

Компьютерная лингвистика и интеллектуальные технологии

По материалам ежегодной международной
конференции «Диалог» (2019)

Выпуск 18

Computational Linguistics and Intellectual Technologies

Papers from the Annual International
Conference “Dialogue” (2019)

Issue 18

УДК 80/81; 004
ББК 81.1
К63

Редакционная
коллегия:

*В. П. Селегей (главный редактор),
В. И. Беликов, И. М. Богуславский, Б. В. Добров,
Д. О. Добровольский, Л. М. Захаров, Л. Л. Иомдин,
И. М. Кобозева, Е. Б. Козеренко, М. А. Кронгауз,
Н. И. Лауфер, Н. В. Лукашевич, Д. Маккарти,
П. Наков, Й. Нивре, Г. С. Осипов, А. Ч. Пиперски,
В. Раскин, Э. Хови, С. А. Шаров, Т. Е. Янко*

Компьютерная лингвистика и интеллектуальные технологии:
По материалам ежегодной международной конференции «Диалог»
(Москва, 29 мая — 1 июня 2019 г.). Вып. 18 (25), 2019.

Сборник включает 64 доклада международной конференции по компьютерной лингвистике и интеллектуальным технологиям «Диалог 2019», представляющих широкий спектр теоретических и прикладных исследований в области описания естественного языка, моделирования языковых процессов, создания практически применимых компьютерных лингвистических технологий.

Для специалистов в области теоретической и прикладной лингвистики и интеллектуальных технологий.

© Редколлегия сборника «Компьютерная лингвистика и интеллектуальные технологии» (составитель), 2019

Предисловие

18-й выпуск ежегодника «Компьютерная лингвистика и интеллектуальные технологии» содержит избранные материалы 25-й международной конференции «Диалог». На основании мнений нашего рецензентского корпуса для публикации в ежегоднике редколлекцией были отобраны 64 доклада из ста работ, которые были приняты к представлению на конференции в 2019 году.

Работы в сборнике отражают те направления исследований в области компьютерного моделирования и анализа естественного языка, которые по традиции представляются на конференции:

- Компьютерные лингвистические ресурсы
- Компьютерный анализ документов (классификация, перевод, поиск, саммаризация, генерация, анализ тональности и т. д.)
- Глубокое обучение в NLP (методики применения, содержательная интерпретация)
- Компьютерный анализ Social Media
- Корпусная лингвистика и корпусометрия (методики создания, использования и оценки корпусов)
- Лингвистический анализ текста (морфология, синтаксис, семантика)
- Лингвистические онтологии и автоматическое извлечение знаний
- Мультимодальная коммуникация (включая лингвистический анализ речи)
- Модели общения и диалоговые агенты
- Компьютерная лексикография

В соответствии с традициями «Диалога», старейшей конференции по компьютерной лингвистике в России, отбор работ основывается на представлении о важности соединения новых методов и технологий анализа языковых данных с полноценным лингвистическим анализом. Диалог является де-факто крупнейшим форумом по проблемам создания современных компьютерных ресурсов, моделей и технологий для русского языка.

Одно из ключевых событий «Диалога» — подведение итогов технологических соревнований между разработчиками систем лингвистического анализа текстов, Dialogue Evaluation. В этом году состоялись четыре соревнования:

- автоматическая генерация заголовков новостей;
- автоматический анализ малоресурсных языков (для которых очень мало данных для машинного обучения);
- автоматическое разрешение анафоры и определение референциальных цепочек (различных упоминаний одного и того же объекта в тексте),
- автоматическое восстановление слов по контексту (гэппинг-эллипсис).

В сборник включены наиболее оригинальные работы участников этих соревнований.

Статьи в сборнике публикуются на русском и английском языках. При выборе языка публикации действует следующее правило:

- доклады по компьютерной лингвистике должны подаваться на английском языке. Это расширяет их аудиторию и позволяет привлекать к рецензированию международных экспертов.
- доклады, посвященные лингвистическому анализу русского языка, предполагающие знание этого языка у читателя, подаются на русском языке (с обязательной аннотацией на английском).

Несмотря на традиционную широту тематики представленных на конференции и отобранных в сборник докладов они не могут дать полной картины направлений «Диалога». Ее можно получить с помощью сайта конференции www.dialog-21.ru, на котором представлены обширные электронные архивы «Диалогов» последних лет и все результаты проведенных тестирований Dialogue Evaluation.

Мы обращаем внимание авторов и читателей сборника, что с 2018 года Редаксовет отказался от печати сборника на бумаге, поскольку бумажный вариант пользуется все меньшей популярностью. Сборник, как и в прошлые годы, размещается на сайте конференции и индексируется Scopus.

Программный комитет конференции «Диалог»

*Редакция сборника «Компьютерная лингвистика
и интеллектуальные технологии»*

Организаторы

Ежегодная конференция «Диалог» проводится при организационной поддержке компании АВВУУ.

Учредителями конференции являются:

- Институт лингвистики РГГУ
- Институт проблем информатики РАН
- Институт проблем передачи информации РАН
- Компания АВВУУ
- Филологический факультет МГУ
- Школа прикладной математики и информатики МФТИ

Международный программный комитет

Богуславский Игорь Михайлович	ИППИ РАН, Россия; Мадридский политехнический университет, Испания
Буате Кристиан	Университет Жозефа Фурье — Гренобль 1, Франция
Гельбух Александр Феликсович	Национальный политехнический институт, Мексика
Иомдин Леонид Лейбович	Институт проблем передачи информации РАН им. А. А. Харкевича, Россия
Кобозева Ирина Михайловна	МГУ им. М. В. Ломоносова, Россия
Козеренко Елена Борисовна	Институт проблем информатики РАН, Россия
Корбетт Гревил	Университет Суррея, Великобритания
Кронгауз Максим Анисимович	НИУ «Высшая школа экономики», Россия
Лукашевич Наталья Валентиновна	НИВЦ МГУ им. М. В. Ломоносова, Россия
Маккарти Диана	Кембриджский университет, Великобритания
Мельчук Игорь Александрович	Монреальский университет, Канада
Нивре Йоаким	Уппсальский университет, Швеция
Ниренбург Сергей	Университет Мэриленда, Балтимор, США
Осипов Геннадий Семёнович	Институт системного анализа РАН, Россия
Райгородский Андрей Михайлович	МФТИ, Школа прикладной математики и информатики, Россия
Раскин Виктор	Университет Пердью, США
Селегей Владимир Павлович	Компания АВВУУ, МФТИ, Россия
Хови Эдуард	Университет Карнеги — Меллон, США
Шаров Сергей Александрович	Университет Лидса, Великобритания

