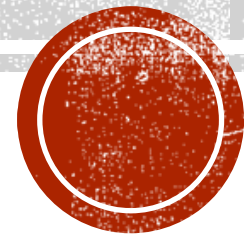


# **РУССКИЙ ЯЗЫК В КОНТАКТНОЙ И ТИПОЛОГИЧЕСКОЙ ПЕРСПЕКТИВАХ: КОРПУСНЫЕ РЕСУРСЫ ИРЯ РАН**

**КОНФЕРЕНЦИЯ К 20-ЛЕТИЮ НКРЯ  
(ИРЯ РАН, 20-21 ДЕКАБРЯ 2024 Г.)**



Егор Кашкин (ИРЯ РАН)

[egorka1988@gmail.com](mailto:egorka1988@gmail.com)

# ВВЕДЕНИЕ

- Корпус контактного русского.
- Параллельные корпуса как источник сведений о языках России.



# КОРПУС КОНТАКТНОГО РУССКОГО

- <http://web-corpora.net/ruscontact/corpus.html>
- Сейчас временный адрес: [https://rus-nonst.web-corpora.net/rus\\_nonst\\_corpus/search](https://rus-nonst.web-corpora.net/rus_nonst_corpus/search)
- Начало: грант РФФИ №17-18-01649 в рамках проекта "Динамика языковых контактов в циркумполярном регионе" (рук. О. В. Ханина, ИЯз РАН).
- С 2018 г. Группа по изучению контактного взаимодействия русского языка с языками коренных народов России в ИРЯ РАН.
- Н. М. Стойнова, И. А. Хомченкова, П. С. Плешак, техподдержка – Т. А. Архангельский [Khomchenkova et al. 2019].
- Контактные языки: ульчский, нанайский, нганасанский, лесной и тундровый энецкий, тундровый ненецкий; выбор подкорпуса.
- Сейчас добавляется горномарийский (Е. В. Кашкин; первичная разметка текстов – Д. Д. Белова, М.-Э. А. Винклер, Т. И. Давидюк, И. А. Хомченкова).





# КОРПУС КОНТАКТНОГО РУССКОГО

- Запись текстов в качестве "бонуса" в экспедициях.
- Носители довольно неоднородны по уровню стандартности русского.
- Влияние социолингвистического фактора; контрольная выборка более грамотных носителей.
- Но: ресурсы для разметки.



# КОРПУС КОНТАКТНОГО РУССКОГО

- Ручная разметка.
- Синхронизация с аудио (верификация, материал по фонетике).
- Стандартная орфография.
- Метаразметка.
- Детализированная разметка явлений, которые могут быть связаны с языковым контактом, тж. другие явно нестандартные явления и переключение кодов.

Cont synt tags:	<input type="text"/>	
Cont morph tags:	<input type="text"/>	
Cont lex tags:	<input type="text"/>	
Cont phon tags:	<input type="text"/>	
Cont pdp tags:	<input type="text"/>	
Substandard tags:	<input type="text"/>	

<b>tooltip_Syntax</b> agr_adj_gender agr_adj_num agr_adj_case agr_verb_gender agr_verb_num agr_verb_person agr_anaph_gender agr_anaph_num	gov gov_caseless gov_dom pp num neg comp	adposition prep_drop poss poss_wo verb_constr valency_change anaph other_synt	<b>tooltip_Polypredication</b> subord_compl subord_adv subord_rel coord	<b>tooltip_Discourse</b> disc_word disc_deixis disc_doubling disc_topic other_pdp	<b>tooltip_Prosody</b> pros_intonation pros_length pros_accent
---	--	---	--	--	---





# КОРПУС КОНТАКТНОГО РУССКОГО

- Стандартный корпусный поиск, например, по лемме: *очень*.

ReceptTala lpd 2017

✓ Но там/ же ... **очень** много ... как его\ червячков/ гврят, да/

kotiki hm\_VVV 2018

✓ [hm\_VVV] **Очень** ло= ловит.

Shaman thl

✓ **Очень** жара говорят\

recept hm\_VVV 2018

✓ А это, на= как блинное тесто-то, можно зелёный лук положить Нарезать лук **очень**. Помешать.

Ср. горномар. корпус:  
*piš nelä* ‘очень трудно’,  
*piš jaratat* ‘очень любят’,  
*piš tumaja* (букв.: очень  
думает),  
*piš rovotajat* (букв.: очень  
работают),  
*piš kačkân šänden* (букв.:  
очень наелся).

→ межъязыковые  
различия в дистрибуции  
интенсификаторов



# КОРПУС КОНТАКТНОГО РУССКОГО

- Поиск по пометам, например, нестандартный морфосинтаксис количественной конструкции (num).

Search result: 195 occurrences, 194 sentences found in approximately 62 documents.

