

RUSKELL 1.3: ОНЛАЙН-ИНСТРУМЕНТ ДЛЯ ИЗУЧЕНИЯ РУССКОГО ЯЗЫКА КАК ИНОСТРАННОГО – СОВЕРШЕНСТВОВАНИЕ И ИСПОЛЬЗОВАНИЕ

Культепина О. (okultepina@gmail.com),

Буйволова О. (ovbuyvolova@edu.hse.ru),

Малолетняя А. (maloletnyaya@gmail.com)

НИУ ВШЭ, Москва, Россия

RuSkELL 1.3: the online language learning tool for Russian language – upgrade and usage

Kultepina O.A. (okultepina@gmail.com), Buivolova O.V. (ovbuyvolova@edu.hse.ru),

Maloletnyaya A.P. (maloletnyaya@gmail.com)

The paper deals with the upgrading process of RuSkELL ("Russian + Sketch Engine for Language Learning") – the online resource for researchers and learners of Russian language (<https://www.sketchengine.co.uk/ruskell-examples-and-collocations-for-learners-of-russian/>).

The tool is based on a specially pre-processed corpus (from the Russian Internet in 2011) and has an intuitive user-friendly interface. Russian language learners can search for phrases in sentences, extract salient collocates and find similar words using RuSkELL.

The student project group focuses on word sketch's adaptation, testing and working out the ways of using the tool in practice. This paper analyses the upgrades which have to be done or are supposed to be done in the short run. It concerns the addition of prepositions, cleaning the outputs, word sketch's renaming and revising sketch grammar. Also, current paper illustrates the probable practical usage of the tool in Russian as a Foreign Language.

Keywords: online language tool; Sketch Engine for Language Learning; collocations; Russian as a Foreign Language.

1. Введение

RuSkELL был создан на основе самого большого корпуса русского языка Russian Web 2011 (ruTenTen11) в системе Sketch Engine по модели английского онлайн-ресурса SkELL с учетом специфики русского языка. Подготовка и чистка корпуса, уменьшение его объема в соответствии с учебными целями ресурса были проведены специалистами¹ до непосредственной работы проектной группы. Деятельность группы с 2015 года по настоящее время заключается в тестировании ресурса, выявлении ошибок в выдаче и её анализ, разработка корректных решений для уменьшения “шума” в выдаче. В процессе работы над улучшением RuSkELL мы находим источники технических ошибок и/или ошибок в скетч-грамматике, пытаемся теоретически обосновать их возникновение с учетом особенностей русской грамматики и семантики, готовим рекомендации по практическому использованию ресурса.

В прошлом году проектная группа работала над созданием скетч-грамматики порядковых и количественных числительных и тестировала пробную версию правила для служебных частей речи (которые впоследствии не были включены в скетч-грамматику из-за нерелевантной выдачи). Кроме того, была проведена работа с небольшой фокус-группой потенциальных пользователей ресурса, по результатам которой было решено изменить названия коллокационных групп и ввести виртуальную клавиатуру.

В этом году группа продолжила работу, и результаты этой работы будут изложены ниже.

2. Краткий обзор функционала

RuSkELL представлен тремя функциями: “Примеры”, “Сочетаемость слова” и “Похожие слова”.

Выдача поискового слова в контексте представлена во вкладке “Примеры” в количестве не больше 40 предложений средней длины со словами средней частотности (алгоритм GDEX – Good Dictionary EXamples). Благодаря разнообразию документов, составивших корпус RuSkELL, и механизму их “чистки и обработки”², пользователь ресурса может ознакомиться как с привычно-нейтральными контекстами, так и специальными терминологическими, жаргонными или сленговыми.

Другая функция RuSkELL – “Сочетаемость слова”, в которой представлены группы различных коллокаций со словом-запросом. Каждая из групп дает полное представление о наиболее частотных грамматических отношениях слова и позволяет ознакомиться с контекстами употреблений. Функция “Сочетаемость слова” демонстрирует выдачу с коллокациями, полуидиоматическими и идиоматическими сочетаниями, которые не всегда прописывают в словарях, что ценно для изучающего русский язык.