Организационный комитет

Селегей Владимир Павлович,
председатель

Беликов Владимир Иванович

Браславский Павел Исаакович

Добров Борис Викторович

Захаров Леонид Михайлович

Иомдин Леонид Лейбович

Кобозева Ирина Михайловна

Козеренко Елена Борисовна

Лауфер Наталия Исаевна

Ляшевская Ольга Николаевна

Пиперски Александр Чедович

Толдова Светлана Юрьевна

Федорова Ольга Викторовна

Шаров Сергей Александрович

Компания АBBYУ

Институт русского языка
им. В. В. Виноградова РАН

Уральский федеральный университет

НИВЦ МГУ им. М. В. Ломоносова

МГУ им. М. В. Ломоносова

Институт проблем передачи информации
РАН им. А. А. Харкевича

МГУ им. М. В. Ломоносова

Институт проблем информатики РАН

Компания Yandex

Институт русского языка
им. В. В. Виноградова РАН

РГГУ

НИУ «Высшая школа экономики»

МГУ им. М. В. Ломоносова

Университет Лидса

Секретариат

Родионова Ольга Игоревна,
координатор оргкомитета

Ульянова Анна Вячеславовна,
секретарь оргкомитета

Компания АBBYУ

РГГУ

Рецензенты

Tania Avgustinova
Vladimir Benko
Anatoly Gersman
Diana Macartney
Preslav Nakov
Piek Vossen
Антонова Александра Александровна
Азарова Ирина Владимировна
Андрианов Андрей Иванович
Апресян Валентина Юрьевна
Артемова (Черняк) Екатерина Леонидовна
Архангельский Тимофей Александрович
Байтин Алексей Владимирович
Богданов Алексей Владимирович
Богданова-Бегларян Наталья Викторовна
Богуславский Игорь Михайлович
Бочаров Виктор Владиславович
Браславский Павел Исаакович
Васильев Виталий Геннадьевич
Галинская Ирина Евгеньевна
Галицкий Борис Александрович
Гельбух Александр Феликсович
Гращенко Павел Валерьевич
Губин Максим Вадимович
Даниэль Михаил Александрович
Добров Борис Викторович
Добровольский Дмитрий Олегович
Добрушина Нина Роландовна
Добрынин Владимир Юрьевич
Дроганова Кира Андреевна
Зализняк Анна Андреевна
Захаров Леонид Михайлович
Иванов Владимир Владимирович
Иомдин Борис Леонидович
Иомдин Леонид Лейбович
Катинская Анисья Юрьевна
Кибрик Андрей Александрович
Князев Сергей Владимирович
Кобозева Ирина Михайловна
Копотев Михаил Вячеславович
Коротаев Николай Алексеевич
Котельников Евгений Вячеславович
Котов Артемий Александрович
Кронгауз Максим Анисимович
Кутузов Андрей
Левонтина Ирина Борисовна
Леонтьев Алексей Петрович
Лобанов Борис Мефодьевич
Лукашевич Наталья Валентиновна
Лютикова Екатерина Анатольевна
Марков Александр Юрьевич
Мисюрев Алексей Владимирович
Недолужко Анна Юрьевна
Новицкий Валерий Игоревич
Пазельская Анна Германовна
Паперно Денис Аронович
Панченко Александр Иванович
Переверзева Светлана Игоревна
Пивоварова Лидия
Пиперски Александр Чедович
Подлесская Вера Исааковна
Смирнов Иван Валентинович
Смуров Иван Михайлович
Селегей Владимир Павлович
Слюсарь Наталия Анатольевна
Сорокин Алексей Андреевич
Тихомиров Илья Александрович
Толдова Светлана Юрьевна
Урысон Елена Владимировна
Усталов Дмитрий Алексеевич
Федорова Ольга Викторовна
Хохлова Мария Владимировна
Циммерлинг Антон Владимирович
Шаврина Татьяна Олеговна
Шаров Сергей Александрович
Шелманов Артём Олегович

Contents*

Апресян В. Ю. Прагматика в интерпретации сфер действия (на материале письменных русских текстов)	1
Апресян В. Ю., Орлов А. В. Семантические типы имплицатур и условия их возникновения (на материале Корпуса газетных заголовков)	17
Badene S., Thompson K., Lorré J-P., Asher N. Learning multi-party discourse structure using weak supervision	30
Баранов А. Н., Добровольский Д. О. Дискурсивные слова в корпусном измерении: одним словом у Достоевского и его современников	41
Baymurzina D. R., Kuznetsov D. P., Burtsev M. S. Language Model Embeddings Improve Sentiment Analysis in Russian	53
Belkin I. BERT finetuning and graph modeling for gapping resolution	63
Богданова-Бегларян Н. В., Блинова О. В., Мартыненко Г. Я., Шерстинова Т. Ю., Зайдес К. Д., Попова Т. И. Аннотирование прагматических маркеров в русском речевом корпусе: проблемы, поиски, решения и результаты	72
Boguslavsky I. M., Frolova T. I., Iomdin L. L., Lazursky A. V., Rygaev I. P., Timoshenko S. P. Knowledge-based approach to Winograd Schema Challenge	86
Bolshakova E. I., Sapin A. S. Comparing models of morpheme analysis for Russian words based on machine learning	104
Bonch-Osmolovskaya A. A., Nesterenko L. V. Multilingual parallel corpora as a source for quantitative cross-linguistic grammar research (the case of voice constructions)	114
Budennaya E. V. Referential choice in multimodal communication	125
Bulygin M. V., Sharoff S. A. Applying an automatic FTD classifier to the annotation of the GICR corpus ..	137

* The reports of each section are ordered by the surname of the first author in compliance with the English alphabet.

