈0.17	23	463	4 %	<i>Serpuxov</i>	60 %

2.2.2 Adjectival base of a PN

- Adjectival PNs are expected to agree more often, as adjectives always agree with their noun heads⁶.

- (10) K konc-u 19 vek-a na gor-e Vysok-oj dobyva-l-o-s'
to end-DAT century-GEN on mountain-LOC Vzsokaja-LOC mine-PST-N-REFL
bole 12 million-ov pud-ov rud-y v god
more million-PL.GEN pood-pl.gen ore-GEN.SG in year
'By the end of the 19th century more than 19 poods of ore were mined on the Vysokaja mountain'.

Table 9. Case agreement of adjectival PNs with the sortal *gora* 'mountain'

Construction	agreement	no agreement	% of agreement cases
<i>gor(y/e/oj)</i> ⁷ <i>Belux(a/i/e/oj)</i> 'Beluxa mountain'	258	662	28 %
<i>gor(y/e/oj)</i> X-(<i>aja/oj</i>)	203	97	67 %

2.2.3 Presence of particular structural elements

- City names with the affix *-sk-* (*Novosibirsk*, *Arkhangelsk*) tend to agree in case more often than the names without this element.

Table 10. Case agreement of city names with affix *-sk-*

Name	Frequency, IPM	agreement	no agreement	% of agreement cases	% of agreement cases with city names having no <i>-sk-</i> element of comparable frequency		
<i>Sovetsk</i>	≈0,21	397	100	79 %	<i>Toržok</i>	≈0,37	35 %
<i>Belozersk</i>	≈0,11	94	15	86 %	<i>Gubkin</i>	≈0,19	35 %
<i>Mcensk</i>	≈0,11	106	55	66 %	<i>Portsmut</i>	≈0,11	18 %
<i>Axtubinsk</i>	≈0,07	43	109	72 %	<i>Olonec</i>	≈0,09	33 %
<i>Abinsk</i>	≈0,06	34	22	60 %	<i>Kanaš</i>	≈0,10	28 %

- sk-* is the most frequent affix in Russian city names (in the list⁸ containing one thousand Russian city names 310 have the affix *-sk-*). Affix *-in-*, being second most frequent (20 names in a thousand list), seems not to lead to the same effect.

⁵ IPM = items per million. Automatically measured in GICR for the nominative form of the given PN.

⁶ By this we do not assume the subordination relation between elements of apposition. This question needs further analysis.

⁷ The absence of the accusative forms is due to the same reasons as above.

⁸ The list is taken from: <https://rusmap.net>

Table 11. Case marking of city names with affixes *-sk-* and *-in-*⁹

Suffix	agreement	no agreement	% of agreement cases
<i>-sk</i>	20102	4973	80 %
<i>-in</i>	1120	1442	44 %

2.3 Syntactic factors

2.3.1 Linear distance

- The ability of appositive constructions with PNs to have “shared” case-marker can be not least due to their linear closeness → what kind of agreement pattern should be expected in case of linear distance?

Table 12. Case agreement of city names with modified sortal *gorod* ‘city’

	agreement	no agreement	% of agreement cases
<i>gorod(a/u/om/e) vojskoj slavy X</i> ‘in the city of military glory X’	129	27	82 %

- More illustrative with masculine country names:

- (11) *sredi posledn-ix osobenno mnogo vyxodc-ev iz stran-y*
among last-PL.GEN especially a.lot nationals-PL.GEN from country-GEN
moej mečt-y Kazaxstan-a
my-SG.GEN dream-SG.GEN Kazakhstan-SG.GEN
‘Of the latter, there are especially many nationals from the country of my dream — Kazaxstan’.
(GICR)
- (12) *Neprivyčno posle stran-y vosxodjašč-ej ulybk-i Tajland-a ...*
Unusual after country-SG.GEN rising-SG.GEN smile-GEN Thailand-GEN
‘It was unusual after the country of rising smile, the Thailand’.

- Rare construction → limited data.
 - only 16 examples with the sortal *strana* ‘country’, followed by heavy genitive modifier, and masculine PN is found, but still some statistics available:

Table 13. Case marking of masculine country names in cases of modified/ unmodified sortal *strana* ‘country’¹⁰

	agreement	no agreement	average % of the agreement cases
sortal without genitive modifier	20	475 ¹¹	4 %
sortal with genitive modifier	14	2	88 %

2.3.2 Case

- In some cases the agreement is more frequent than in others. In the Instrumental and Locative cases the agreement pattern appears to be different from the Genitive, Dative and Accusative.

Table 14. The distribution of agreement and no agreement forms of PNs *Moskva* ‘Moscow’ and *Kiev* in appositional constructions with the sortal *gorod* ‘city’ in five cases¹²

Case	agreement	no agreement	% of the agreement cases
<i>Moskva</i>			
Genitive	31674	1512	95 %
Dative	2636	259	91 %
Accusative	1119	27	97 %
Instrumental	108	59	64 %

⁹ Fishers exact test, $P < 0,001$.