Третьей функцией RuSkELL являются “Похожие слова”, которые позволяют пользователю посмотреть близкие по семантическим и сочетаемым характеристикам слова. Слова представлены как списком, так и в облачном формате и распределены по

¹ Проведены Андреем Шестаковым, Екатериной Черняк и Тимуром Исхаковым. См. подробнее: <https://www.sketchengine.co.uk/ruskell-examples-and-collocations-for-learners-of-russian/>

² См. предыдущую сноску.

величине схожести и частотности появления в корпусе. В список “Похожих слов” попадают гипонимы, гиперонимы, синонимы и антонимы, аналоги, конверсивы, видовые корреляты (для глаголов), найденные по семантической близости и одинаковым грамматическим отношениям со схожими коллокатами.

Более подробно о функциях RuSkELL можно прочитать в статье: Apresjan V., Baisa V., Buivolova O., Kulterina O., Maloletnjaja A. *RuSkELL: Online Language Learning Tool for Russian Language* (2016).

RuSkELL может быть полезен для студентов РКИ, которые хотят более углубленно изучать лексику русского языка, для преподавателей РКИ, которые хотят использовать новые корпусные методики в преподавании, и для лингвистов, которые могут найти в ресурсе интересный материал для исследований в области лексикографии, лексической семантики и синтагматики.

3. Количественные и качественные изменения в работе RuSkELL

3.1. Интерфейс

Один из самых активно обсуждаемых в проекте вопросов касается перевода технических названий коллокаций в удобный для пользователя вид. В прошлом году мы разработали два варианта названий, один из которых был запущен на платформу (Таблица 1, графа “Вариант 2.2”). Но после сбора мнений преподавателей и студентов РКИ, которые нашли названия слишком громоздкими, был создан третий вариант названий. Мы считаем, что такой вариант перевода технического названия будет корректно и лаконично передавать суть грамматических отношений коллокации, а также будет одинаково понятен и удобен как для студентов РКИ, так и для лексикографов.

Техническое название	Вариант 2.2.	Вариант 3.1.
и/или	и/или	%w и/или + существительное _{nom}
subject/subject_of	подлежащее при %w/глагол с %w в роли подлежащего	%w _{nom} + глагол / %w + существительное _{nom}
object2/object2_of	дополнение в родительном падеже/дополнение в родительном падеже при глаголах	глагол + %w _{gen} / %w + существительное _{gen}

Таблица 1. Примеры названия групп сочетаемости в RuSkELL

3.2. Работа со скетч-грамматикой

Поисковое слово получает свой список коллокаций благодаря работе скетч-грамматики, своду правил для соединения запросного слова со словом с конкретными грамматическими свойствами, по принципу: w1 gram_rel w2. Скетч-грамматика – главный объект практической деятельности проекта, так как все изменения, связанные с улучшением выдачи, вносятся именно в ее правила. Изменения в скетч-грамматике идут

по двум направлениям: добавление правил для увеличения выдачи в группах “Сочетаемости слова”; изменение и корректировка существующих правил для “чистки” выдачи.

3.2.1. Добавление предлогов в скетч-грамматику

Экспериментальным шагом в работе с RuSkELL стало составление правила для сочетаний с предлогами-запросами. Анализируя корпус, мы обнаружили омонимию в выдаче глаголов, наречий и существительных, от которых образованы производные предлоги (например, предлог *несмотря* получил тег “глагол”), и отсутствие скетчей для предлогов как отдельных частей речи. Хотя в корпусе все предлоги имеют разметку, до данного момента у них не было собственных правил в скетч-грамматике. Поэтому они не могли иметь свой упорядоченный лист сочетаний и из-за автоматической лемматизации могли получать теги других частей речи. Мы решили разработать правило, чтобы предлоги получили выдачу в “Сочетаемости слова”.

Создание правил для всех предлогов невозможно. Особенности поисковой системы RuSkELL не позволяют использовать составные лексикализованные единицы в качестве запросов, поэтому сложные предлоги, вроде *несмотря на*, пока что игнорируются RuSkELL. Однако включение правила для односоставных предлогов с простыми отношениями *предлог + существительное*, *глагол + предлог + существительное* остается возможным. Правило имеет следующий вид и строит сочетания для а) любого простого предлога при существительном в любом падеже, б) любого простого предлога после глагола и перед существительным:

**DUAL*

= %w + существительное

#=prep_object

1: "Sp.*" **2:** noun_nom

1: "Sp.*" **2:** noun_gen

1: "Sp.*" **2:** noun_dat

1: "Sp.*" **2:** noun_acc

1: "Sp.*" **2:** noun_inst

1: "Sp.*" **2:** noun_loc

3: [tag="Sp.*"] **2:** adj [word=", "]? **1:**noun & agree (1,2)

**TRINARY*

= глагол + %w + существительное

#=pp_%(3.lemma)

1: "V.*" **4:** [tag!="SENT" & tag="Sp.*"] **3:** adj [word=", "]? **2:** noun & agree(2,3)

Практическое значение выдачи предлогов – показ предлогов как многозначных лексических единиц. Самый простой пример – предлог *в*, имеющий пространственное и временное значения (*в доме* и *в 2011 году*). При этом мы понимаем, что студенты, изучающие РКИ, могут работать со скетчами предлогов только на высоком уровне владения русским языком или при наличии подробных инструкций и тьютора.

3.2.2. Корректировочные добавления в скетч-грамматике

Мы включили в новую версию скетч-грамматики RuSkELL расширение для правила с грамматическим отношением *глагол+ сущ в Тв.п.*, добавив в него прилагательное в Тв.п.:

DUAL

#=inst_modifier/inst_modifies

2: *adj_ins* **1:** "Vm.*"

Необходимость этого правила обсуждалась ещё в прошлом году, так как депиктивные конструкции часто вызывают сложности у изучающих РКИ, особенно в отношении глаголов, с которыми возможно употребление депиктивов. Поэтому в этом году мы включили правило в скетч-грамматику и получили релевантную выдачу. Результаты выдачи по некоторым глаголам в группе *inst_modifier/inst_modifies* с дополненным правилом приведены в Таблице 2:

Глагол-запрос	Выдача
выглядеть	<i>странным выглядит отсутствие; логичным выглядит сохранение; интересным выглядит включение; перспективным выглядит участие</i>
казаться	<i>странным казался этот вопрос</i>
оказаться	<i>эффективным оказался режим; успешной оказалась работа; интересным оказался раздел; удачным оказался спектакль; лучшими оказались рисунки</i>

Таблица 2. Выдача коллокация по правилу: **2:** *adj_ins* **1:** "Vm.*"

Таким образом, в выдачу попали самые частотные в корпусе сочетания *глагол + прилагательное в Тв.п.* Это дополнение интересно для изучающих русский язык, особенно иностранных студентов, часто ошибочно выбирающих после указанных глаголов прилагательные в форме Им.п.

Другая корректировка связана с идеей внести список глаголов с управлением в скетч-правила для того, чтобы избежать грамматической омонимии в выдаче. Подобный подход может решить проблему генитивных глаголов (т.е. глаголов, управляющих прямым дополнением в Р.п.). Их список невелик, но они имеют определенный “шум” (из-за омонимии с Д.п. и с В.п.) в выдаче. Например, глагол *достигнуть* имеет группы сочетаний и с Д.п., и с В.п., в выдаче которых на самом деле оказываются грамматические омонимы в Р.п.

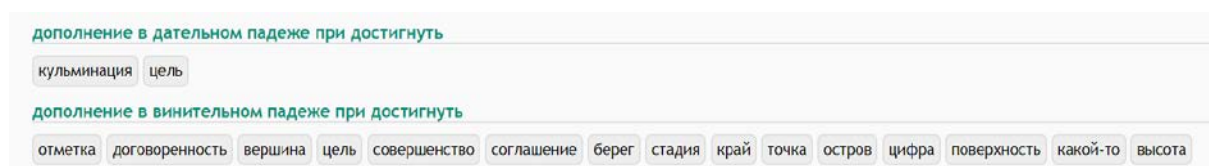


Иллюстрация 1. Выдача по запросу *достигнуть*: грамматическая омонимия

Для решения этой проблемы было предложено ввести небольшой список генитивных

глаголов в правило скетч-грамматики. Следующий код в скетч-правиле, регулирующем коллокации запросного слова с дополнением в В.п. (выделено полужирным; приведена только часть правила), поможет убрать омонимию в выдаче, так как он включает запрет на сочетание генитивного глагола с существительным, имеющим одинаковое окончание в В.п. и Р.п.:

#=object4/object4_of

1:"Vmi.*" & [lemma!="стоять|казаться|достигнуть|держаться|добиваться|добиться"]
[pos="R"]? adj_acc {0,3} **2:**noun_acc

Подобное правило только с некоторыми из указанных глаголов также будет добавлено и в группу коллокаций с Д.п.