Chechuro I. Yu., Lyashevskaya O. N. A Simple Fingerprint Approach to Extracting the Global Prosodic Properties from Field Data	147
Chistova E. V., Shelmanov A. O., Kobozeva M. V., Pisarevskaya D. B., Smirnov I. V., Toldova S. Yu. Classification Models for RST Discourse Parsing of Texts in Russian	163
Dikonov V. G. Simulation of background knowledge and bridging in Russian	177
Dudarin P. V., Tronin V. G., Svyatov K. V. An Approach to Customization of Pre-Trained Neural Network Language Model to Specific Domain	194
Emelyanov A. A., Artemova E. L. Gapping parsing using pretrained embeddings, attention mechanism and NCRF	203
Fomin V., Bakshandaeva D., Rodina Ju., Kutuzov A. Tracing cultural diachronic semantic shifts in Russian using word embeddings: test sets and baselines	213
Gusev I. O. Importance of Copying Mechanism for News Headline Generation	228
Инькова О. Ю. Аннотирование параллельных текстов: понятие «дивергентный перевод»	237
Inshakova E. S. An anaphora resolution system for Russian based on ETAP-4 linguistic processor	249
Иомдин Л. Л. В копилку микросинтаксических неожиданностей: две русские антонимичные синтаксические фраземы с компаративами	262
Khomchenkova I. A., Pleshak P. S., Stoynova N. M. The corpus of contact-influenced Russian of Northern Siberia and the Russian Far East	276
Кибрик А. А., Коротаяев Н. А., Федорова О. В., Евдокимова А. А. Единая мультимедийная аннотация как инструмент анализа естественной коммуникации	288
Князев С. В., Малыгина П. А. Эволюция диалектной системы безударного вокализма в речи жителей Москвы: 4 поколения	304

Кривнова О. Ф., Смирнова О. С. Интроспективная просодическая разметка письменного текста и его реальное озвучивание (сравнительный анализ на материале коллекции текстов Р. И. Аванесова)	318
Kuratov Yu., Arkhipov M. Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language	333
Кустова Г. И. Концептуализация не полностью контролируемых ситуаций: глаголы и местоимения	340
Лапошина А. Н., Веселовская Т. С., Лебедева М. Ю., Купрещенко О. Ф. Лексический состав текстов учебников русского языка для младшей школы: корпусное исследование	351
Le T. A., Petrov M. A., Kuratov Y. M., Burtsev M. S. Sentence Level Representation and Language Models in the task of Coreference Resolution for Russian	364
Левонтина И. Б. Языковые механизмы расширения сочетаемости: сочетаемость частицы -ка	
Левонтина И. Б., Полинская М. С. <i>Достали так употреблять инфинитив!</i> О новой каузативной конструкции в русском языке	384
Likhonosov A., Indenbom E., Yudina M. Automatic vocabulary positioning in a thesaurus	397
Лобанов Б. М., Житко В. А. Анализ просодических признаков эмоциональной интонации с использованием системы «IntonTrainer» (на примере русскоязычных фраз)	408
Lyashevskaya O. N. A Reusable Tagset for the Morphologically Rich Language in Change: a Case of Middle Russian	422
Лютикова Е. А., Герасимова А. А. Послеложные конструкции татарского языка: методики оценки внутриязыкового варьирования	435
Микаэлян И. Л., Зализняк Анна А. Производные значения русского неопределенного наречия как-то: опыт корпусного анализа	458
Movsesyan A. A. An Attention-based Approach to Automatic Gapping Resolution for Russian ...	472

Пекелис О. Е. Слово это в частном вопросе: о признаках, отличающих частицу от местоимения	484
Pereverzeva S. I. Tense and lax body parts in the Russian deictic gestures: the case of index finger pointing	497
Pisarevskaya D., Galitsky B. An Anatomy of a Lie: Discourse Patterns in Ultimate Deception Dataset	513
Подлеская В. И. Просодия и грамматика предикативного сочинения: конструкции с союзом И по данным просодически размеченного корпуса	532
Подлеская В. И., Коротаев Н. А., Мазурина С. И. Самоисправления говорящего в русском монологическом и диалогическом дискурсе: опыт корпусного исследования	547
Rossyaykin P. O., Loukachevitch N. V. Measure clustering approach to MWE extraction	562
Shavrina T. O. Word vector models as an object of linguistic research	576
Шмелев А. Д. Передача церковнославянского текста средствами гражданской графики: можно ли получить ее при помощи формальной процедуры?	589
Smurov I. M., Ponomareva M., Shavrina T. O., Droганova K. AGRR-2019: Automatic Gapping Resolution for Russian	600
Sokolov A. M. Phrase-Based Attentional Transformer for Headline Generation	615
Sorokin A. A. Filling the gaps with rules and networks	622
Sorokin A. A. Morphological parsing of low-resource languages	636
Stankevich M. A., Smirnov I. V., Kuznetsova Y. M., Kiselnikova N. V., Enikolopov S. N. Predicting Depression from Essays in Russian	647
Stepanov M. A. News headline generation using stems, lemmas and grammemes	658
Stoynova N. Some features of the completive prefix do- in Russian: theory faces empirical data	667

Tarasov D., Matveeva T., Galiullina N. Language models for unsupervised acquisition of medical knowledge from natural language texts: Application for diagnosis prediction	677
Tikhomirov M. M., Loukachevitch N. V., Dobrov B. V. Assessing Theme Adherence in Student Thesis	688
Тискин Д. Б. Притяжательные местоимения в русских объектных именных группах	701
Toldova S., Davydova T., Kobozeva M., Pisarevskaya D. Contrast and Comparison Relations in RST framework: the case of Russian ...	714
Vossen P., Baez S., Bajcetić L., Basić S., Kraaijeveld B. A communicative robot to learn about us and the world	728
Вознесенская М. М., Шмелева Е. Я. О проекте словаря «Интертекстуальный тезаурус современного русского языка»: книжный vs. мультимедийный	744
Янко Т. Е. Просодия вопросов с частицей ЛИ	
Зализняк Анна А., Падучева Е. В. Русское что-то как дискурсивное слово	765
Циммерлинг А. В. Корпусная грамматика количественных групп в русском языке	781
Zinina A., Arinkin N., Zaydelman L., Kotov A. The role of oriented gestures during robots communication to a human	800
Zubarev D. V., Sochenkov I. V. Cross-language text alignment for plagiarism detection based on contextual and context-free models	809
Abstracts	821
Авторский указатель	841
Author Index	842