Document	Search Term	Year	Context
ThreeBrothers2 tkf	1999	В сочетании с количественным числительным существительное оформляется показателем единственного числа (30, 31).	
<input checked="" type="checkbox"/>	Два олень передний санка место...	(30)	Хасава һацекэцяр тад' хонаванда си"ив' яля хоны [ТЛ].
RazgovoryKakVSkazkax oab	201	хасава      һацекэ-ця-р      тад'      хона-ва-нда      си"сив      яля мужчина      ребенок-DIM <sub>2</sub> -NOM/SgPOSS/2Sg      затем      лечь.спать-VN <sub>сонс</sub> -3Sg      семь      день	
<input checked="" type="checkbox"/>	Эту ... "Трое сыновья"	хоны-Ø спать-SUBJ/3Sg	
SkazkaSusu spk	2017	'Мальчик лёг спать, проспал так семь дней'. Ненецкий [Буркова 2010: 237]	
<input checked="" type="checkbox"/>	Вот это дети .... трое\ сыновья		
Klad vsg	2012	The noun following a numeral or a quantifier (such as bənsə 'all, the whole', bənjika 'every, all', ɲɯŋkəgə ~ ɲukəgə 'a lot of, many, much', kanəraa ~ kanəraaʔky 'some', tanua 'few, little', tanəʔiky 'few, small in number') is in the singular, as in Uralic languages in general: nagyɾ basutuə 'three hunters', ɲukəgəə kəntə 'many sledges'. Occasionally, however, dual and plural nouns can occur after a numeral, e.g. sʲiti ni 'two women' (lit. two woman) vs sʲiti ni-gəj two woman-DU 'two women'. The head and the quantifier must partially agree in case, thus the modifier takes the genitive ending while the head noun is in a local case, cf. sʲiði ma-tənu two.GEN tent-LOC 'in two houses'.	
<input checked="" type="checkbox"/>	Т(аньк)е\-то привезла три годик\ уже на пенсии		
Biografija3 vsg	2012		
<input checked="" type="checkbox"/>	Там сколько погибли\ наши, да/		
Biografija3 vsg	2012		
<input checked="" type="checkbox"/>	[vsg] Вот три парни были/ - я хорошо\ помню		

Нганасанский  
[Wagner-Nagy  
2022: 771]





# КОРПУС КОНТАКТНОГО РУССКОГО

- Горномарийский подкорпус: употребление частицы *-то*.

testo\_cypljata hm\_TGA 2018

- ✓ А вы не оставляете яйца, чтобы остались там цыплятки, утятки?  
[hm\_TGA] Ааа, это... Если сядет, так э= **Курица-то сядет-то**. Сядет, так оставим это. Оставим. Потом выводит...

elektriki hm\_EVI 2018

- ✓ В новых домах-то обязательно... В сельской местности... Наружные счётчики ставят. А в **старых-то** можно так было ставить. Ну ладно, чё теперь, сделали - сделали.  
[Interviewer\_EK] Угу, угу, угу.  
[hm\_EVI] Полпенсии забрали.

igry2 hm\_KVK 2017

- ✓ И... На эту... На конец доски наступает... И **палки-то** разлетаются в разные стороны. Который вот охраняет эти палки... Начинает их собирать. А в это время...



# КОРПУС КОНТАКТНОГО РУССКОГО

- Горномарийский подкорпус: употребление частицы *-то*.
- Коммуникативные (артиклеподобные?) функции в русских говорах, ареальные параллели (см. [Селищев 1939; Кузьмина, Немченко 1962; Трубинский 1970; Уздинская 1996; Leinonen 1998, 2002; Stadnik-Holzer 2006; Касаткина 2007, 2008; Мызникова 2019] и др.).
- Корпусные данные позволяют сравнить употребления с дискурсивными показателями POSS.3SG в горномарийском (см. о них [Хомченкова 2022, 2023]).

(11) *p'atačok, tãn'-žã xãna-em li-ät, a mãn'-žã xoza êl-am=êš.*  
Пятчок ты-POSS.3SG гость-POSS.1SG становиться-NPST.2SG а я-POSS.3SG хозяин быть-NPST.1SG=PTCL  
'Пятчок, ты-то будешь моим гостем, а я-то буду хозяином'. [ГК: «Пересказ мультфильма о Винни-Пухе», 17]

[Хомченкова 2022: 152]



# КОРПУС КОНТАКТНОГО РУССКОГО

- В близких планах русская речь носителей татышлинского удмуртского.
- Потенциально другие экспедиции участников группы по языковым контактам.
- Другие коллекции можно добавлять при Вашем желании и участии.



# ПАРАЛЛЕЛЬНЫЕ КОРПУСА

- Обширный проект внутри НКРЯ ([Сичинава 2012, 2015, 2019; Добровольский 2015, 2016; Sitchinava, Perkova 2019; Добровольский, Зализняк 2021, 2022, 2023] и др.).
- Крупные языки (английский, немецкий, хинди и др.) – независимые параллели в переводах, в некоторых случаях влияние за счет культурных контактов.
- Также языки России, плотно контактирующие с русским (башкирский, бурятский, вепсский, карельский, хакасский, чувашский); языки ряда сопредельных государств.
- Взаимовлияние русского и "параллельного" языка.
- "Высвечивание" тех или иных особенностей языка, дальнейшие типологические построения.



# ПАРАЛЛЕЛЬНЫЕ КОРПУСА

- Пример: русско-башкирский параллельный корпус.
- Словоформа *булып* (деепричастие бытийного глагола).
- Различия в конструкциях депиктивного типа; типологическая оценка.

*Выросла* репка *большая-пребольшая*. – Шалкан *зур* *булып* *үскән*.

После окончания в 1956 г. Уфимского нефтяного техникума *работал оператором* на Уфимском нефтеперерабатывающем заводе – 1956 йылда Өфө нефть техникумын тамамлағандан һуң Өфө нефть эшкәртеү заводында *оператор* *булып* *эшләй*.

Мы планировали профинансировать в этом году 300 проектов, но на практике их *оказалось* *в полтора раза больше* – Быйыл 300 проектты финанслау күзалланғайны, әммә улар *бер ярым тапкырға күберәк* *булып* *сыкты*.

Был тауыш тиз генә тыныр өсөн *яралмаған* *булып* *сыкты* – Нет, *не* для того *народился* этот голос, чтобы так скоро замолкнуть (букв.: *вышел, будучи не возникшим*)



# ПАРАЛЛЕЛЬНЫЕ КОРПУСА

- Возможна разработка новых корпусов при Вашем желании и участии.
- Подбор, распознавание, разметка текстов – силами инициаторов корпуса.
- Образцы разметки, техподдержка – команда НКРЯ.

