

Е. В. Кашкин, И. А. Хомченкова
(Москва, Россия)

Институт русского языка им. В. В. Виноградова РАН
egorka1988@gmail.com, irina.khomchenkova@yandex.ru

КРУГЛЫЙ СТОЛ «КОРПУСНЫЕ МЕТОДЫ В ИССЛЕДОВАНИИ ЯЗЫКОВЫХ КОНТАКТОВ»

В работе представлена ключевая информация о круглом столе «Корпусные методы в исследовании языковых контактов». Он посвящен корпусам нестандартных вариантов русского языка, контактирующих с иными языками; корпусам некоторых миноритарных языков России и возможностям их использования для анализа контактно обусловленных явлений (в том числе в изучении переключения кодов); параллельным корпусам.

Ключевые слова: корпусная лингвистика, языковые контакты, миноритарные языки, полевая лингвистика, социолингвистика, переключение кодов.

1. Постановка проблемы

Круглый стол, организованный авторами данной работы, посвящен использованию корпусов при исследовании языковых контактов: методологическим и техническим аспектам подготовки корпусов, служащих инструментом таких исследований, подходам к разметке, примерам анализа конкретных явлений.

Тематика языковых контактов приобретает всё большую важность, находясь на стыке теоретической лингвистики и многих других областей (социологии, культурологии, истории, географии и др.) и вписываясь тем самым в современный междисциплинарный контекст, см., например, [Bullock, Toribio (eds.) 2009; Bhatia, Ritchie (eds.) 2013; Grant (ed.) 2019; Matras 2020]. Исследовательские вопросы состоят в том, какие явления и модели чаще, а какие реже подвергаются заимствованию, чем объясняются выявленные закономерности, какие теоретические и когнитивные механизмы скрываются за контактно обусловленными процессами. Изучение вариантов русского языка, развившихся во взаимодействии с другими языками, дополняет эмпирическую картину того, как в принципе может быть устроен русский язык и каковы пределы его внутреннего варьирования.

Важность корпусного метода обусловлена в данном случае тем, что исследования языковых контактов зачастую ведутся методами наблюдения или анкетирования без четко структурированных корпусных данных (ср. традиционные исследования «интерференции» и ее «преодоления» в русском языке), или же корпуса выполняют вспомогательную роль (например, исследователь может найти в обычном корпусе ту или иную конструкцию, а далее с не всегда достаточно эксплицированной аргументацией трактовать ее как появившуюся под контактным влиянием).

Некоторые контактные явления (например, заимствования, переключение кодов) можно исследовать по корпусам, создаваемым в рамках документационных проектов. При этом требует решения вопрос того, как отражать в корпусах такие явления оптимальным для пользователя образом. Кроме того, в последние десятилетия были разработаны корпуса, нацеленные на систематизацию контактно обусловленных явлений (см. раздел 2). При создании этих корпусов принимались различные решения о подборе материала и о принципах его разметки; ср., например, наборы тегов (при их наличии) для нестандартных явлений. Обмен опытом в этой сфере полезен как для совершенствования разработки подобных ресурсов, так и для поиска общих закономерностей и различных черт в лингвистических характеристиках нестандартных вариантов русского языка.

2. Доклады

Программа круглого стола включает пять докладов. Выступление **Е. В. Кашкина** посвящено корпусным ресурсам ИРЯ РАН, предназначенным для изучения русского языка в контактной и типологической перспективах. Во-первых, это корпус русской речи носителей языков России [Ruscontact], см. также [Khomchenkova et al. 2019]. Он включает образцы устной русской речи носителей энецкого, нганасанского, ненецкого, нанайского, ульчского, эвенского, чукотского языков. В настоящее время в корпус интегрируется коллекция русских текстов, записанных от носителей горномарийского языка; эта работа отдельно обсуждается в докладе. В корпусе представлена детализированная разметка языковых явлений, предположительно имеющих контактную природу: нестандартное употребление фонетических единиц (например, озвончение/оглушение), морфологических форм (числа, вида и др.), синтакси-

ческих конструкций и дискурсивных единиц (сбои согласования и управления, особое использование дискурсивного маркера и др.), лексические кальки.