ПРАГМАТИКА В ИНТЕРПРЕТАЦИИ СФЕР ДЕЙСТВИЯ (НА МАТЕРИАЛЕ ПИСЬМЕННЫХ РУССКИХ ТЕКСТОВ)¹

Апресян В. Ю. (valentina.apresjan@gmail.com,
vapresyan@hse.ru)

Национальный исследовательский университет
«Высшая школа экономики»,
Институт русского языка им. В. В. Виноградова РАН

PRAGMATICS IN THE INTERPRETATION OF SCOPE IN WRITTEN RUSSIAN TEXTS

Aprésyan V. Ju. (valentina.apresjan@gmail.com,
vapresyan@hse.ru)

National Research University Higher School of Economics,
Vinogradov Russian Language Institute of the Russian Academy
of Sciences

The paper is a corpus study of pragmatic factors involved in disambiguating sentences with negation and universal quantifier in written Russian and English, such as *Ja ne pozval vseh svoih dal'nih rodstvennikov*, 'I haven't invited all of my distant relatives.' Ambiguity results from differences in scope. If negation scopes over the quantifier, we get partial negation: 'I have invited some, but not all of my distant relatives.' If negation scopes over the verb, we get total negation: 'I haven't invited any of my distant relatives.' Our study is based on Russian and English data extracted from a variety of corpora.

We demonstrate that despite syntactic differences, Russian and English rely on similar mechanisms of disambiguation via pragmatic reasoning. We show that quantifier 'all' has different interpretations with verb vs. quantifier negation: emphatic in the former case and quantificational in the latter. Contextual markers for each reading are consistent with this difference. V-negation occurs with demonstrative pronouns, negatively connoted nouns and temporal modifiers, which add emphasis (*I don't want to talk to all these idiots; I haven't eaten all day*), while Q-negation occurs in the context of quantitative verbs that consolidate the interpretation of quantity (*I haven't listed all the options*).

¹ Публикация подготовлена в ходе проведения исследования по проекту «Factors in resolving scope ambiguity» (№ 18-01-0007) в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2018–2019 гг. и в рамках государственной поддержки ведущих университетов Российской Федерации «5–100».

SOME FEATURES OF THE COMPLETIVE PREFIX *DO-* IN RUSSIAN: THEORY FACES EMPIRICAL DATA¹

Stoynova N. (stoynova@yandex.ru)

Vinogradov Russian Language Institute & Institute of Linguistics,
RAS; NRU HSE; Moscow, Russia

The paper deals with some formal features of the completive prefix *do-* ('to finish, to complete'). It was claimed in previous studies, that this prefix along with some others, has a range of formal properties that differ both from formal properties of productive "superlexical" prefixes (such as the cumulative *na-*, the distributive *po-*) and "lexical" (highly integrated) ones. Two important features were mentioned among others. 1) It can attach both to the perfective stem and to the imperfective one. 2) It cannot attach to secondary imperfectives. In the paper, I verify and develop these claims on corpus data. 1) I propose the rules of choice between the perfective vs. imperfective stem and describe the pool of variation. 2) I show, that, contrary to expectations, in informal speech *do-* attaches to secondary imperfectives quite easily.

Key words: Russian, verb, prefixation, intermediate prefixes, superlexical prefixes, completive

1. Introduction

The paper deals with some formal features of the completive prefix *do-* ('to finish, to complete'). It was postulated in previous theoretical studies on Russian prefixation, that this prefix belongs to a small group of Russian prefixes, intermediate between productive "superlexical" prefixes (which are semantically transparent and occupy the external position within the stem) and "lexical" ones (which are lexicalized and attach directly to the root), cf. [Tatevosov 2008]; [2009]; [2013]. They are predicted to have a range of formal features that distinguish them from both these types.

This paper is empirically-oriented, rather than theoretical. Its aim is to consider these features on quite a large massive of empirical data (dictionary data and corpus data), to verify the theoretical predictions on this prefix and to describe its more specific properties, which do not follow from general theoretical assumptions.

The paper has the following structure. In **Section 2**, I observe briefly the existing classification of Russian prefixes with a special focus on the completive *do-*. In **Section 3**, I differentiate between the completive *do-*, which is in focus of the study,

¹ The study was funded by RFBR, project 17-29-09154 (Trends in the development of language system: a corpus-based study of synchronic variation and diachronic change in different text types).

and some other similar uses of this prefix. In **Sections 4 and 5**, I test on corpus and dictionary data two formal features of *do-*: the ability to attach both to perfective and imperfective stems (4) and the position with respect to the suffix of secondary imperfectivization (5). Section 6 contains a brief conclusion.

2. The completive *do-* within the classification of Russian prefixes

Russian verbal prefixes can be divided at least into two formal types. Prefixes of the first type are less productive, they express more concrete and less transparent meanings, and they are closer to the root within the verb stem: cf. *при-йти*, *вы-йти*, *обо-йти*, derived from *идти* ‘to go’; *при-думать*, *вы-думать*, *об-думать* derived from *думать* ‘to think’. Prefixes of the second type (e.g. the cumulative *на-*, the distributive *по-* and *пере-*) are very productive, they have abstract, transparent meanings, within the verb stem they can attach above prefixes of the first group, cf. *на-* in *на-при-думывать* (CUMUL-LEX.PREF-think.IPFV ‘to think out a lot of things’), *по-* in *по-у-ходить* (DISTR-LEX.PREF-go ‘to go one by one’). This opposition was formulated in impressionistic semantic terms of “qualifiers vs. modifiers” already in [Isatchenko 1965/2003: 222–224]. Later, this idea was developed within the modern formal approaches by [Babko-Malaya 1999]; [Ramchand 2004]; [Svenonius 2004]; [Romanova 2004]; [2006] and others (inner vs. outer prefixes, lexical vs. superlexical prefixes).