¹⁰ Fisher’s exact test, $P < 0,001$.

¹¹ The data on the following masculine PNs examined: *Kitaj* ‘China’, *Izrail* ‘Israel’, *Irak* ‘Iraq’, *Kazaxstan* ‘Kazakhstan’, *Afganistan* ‘Afganistan’ (five most frequent, according to (Ляшевская, Шаров 2009)), *Ljuksemburg* ‘Luxembourg’, *Kuvejt* ‘Kuwait’, *Alžir* ‘Algeria’, *Tunis* ‘Tunisia’, Turkmenistan ‘Turkmenistan’ (five least frequent).

¹² Only minimal manual filtration applied.

Locative	5510	4256	56 %
Total	47160		
<i>Kiev</i>			
Genitive	5663	605	90 %
Dative	379	3	99 %
Instrumental	83	37	69 %
Locative	1844	517	78 %
Total	9131		

- Case hierarchy:
nom acc/erg gen dat loc abl/inst others (Blake 2004: 156)

2.4 Lexical-semantic category

- There is obligatory case-agreement in close appositions with human names (Matushansky 2013: 5).
- In case of toponyms a considerable variation between different sortals can be observed (Граудина 1976: 140; Розенталь 1989: 165 and others).
- Sometimes case agreement is not a question → appositive oblique construction:
 - mainly with sortals like *sozvezdie* ‘constellation’ and *galaktika* ‘galaxy’

- (13) sozvezdi-em Tel’c-a/ ?Telec/ *Tel’c-om
constellation-INS Taurus-GEN Taurus Taurus-INS
‘(with) the Taurus constellation’

- (14) v galaktik-e Andromed-y/ *Andromeda
in galaxy-LOC Andromeda-GEN Andromeda-NOM
‘in the Andromeda galaxy’

- Latin constructions constellation_{NOM} Virginis_{GEN}/ *constellation_{NOM} Virgo_{NOM} (Google Books)
- in some other cases also possible

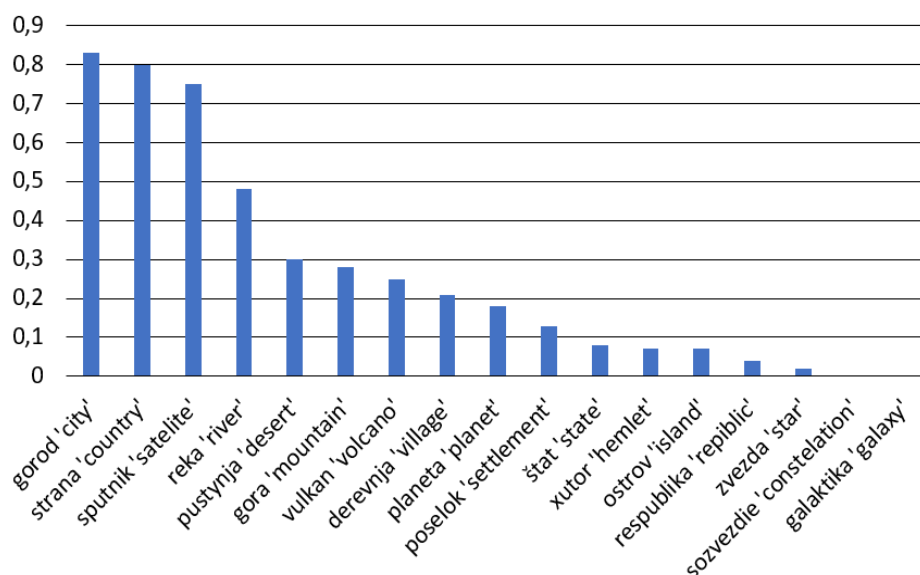
- (15) zeml-ja Egipt-a
land-NOM Egypt-GEN
‘the land of Egypt’

- (16) prazdnik Pasx-i/ Roždestv-a
festival Easter-GEN Christmas-GEN
‘Easter/ Christmas holiday’

- (17) cvetok azali-i neploxo rast-et pri temperatur-e +18... +22
flower azalea-GEN well grow.up-PRS[3SG] at temperature-LOC
‘azalea flower grows up good at the temperature of +18... +22’ (Yandex)

- Going back to the case agreement discussion:

Fig.1. The frequency of agreement in relation to the lexical-semantic category of the sortal



- The obtained continuum is not homogeneous (sortals, denoting places of human living do not form a single group, some sortals are unexpectedly high).
- The result is in consonance with toponym properhood “hierarchy” proposed in (Langendonck 2007: 204-212; Langendonck, de Velde 2016) and based on the markedness parameter:

- (18) human names > city names > country names > names of landscape objects > names of water bodies
- | | | | | |
|-------------|---------------|-----------------|-----------------------|-----------------------------|
| no marker | no marker | classifier | article + classifier | obligatory sortal + article |
| <i>Mary</i> | <i>London</i> | <i>Fin-land</i> | <i>the High-lands</i> | <i>the Northern Sea</i> |