Во-вторых, обсуждаются параллельные корпуса, размещенные на основной платформе НКРЯ (см. о них [Сичинава 2019]). В них размещены тексты разнообразных жанров на литературном русском языке и переводные (в одном или другом направлении) соответствия на других языках. На данный момент в параллельных корпусах обработаны данные более чем 30 языков — как крупнейших языков мира (английского, испанского, хинди, японского и др.), так и языков, носители которых активно контактируют с русским языком в той или иной социолингвистической ситуации (армянского, башкирского, вепсского, хакасского и др.). Использование этого материала дает возможность сопоставительно-типологических исследований русского языка, а также анализа контактных процессов.

Доклад **И. А. Хомченковой** посвящен контактным явлениям, которые можно исследовать с помощью корпусов миноритарных языков, — на материале корпусов горномарийского языка [КГМ] и татышлинского говора удмуртского языка [КТУ]. В качестве иллюстраций рассматривается несколько результатов контактов между соответствующими идиомами (реципиентами) и русским языком (донором). Во-первых, корпуса позволяют узнать, какие русские элементы используются в горномарийской и удмуртской речи (и сравнить результаты, например, с иерархиями заимствуемости по [Matras 2007]). Во-вторых, можно не только оценить их абсолютную частотность, но и сравнить с аналогами (при их наличии) в языке-реципиенте. В докладе внимание фокусируется на русских числительных, союзах и частицах. Показано, что не все иерархии по [Matras 2007] подтверждаются на проанализированном материале. Также обсуждаются ограничения на использование русских элементов с точки зрения (ин)конгруэнтности конструкций языков-реципиентов и языка-донора. Например, условные конструкции и в русском, и в горномарийском включают финитный глагол, однако позиция союзов различается (препозитивный в русском, постпозитивный в горномарийском). Такая неполная конгруэнтность позволяет совмещать структуры двух языков, в результате чего может происходить дублирование союза (*если... збiнь*).

Тематика доклада **Г. А. Мороза** — исследование вариативности в устных корпусах Международной лаборатории языковой конвергенции НИУ ВШЭ (см. сайт [Lingconlab]). Коллектив лаборатории в сотрудничестве с лингвистами из других организаций ведет разработку корпусов ряда малых языков России (абазинского, башкирского, даргинских языков и др.), русских диалектов, корпусов билингвального русского (варианты русского языка, находящиеся в контакте с башкирским, луговым марийским, чувашским, хантыйским и нек. др. языками), других электронных ресурсов. На этом материале с применением статистических методов исследуются, в частности, особенности русской речи с учетом контактно обусловленных, внутриязыковых, социолингвистических факторов, например, нестандартное опущение предлогов [Panova, Philippova 2021; Яковлева (в печати)], отличный от литературного русского синтаксис количественных конструкций [Naccarato, Moroz 2023], ряд других морфосинтаксических явлений.

Выступление **Е. В. Рахилиной** и **А. К. Казкеновой** посвящено еще одному ресурсу, развиваемому в НИУ ВШЭ, — Русскому учебному корпусу [RLC], см., например, [Rakhilina et al. 2016]. Корпус содержит русские тексты, созданные людьми, которые не владеют русским языком как родным и изучают его либо являются эритажными говорящими (т. е. начали усваивать русский в детстве как первый язык и затем, как правило, в результате эмиграции перешли на другой доминантный язык). Как и в упомянутом выше корпусе [Ruscontact], в этом корпусе также размечаются явления, нехарактерные для стандартного русского языка. В качестве иллюстраций и аргументации ряда теоретических выводов авторы представляют результаты работы с казахским подкорпусом Русского учебного корпуса, см., например, [Рахилина, Казкенова 2020; Рахилина и др. 2021; Казкенова, Рахилина 2024].

Н. М. Стойнова рассматривает в своем докладе подходы к отражению переключения кодов в устных корпусах малых языков (ср., например, сочетание лексических и грамматических фрагментов русского и горномарийского языков во взятом из естественного текста предложении *Тӱ сельсӱкохӱзӱштвенный техникумышкыжы мӱнӱӱ после седьмого класса даже поступаенӱм ылы, по конкурсу не прошла* ‘В тот сельскохозяйственный техникум я после седьмого класса даже

поступала, по конкурсу не прошла'). Обсуждаются проблемы документации, транскрибирования, лингвистической разметки подобного материала с опорой, в частности, на опыт разработки корпуса с переключением кодов в ИРЯ РАН под руководством автора доклада [CS corpus] и на опыт корпусов Гамбургского проекта [INEL].