A more detailed classification was proposed by S. G. Tatevosov. In [Tatevosov 2008] he introduced one more “intermediate” type of prefixes consisting of the completive *do-*, which is in focus in the paper, and the repetitive *пере-* (and also *под-* which was added to the list in later papers). In [Tatevosov 2009] and [Tatevosov 2013] he also divided other superlexical prefixes to several groups.

In [Tatevosov 2008] proposes a list of 6 formal features of “intermediate prefixes”: some of them are shared with lexical ones, some of them are shared with superlexical ones, and some others are specific for this type.

Two of these features are relevant for this paper.

1) “Aspectual selection”. Superlexical prefixes attach to the imperfective stem: *по-спрашивать* (*po*-ask.IPFV), lexical prefixes behave differently: *от-дать* *ot*-give.PFV, *но за-шить* (*za*-sew.IPFV), while intermediate prefixes are predicted to combine with both types of stems without restrictions: *до-шить* (*do*-sew.IPFV) and *до-подшить* (*do*-shorten.PFV). If both options are allowed by the general model of Russian prefixation, the question arises, on which grounds the stem is chosen in practice.

2) “Position with respect to the secondary imperfective”. Following [Tatevosov 2008], superlexical prefixes can attach above the secondary imperfective (*украсить* (decorate.PFV > *украш-а-ть* (decorate-IPFV) > *по-[у-краш-а]-ть* (*po*-decorate-IPFV)). Intermediate prefixes, including *do-* are, in contrast, predicted to attach only below the secondary imperfective. However, in [Tatevosov 2013] notes that this feature has no connection to other formal properties of this group of prefixes and it does not follow from the general architecture of Russian prefixation, i. e. the opposite situation would not break anything in the general model. If so, it is interesting to check, whether this feature is in fact so strict or the opposite option also can be realized under some conditions.

3. The completive use among other uses of *do-*

In the paper, I discuss the formal properties of *do-* only in its completive uses, i.e. the uses with the meaning 'to finish, to complete, to carry through', as in (1). I include in this type also uses with the meaning 'to complete after a break', as in (2):

- (1) До начала первого писал и, наконец, дописал статейку для «Смены». [RNC] 'I had been writing a paper for "Smena" till 12 o'clock, and then I finished (*do*-+write) it at last.'
- (2) Потом, потом допишу, говорил я себе. А теперь пусть подсохнут листки. [RNC] 'I said to myself: I'll finish writing (*do*-+write) later. And now the papers must dry up.'

There are some other productive uses of *do-*, which are semantically very close to the completive one. However, I exclude them, because their formal properties differ in some respect from those of the completive *do-* and must be described separately. These are:

- 1) spatial uses 'to reach the (spatial) Goal', as in *дойти* (*do*-+go), *долететь* (*do*-+fly), *добежать* (*do*-+run);
- 2) uses with the explicitly expressed endpoint (*до* 'to' + GEN), as in: *досидеть до обеда* 'to be sitting till the lunch-time (*do*-+sit)', *дожить до старости* 'to live to an old age (*do*-+live)', *догореть до середины* 'to burn half (*do*-+burn)';
- 3) additive uses 'V with additional portion of object / with a new object', as in: *долить бензину* 'to pour more petrol (*do*-+pour)', *дорисовать усы* 'to draw a moustache {= to add it to the existing portrait} (*do*-+draw)', *докупить продуктов* 'to buy more food' (*do*-+buy).

The borderline between the completive *do-* and the additive *do-* is especially relevant for the further discussion. In the majority of cases it is quite clear. Usually, the object of a completive *do*-verb is definite, while the object of an additive *do*-verb is indefinite (it follows directly from their semantics). Since Russian does not express definiteness overtly, the following substitution test can be used to differentiate between these two uses. The additive *do-* allows the object modifiers, which show explicitly that a new object (or a new portion of the object) is involved: *еще* 'more', *еще один* 'one more', *новый* 'new' etc., while the completive *do-* does not. Also, only the additive *do-* can take the object in the genitive case² instead of the accusative case. Cf.:

- (3) а. *Допив воду* / **воды*, он вдруг спросил... [RNC] — completive 'Having drunk up (*do*-+drink) the water (water.ACC / *water.GEN)'
- б. Лучшее всего *долить* в смесь *воду* / ^{OK}*воды*. [RNC] — additive 'The best option is to pour (*do*-+pour) some water ('water.ACC / ^{OK}water.GEN') to the mixture'

² Ø-stem masculine nouns take the special partitive case form *-u* in this context (*чаю* 'tea, PART', *сахар-у* 'sugar.PART').

However, there are uses, in which the distinction between the completive *do-* and the additive *do-* is more subtle. A problematic class is that of verbs with the incremental object (such as ‘pour’). For these verbs, a temporal phase of the event (referred to by the completive *do-*) corresponds to the degree of involvement of the object (referred to by the additive *do-*). A special case within this class is verbs with the semantic incremental object, which is referred to by the prefix, but which is not overtly expressed in the syntactic structure. Cf. *досоли́ть* (*do-*+salt) in (4). The verb *соли́ть* ‘to pour salt to sth’ (and its English equivalent ‘to salt’) takes food names as the direct object (*соли́ть суп* ‘to salt the soup’), and its semantic object ‘salt’ is not expressed. Example (4) is ambiguous, since it is unclear, whether ‘salt’ is definite or not and whether we deal with a new portion of salt or with the full normal portion of salt.

- (4) Попробовав с ложки еще раз, Кытин засомневался, подумал, *досолил* и помешал. [RNC]
 ‘Kytin tried (the soup) again, hesitated, salted (the soup) and stirred it.’ =
 ‘added more salt’ (additive) or ‘made salty enough’ (closer to completive)?