3. Multifactorial analysis of case agreement in appositive constructions with sortal *gorod* ‘city’

- Material:
 - 100 Russian city names (high variety of agreement patterns, diverse morphology, great probability to find a lot of data, «sore point» of Russian grammar);
 - the data on their relative frequency (IPM, measured automatically in GICR);
 - the data on their agreement pattern (% of the cases with no agreement);
- Predictors:
 - gender congruency;
 - number congruency;
 - morphological complexity;
 - adjectival base;
 - presence of *-sk-* element;
 - frequency. ←

Supposed to correlate with the parameter of ‘user’s knowledge’ (Розенталь 1989: 266): the more familiar PN is for the user, the higher is the probability of its case declension (cf. Russian speakers are said to avoid to inflect the foreign PNs).
- Statistic model: multifactorial linear regression.
- Results:
 - Coefficient of the determination of the obtained model equals to $\approx 60\%$ ($R^{2\ 13} = 0.63$, $R^2\text{-adjusted}^{14} = 0.6059$) → about 60 % percent of variation is explained by the chosen model.

Table 15. The results of the statistical model implementation

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.7888	18.7654	0.308	0.758414
Frequency	-1.4922	0.2322	-6.428	5.67e-09
Gender	-8.3130	4.8984	-1.697	0.093057
Number	41.5637	7.8592	5.289	8.28e-07
Affix	-28.5882	4.7178	-6.060	2.96e-08
Two	28.6505	7.8763	3.638	0.000454
Adjective	13.2753	6.4235	2.067	0.041576

- inverse relationship between the frequency of use and the absence of case agreement;
- inverse relationship between the presence of *-sk-* affix and the absence of case agreement (presence of *-sk-* → higher probability of agreement);
- direct correlation between the inherent plurality and the absence of case agreement;
- direct correlation between internal complexity and absence of case agreement;
- direct correlation between adjectival character of the PN’s base and absence of agreement

⇒ Possible explanation: adjectival city names are less adjectival than other names of this type:

- do not appear on the left of the sortal (*gora Zmeinaja* ~ ^{OK}*Zmeinaja gora* ‘Zmeinaja mountain’ vs. *gorod Žukovskij* ~ **Žukovskij gorod* ‘the city of Žukovskij’);

¹³ Coefficient of determination (R^2) shows, how much of the variance in the dependent variable is predictable from the independent variable(s).

¹⁴ Is used in statistical models with multiple predictions. Helps to avoid an increase in the coefficient of determination only by increasing the number of predictors under consideration.

- can be used without the sortal (*my živem v Žukovskom* ‘we live in Žukovskij’ vs. *oni ostalis’ na Zmeinoj* ‘they remained on Zmeinaja’)

PN adjectival continuum:

Černoe more > *gora Zmeinaja* > *gorod Žukovskij*
 ‘the Black sea’ ‘Zmeinaja mountain’ ‘the city of Žukovskij’,
 where > is for “more adjectival”.

- no correlation between gender incongruency and absence of agreement → both feminine and masculine city names agree with the sortal with approximately equal probability.

4. Summary

- Case agreement in appositive constructions appears to be influenced by a range of factors:
 - lexical-semantic category of sortal;
 - morphological properties of the elements in apposition
 - ✓ gender and number congruency
 - ✓ internal complexity
 - ✓ adjectival character of the base
 - ✓ presence of particular classifier morpheme
 - syntactic properties of the construction
 - ✓ linear distance between elements
 - ✓ case
 - frequency of the PN
- A lot of factors appear to be not indicated yet.

References

- Голуб И. Б. Стилистика русского языка. М.: Айрис-пресс, 2016.
- Граудина Л.К., Ицкович В.А., Катлинская Л.П. Грамматическая правильность русской речи. Опыт стилистического словаря вариантов. М.: Наука, 1976.
- Ляшевская О. Н., С. А. Шаров, Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009.
- Русская грамматика / Под ред. М. Ю. Шведовой. М.: Наука, 1980.
- Розенталь Д.Э. Справочник. Правописание и литературная правка. М.: Книга, 1989.
- Суперанская А. В. Общая теория имени собственного. М.: Наука, 1973.
- Суперанская А. В. Словарь географических названий. М., 2013.
- Blake, B. J. Case. Cambridge: Cambridge University Press, 2004. First published 2001.
- Langendonck W. Theory and typology of proper names. Berlin, New York: Mouton de Gruyter, 2007.
- Langendonck, W. & M. Van de Velde. Names and grammar. / Carole Hough & Daria Izdebska (eds.) // The Oxford handbook of names and naming. Oxford: Oxford University Press, 2016. P. 17–38.
- Matushansky O. Sorts of proper names. The talk at Semantics and Philosophy in Europe (SPE) 6, St. Petersburg, June 10-13, 2013.
- Moltmann, F. Names and the Mass-Count Distinction. Ms., Université Paris, 2015.