Литература

Казкенова А. К., Рахилина Е. В. От глагольных форм — к служебным словам: грамматикализация в русском и казахском языках сквозь призму текстов билингов // *Russian Linguistics*. 2024. № 48 (16). С. 1–23.

КГМ — Корпус горномарийского языка [Электронный ресурс]. URL: <https://hillmari-exp.tilda.ws/corpus> (дата обращения: 03.12.2024).

КТУ — Корпус татышлинского удмуртского [Электронный ресурс]. URL: <https://udmurt.web-corpora.net/tatyshly/> (дата обращения: 03.12.2024).

Рахилина Е. В., Казкенова А. К. Маркирование итератива в русской речи носителей казахского языка // *Труды Института русского языка им. В. В. Виноградова*. 2020. №4. С. 168 — 192.

Рахилина Е. В., Казкенова А. К., Ахапкина Я. Э. После, через, спустя во временных контекстах: из наблюдений над текстами казахско-русских билингов // *Вестник Томского государственного университета. Филология*. 2021. № 73. С. 93–113.

Сичинава Д. В. Параллельные тексты в составе Национального корпуса русского языка: новые языки и новые задачи // *Труды Института русского языка им. В. В. Виноградова*. 2019. № 3. С. 41–61.

Яковлева А. В. Опускание предлогов в русской речи марийских и бесермянских билингов: исследование на основе устных корпусов // *Труды института русского языка им. В. В. Виноградова*. В печати.

Bhatia T., Ritchie W. (eds.). *The handbook of bilingualism and multilingualism*. Oxford: Wiley Blackwell, 2013. 976 pp.

Bullock B., Toribio A. (eds.). 2009. *The Cambridge handbook of linguistic code-switching*. Cambridge: Cambridge University Press. 422 pp.

CS corpus — Корпус устных текстов с разметкой переключения кодов [Электронный ресурс]. URL: <http://web-corpora.net/ruscontact/CS.html> (дата обращения: 03.12.2024).

Grant A. (ed.). The Oxford handbook of language contact. Oxford: Oxford University Press, 2019. xxix + 757 pp.

INEL — INEL Ressourcen Portal [Online resource]. URL: <https://inel.corpora.uni-hamburg.de/portal/#en> (date of access: 03.12.2024).

Khomchenkova I., Pleshak P., Stoynova N. The corpus of contact-influenced Russian of Northern Siberia and the Russian far East // Computational linguistics and intellectual technologies: Proceedings of the international conference “Dialogue 2019”. M.: RSUH, 2019. P. 253–264.

Lingconlab — Ресурсы Международной лаборатории языковой конвергенции НИУ ВШЭ [Электронный ресурс]. URL: <https://lingconlab.ru/> (дата обращения: 03.12.2024).

Matras Y. The borrowability of grammatical categories // Grammatical borrowing in cross-linguistic perspective / Y. Matras and J. Sakel (eds.). Berlin, NY: Mouton de Gruyter, 2007. P. 31–74.

Matras Y. Language contact. Cambridge: Cambridge University Press, 2020. xix + 409 pp.

Naccarato Ch., Moroz G. Non-standard numeral constructions in L2 Russian: A corpus-based study // Talk at the conference "Indigenous languages of Russia in contact with Russian" (Moscow, Vinogradov Russian Language Institute of the RAS, 15–16 September 2023).

Panova A., Philippova T. When a cross-linguistic tendency marries incomplete acquisition: Preposition drop in Russian spoken in Dagestan // International journal of bilingualism. 2021. № 25(3). P. 640–667.

Rakhilina E., Vyrenkova A., Mustakimova E., Ladygina A., Smirnov I. Building a learner corpus for Russian // Proceedings of the joint workshop on NLP for computer assisted language learning and NLP for language acquisition at SLTC, Umeå, 16th November 2016. [Online resource]. URL: <http://aclweb.org/anthology/W16-65> (date of access: 03.12.2024).

RLC — Русский учебный корпус [Электронный ресурс]. URL: <http://web-corpora.net/RLC/> (дата обращения: 03.12.2024).

Ruscontact — Корпус русской речи носителей языков России [Электронный ресурс]. URL: <http://web-corpora.net/ruscontact/corpus.html> (дата обращения: 03.12.2024).