4. The choice between the perfective vs. imperfective stem

As mentioned above, an important feature of the completive prefix *do-* is its ability to attach both to perfective and imperfective stems. However, the general model of Russian verbal prefixation, observed in **Section 2**, does not predict the choice of a stem for any particular verb. In this section, I formulate general empirical-based rules of the stem choice. The following data were used: first, the sample of *do-*verbs extracted from Minor Academic Dictionary (678 items); second, the corpus data of RNC (www.ruscorpora.ru) and GICR (<http://www.webcorpora.ru>). The first (dictionary) data source provides information on more conventionalized derivatives which belong to Standard Written Russian. The second (corpus) one provides additional information on the productive derivation in the real use, particularly on occasional derivatives and on those rejected by the prescriptive norm, but attested in informal speech. Such derivatives are especially in focus in the study. They are checked on the data of GICR (the search on blogs: livejournal.com and vk.com).

The rules of stem choice are different for different formal types of aspectual verb pairs. I will describe them separately.

A) Aspectual pairs “unprefixed imperfective + prefixed perfective”

(*писа́ть* — *написа́ть* ‘write’)

Within this type, the completive *do-* attaches consistently to the imperfective stem:

- (5) a. ^{OK}*до-писа́ть* (*do-*+write.IPFV) — ^{???}*до-на-писа́ть* (*do-*+PREF-write.PFV)
 b. ^{OK}*до-шито́* (*do-*+sew.IPFV) — ^{???}*до-с-шито́* (*do-*+PREF-sew.PFV)

Very few counter-examples are attested in written informal texts of blogs: 1–2 uses of *до-про-чита́ть* (*do-*+PREF-read.PFV) и *донаписа́ть* (*do-*+PREF-write.PFV) compared to more than 10,000 uses of *дочита́ть* (*do-*+read.IPFV) and *дописа́ть* (*do-*+write.IPFV) in GICR:

- (6) *Допрочитал-таки* The Probability of God... [GICR]
'In the end (I)'ve finished reading (*do*-+PREF-read.PFV) "The Probability of God"...'
- (7) ... а сегодня *донаписал* кое-как распознавание маджонговых фишек. [GICR]
'...and today, I've somehow finished writing (*do*-+PREF-write.PFV) the code for the recognition of mahjong tokens.'

B) Aspectual pairs "unprefixed perfective + unprefixed imperfective"

(*решить* — *решать* 'solve, decide')

Within this type, the completive *do*- also usually attaches to the imperfective stem:

- (8) Игорь сдал работу и ушел, я *дорешал* свою за двадцать минут. [GICR]
'Igor handed in his exercise and went away, I finished solving (*do*-+solve.PFV) my task in twenty minutes.'

Uses with the perfective stem are also attested in informal speech, though they are rare (9% of uses for the verb *решить* 'solve/decide.PFV', mostly with the meaning 'to decide')³.

- (9) Ну что такого, что я не смогла *дорешить* задачу? [GICR]
'What does it matter, that I failed to solve the task completely (*do*-+solve.PFV)?'

C) Aspectual pairs "prefixed perfective + prefixed secondary imperfective"

(*переписать* — *переписывать* 'rewrite')

This is the most interesting type. For aspectual pairs, in which the prefixed imperfective is derived from the prefixed perfective by means of the imperfectivizing suffix *-ува*(~*-ва*~*-а*), a variation is attested. For one and the same aspectual pair, both the perfective stem and the imperfective one can take the completive *do*-. Cf.:

- (10) – *доуложить* (*do*-+put.to.bed.PFV) and *доукладывать* (*do*-+put.to.bed-IPFV)
а. ...*вот доуложу* сейчас детей и узнаю кто убийца. (www.hv-info.ru)
'Now I will finish putting children to bed and find out who is a murderer.'

– *доуложить*

б. ...ты уложи Андрея, а я тогда *доукладываю* Костю.

(<https://www.babyblog.ru>)

'Put to bed Andrej, and I will finish putting to bed Kostya.' — *доукладывать*

D) Other cases

There are also some marginal cases that do not fit in this picture. One of them is a case of "aspectual triplets". For some verbs, the unprefixed imperfective and the prefixed secondary imperfective are synonymous or quasi-synonymous, cf. *мести* > *подмести* > *подметать* ('to sweep'). In this case, all three stems can take *do*-:

³ This is a small heterogeneous exceptional verb class. Only *решать* — *решить* was consistently checked on corpus data. It is possible that the quantitative data on other verbs differ from those on *решать* — *решить*.

до-мести ~ *до-под-мести* ~ *до-под-мет-а-ть*. Not only two prefixed stems, but also the initial unprefixed one is involved in competition⁴.

One more exceptional case is a narrow class of verbs of attachment with spatial prefixes: *при-/в-/за-крутить*; *при-/в-/за-винтить*; *при-/в-/за-вернуть* ‘to screw on/in’⁵. The unprefixed imperfective stem (*крутить* etc.) is not used with the same meaning. However, *до-* can attach not only to the prefixed stem, but also to this unprefixed imperfective stem, cf.:

(11) Ну не *до-вернули* гайку, ну выпил пивка или еще чего?

[RNC] = *до-за-вернули*, the unprefixed verb **вернуть* does not exist at all

‘We have not screwed on the nut completely, I’ve drunk a bit of beer, what else?’

Table 1 contains the quantitative data on competing derivatives for 2–3 verbs of each type. These are the data on informal Internet-communication taken from GICR.

The table shows, that a real competition is in fact attested only within Type C (prefixed perfective + prefixed secondary imperfective). The majority of such competing completives (both derived from the imperfective and perfective stem) are occasional. If one of them is conventionalized and does not contradict to the prescriptive norm, then it is usually the derivative from the perfective stem (*дорассказать*). However, derivatives from the perfective stem are not obviously more frequent in informal speech (cf. *довышивать* with the opposite distribution). The frequency distribution of competing variants varies a lot across particular verbs. Using such a little sample, I cannot explain which features of a verb predispose to one or another distribution.

Table 1. Competing *до-*derivates from the perfective vs. imperfective stem: GICR⁶

verb type	<i>до-</i> +PFV vs. IPFV	verb	translation	+IPFV	+PFV	% of IPFV
type A	IPFV	<i>читал-прочитал</i>	‘read’	26,770	1	100%
		<i>писал-написал</i>	‘write’	9,192	2	100%
type B	IPFV (~PFV)	<i>решил-решил</i>	‘solve’	71	7	91%

⁴ A more difficult case is a triplet *чесать* > *при-чесать* > *при-чес-ыва-ть* ‘to brush’. It also has three competing derivatives: *до-при-чесать*, *до-при-чес-ыва-ть* and *до-чесать*. The unprefixed *чесать* had been used with the meaning ‘to brush’ till the beginning of the XX cent. (cf. *Марьянка в одной рубашке чесала косу, собираясь спать*. [Л. Н. Толстой. *Казак* (1863), RNC]). However, in modern Standard Russian it is not a neutral synonym of *при-чесывать*. *Мести* is also archaic compared to *подметать* ‘to sweep’. At the same time, the derivatives *до-мести* и *до-чесать* do not seem to be archaic.

⁵ These three verbs are conventionalized and mentioned in dictionaries. Cf. also rare occasional uses of the same class, which are outside the norm, but attested in informal Internet communication: *до-вязать шнурки*, *до-стегнуть крепление*, *до-крепить унитаз*.

⁶ The search on blogs: vk.com (8,720 millions) and livejournal.com (9,820 millions): the forms *рст.ф* and *рст.м*. All the results were looked through manually, only completive uses were counted (see above on the distinction between different meanings of *до-*).

verb type	do-+PFV vs. IPFV	verb	translation	+IPFV	+PFV	% of IPFV
type C	IPFV (~PFV)	вышила-вышивала	'embroider'	316	56	85%
		перечитал(а)-перечитывал(а)	'read again'	22	65	25%
		переписал(а)-переписывал(а)	'write again'	3	10	23%
		рассказал(а)-рассказывал(а)	'tell'	10	202	5%

Competing completives can be used in absolutely identical contexts, (12):

- (12) ...потому что вензель на наволочке не успела довышивать :) Ну что, довышила и поехала! [GICR]
 '... because I didn't have time to finish embroiding (do-+IPFV) a monogram on the pillow-case. Well, I finished it (do-+PFV) and went away.'

However, the following non-strict tendencies in their distribution are attested.

a) The presence of the corresponding perfective vs. imperfective base verb may predispose to the choice of stem (*укладывал, но так и не доукладывал; нужно уложить, вернее доуложить ребенка*).

b) There is a correlation with the frequency of a secondary imperfective derived from the completive verb, which is homonymous to the derivate from the secondary imperfective (*дорассказывать = do-+tell.IPFV 'to finish telling' vs. do-tell+-IPFV 'to be finishing telling'*), see [Section 5](#).

c) The main factor is which component of the event is in focus. If the temporal semantics is in focus, then the derivate from the imperfective stem is more likely. If the argument semantics (the change/involvement of the object in the course of the event) is in focus, then the derivate from the imperfective stem is more likely. Cf.:

- (13) *Контрольные допроверяла!!! Ура!!! Еще чуть-чуть приблизилась к концу семестра.* [GICR] — the imperfective stem
 'I've finished checking the test. Hurrah! Now, I'm a bit closer to the end of the semester!' (the temporal semantics is more in focus⁷)

- (14) *Доделала, допроверила, подправила и выслала клиенту итог двухнедельного труда — 5 свеженарисованных нарядных отчета, каждый на двух языках...* [GICR] — the perfective stem
 'I finished, checked and sent to the client the result of my work of the last two weeks, 5 just written accurate reports, each in two languages.' (the argument semantics is more in focus⁸)

⁷ The temporal interpretation is supported by the next sentence, in which the speaker refers to the time period.

⁸ The speaker focuses not on the time period her work took, but rather on the result of her work, namely on the positive changes in her reports, cf. the other verbs of changing in the chain: *доделала, подправила*.

d) One more argument for the previous point is the fact that the additive *do-*, which is closer to uses that focus the object, than to uses that focus the temporal semantics (see Section 3 above on the semantics of additive uses), are also likely to chose the perfective stem. Cf.:

(15) a. — Ну, эти крупные планы мы *доснимем* в павильоне в Москве.

[RNC] — additive & the perfective stem

‘We will make photo of these close-ups in the studio in Moscow.’

b. Завтра *доснимаю* и *выложу*. — completive & the imperfective stem

‘Tomorrow, I will finish making photos and publish them.’

It is interesting, that the rules regulating the stem choice for *do-* are quite different from those regulating it for *pere-* (which is expected to belong to the same formal type), cf. for the rules for *pere-* [Stoynova 2014].

5. The prefix *do-* and secondary imperfectivization

One more point that requires empirical verification is the position of *do-* with respect to the suffix of secondary imperfectivization *-yva(-va-~a)*.

Completive *do-*verbs, which themselves are perfective, attach the suffix *-yva* without any restrictions, cf.:

– imperfective: *мыть* (wash-IPFV) > perfective: *до-мыть* (*do-wash*) > secondary imperfective: *до-мы-ва-ть* (*do-wash-IPFV*)

(16) Разделся в прихожей и прошел на кухню — там престарелая

домработница Клава *домывала* посуду после ужина. [RNC]

‘(I) took off my clothes and entered the kitchen — there the old housemaid Klava was finishing washing (*do-wash-IPFV*) dishes after dinner.’

On the contrary, the attachment of the completive *do-* itself to secondary imperfectives is claimed to be forbidden. [Tatevosov 2008] considers this feature as one of the arguments to attribute this prefix to a separate intermediate formal type along with *pere-* and *pod-* (see Section 2 above).

In fact, for *pere-* and *pod-* this restriction takes place:

– perfective: *записать* (write.down) > secondary imperfective: *записыва-ть* (write.down-IPFV) > *perfective with *pere-*: **пере-[записыва]-ть* (*pere-write.down-IPFV* ‘to write down again’)⁹

The case of *do-* is more complicated. Derivates from secondary imperfectives are not attested in the sample extracted from Minor Academic Dictionary. However, they are actively used in informal speech. Cf. the verbs *довышивать*, *доукладывать*, *до-рассказывать* and others, mentioned and exemplified in Section 4. Moreover, for

⁹ The homonymous secondary imperfective derived from the perfective verb with *pere-* is attested: *пере-записать* ‘to write again’ > [*пере-запис*]-*ыва-ть* ‘to be writing again’.

some verbs these variants are more frequent, than the expected derivatives from the perfective stem (see **Table 1** above). Cf. one more example:

– perfective: *перечитать* ‘to read again’ > secondary imperfective: *перечит-ыва-ть* (read.again-IPFV) > *до-перечитывать* *do-read.again-IPFV* ‘to finish re-reading’, (17):

- (17) ...надеюсь *доперечитывать* и отправить с приведением любимых фрагментов. [GICR]
 ‘I hope to finish re-reading (*do-read.again-IPFV*) (it) and to send (it) with my favorite fragments marked.’

As both derivational scenarios: *do-V* > [*do-V*]-*ыва* and [*V-ыва*] > *do*-[*V-ыва*] are available, homophonous verbs with different structures and aspectual interpretations are imaginable. And they are in fact attested, cf. *до-рассказ-ыва-ть* (*do-tell-IPFV*) in (18) and (19):

- (18) Вы не дорассказали! Зеро. *Дорассказываю*. Однажды... [RNC] — *до-рассказ-ыва-ть* is the secondary imperfective derived from *дорассказать* (*do-tell*)¹⁰
 ‘— You haven’t finished your story! Zero: I’m telling (*do-tell-IPFV*): Once...’

- (19) Я же ещё не *дорассказывал* сказку=(risovach.ru) — *до-рассказывать* is the perfective completive derivative from *рассказ-ыва-ть* (*tell-IPFV*)¹¹
 ‘But I haven’t finished (*do-tell-IPFV*) my tale yet’.

6. Conclusion

Thus, the empirical data involved in the study, and especially the data of informal speech which lies beyond the prescriptive norm, give the possibility to verify and enlarge the existing assumptions on the completive *do*- and on the whole system of Russian prefixes.

- 1) The prefix *do*- is predicted to attach both to perfective and imperfective stems. The empirical data confirm this prediction and give a possibility to formulate the rules regulating the stem choice, which do not follow from the general theoretical assumptions. The rules are complex, a large pool of variation takes place.

¹⁰ The form *дорассказываю* has the progressive performative interpretation in this sentence, so this is the imperfective verb, derived from *дорассказать*. If it was the perfective derivative from *рассказывать*, it would have the future time reference, which is very improbable in this context (cf. the following sentence with unprefixing verbs: *Рассказываю* (IPFV). *Однажды...* / *???Расскажу* (PFV). *Однажды...*).

¹¹ According to the general rule of interpretation of perfective vs. imperfective verbs in the not-yet context, the imperfective derivative from *дорассказать*, would have the meaning ‘(I) have not begun to finish telling’, and not the meaning ‘I (began) and have not finished telling’ expected from the broader context, cf. the following contrastive pair with unprefixing verbs: *Я же еще не рассказывал сказку* ‘(I) have not even begun to tell the tale’) / *не рассказал сказку* ‘(I) have not finished the tale yet’).

- 2) The prefix *do-* is predicted to attach below the secondary imperfective. It is true for more conventionalized uses. However, numerous occasional uses in modern informal speech totally break this prediction.
- 3) The small class of “intermediate prefixes” appears to be heterogeneous: there are some features that distinguish *do-* from *pere-*, including the interaction with the secondary imperfective and the rules of stem choice.
- 4) So, the detailed empirically-based study does not create considerable problems for the existing theory of Russian prefixation. However, more new data involved — more detailed becomes the classification, up to one prefix classes.

References

1. Babko-Malaya O. (1999), Zero Morphology: A Study of Aspect, Argument Structure, and Case, Ph.D. dissertation, Rutgers University.
2. Isatchenko A. V. (1965/2003), Grammatical structure of Russian compared to Slovak. Morphology [Grammatičeskij stroj russkogo jazyka v sopostavlenii s slovackim. Morfologija], I-II, Moscow, JaSK.
3. Ramchand G. (2004), Time and the event: The semantics of Russian prefixes, Nordlyd, 32(2). Special issue on Slavic prefixes, 2004. pp. 323–366.
4. Romanova E. (2004), Superlexical vs. lexical prefixes, Nordlyd, 32(2). Special issue on Slavic prefixes, pp. 255–278.
5. Romanova E. (2006), Constructing Perfectivity in Russian, Ph.D. dissertation, University of Tromsø.
6. Stoynova N. (2014), Repetitive *pere-* in Russian: single vs. multiple prefixation competing, 47th Annual Meeting of the Societas Linguistica Europaea. Book of abstracts, available at: http://sle2014.eu/downloads/SLE2014BookofAbstracts_FINAL_2.pdf.
7. Svenonius P. (2004), Slavic prefixes inside and outside VP, Nordlyd. Special issue on Slavic prefixes, 32, pp. 205–253.
8. Tatevosov S. G. (2008), Intermediate prefixes in Russian, Antonenko A., Bethin C., Baylin J. (eds.), Formal approaches to Slavic linguistics, New York, Ann Arbor, MI: University of Michigan Press, pp. 423–442.
9. Tatevosov S. G. (2009), Multiple prefixation and anatomy of Russian verb [Množestvennaja prefiksacija i anatomija russkogo glagola], Kiseleva et al. (ed.), Corpus-based studies on Russian grammar [Korpusnyje issledovanija po russkoj grammatike], Moscow, Probel-2000.
10. Tatevosov S. G. (2013), Multiple prefixation and its outcome [Monožestvennaja prefiksacija i ee sledstvija], Voprosy Jazykoznanija, 3, pp. 42–89.