4. Изучение языка при помощи RuSkELL: возможности и подходы.

Главное условие корпусного подхода в преподавании, отмеченное во многих работах (Kilgarriff et al. 2015, Thomas 2016), – наличие преподавателя-тьютора и подробных наводящих инструкций. В случае с RuSkELL работа студента с корпусом может дать ему множество возможностей, например:

- 1) найти новые значения уже известных слов в контекстах (например, для глагола *кинуть* – *кинуть на деньги* или *кинуть в личку*). В этом случае RuSkELL может служить и как справочным ресурсом, и как источником контекстов для работы. Например, с его помощью можно составить задание «Заполните пропуски...» и предложить список контекстов с пропусками, которые студент восстанавливает из «Примеров» или «Сочетаемости слова»;
- 2) понять возможности подстановки синонимов и запретов на замены в контекстах. При помощи функции «Похожие слова» студент знакомится с близкими по семантическим и коллокационным свойствам словами. Благодаря этому студент может научиться правильно подбирать синонимы и антонимы в заданных контекстах, различать оттенки значений и пр. Например, при сравнении выдачи похожих слов для однокоренных прилагательных *экономический* и *экономичный* студент понимает разницу значений и употреблений, а также (подключив сравнение скетчей) отличия в сочетаемости;
- 3) расширить познания в лексической (не)сочетаемости слов и ограничений на использование синонимов (например, *заядлый* и *закадычный*).

Работа с инструментом должна преподноситься как новый формат работы со «словарями»: RuSkELL дает разъяснения и примеры там, где учебный или толковый словари не предоставляют достаточно информации. Наиболее привлекательной для образовательных целей является функция «Сочетаемость слова». Как правило, иностранные студенты, особенно на продвинутом уровне, испытывают сложности при выборе формы слова в словосочетании (*доверять рекламу* или *доверять рекламе*), а также непосредственно при построении словосочетания. Например, вопросы могут вызвать существительные, сочетающиеся с прилагательным *низкий*, особенно редкие коллокации вроде *низкий поклон*.

Использование RuSkELL в качестве дополнительного материала по объяснению

особенностей грамматики и лексики русского языка возможно только при руководстве тьютора или преподавателя.

Библиография

Apresjan V., Baisa V., Buivolova O., Kultepina O., Maloletnjaja A. (2016), RuSkELL: Online Language Learning Tool for Russian Language. Proceedings of the XVII EURALEX International Congress (Tbilisi, 6-10 September 2016), pp. 292-300.

Atkins BT Sue, Rundell M. (2008), The Oxford guide to practical lexicography. Oxford University Press, available at:

<http://site.ebrary.com/lib/hselibrary/reader.action?docID=10246237&ppg=6;>

Baisa V., Suchomel V. (2014), SkELL: Web Interface for English Language Learning. // *Eighth Workshop on Recent Advances in Slavonic Natural Language Processing*, 2014.

Baisa, V., Suchomel V. (2014), SkELL: Web Interface for English Language Learning, available at:

[https://www.researchgate.net/publication/275007862_SkELL_Web_Interface_for_English_Language_Learning;](https://www.researchgate.net/publication/275007862_SkELL_Web_Interface_for_English_Language_Learning)

Kilgarriff A., Baisa V., Bušta J., Jakubíček M., Kovář V., Michelfeit J., Rychlý P., Suchomel V. (2014), The Sketch Engine: ten years on. *Lexicography*. July 2014, Volume 1, Issue 1, pp 7-36

Kilgarriff A., et al (2008). GDEX: Automatically finding good dictionary examples in a corpus. *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*

Kilgarriff A., Marcowitz F., Smith S., Thomas J. (2015), Corpora and Language Learning with the Sketch Engine and SKELL. // *Revue Française de Linguistique Appliquée* 20(1), pp. 61-80.

Rusgram - Проект корпусного описания русской грамматики (<http://rusgram.ru>)

RuSkELL <http://ruskell.sketchengine.co.uk/run.cgi/concordance?lpos=&query=1>.

Russian National Corpus <http://www.ruscorpora.ru/>.

SkELL - What is SkELL? <https://www.sketchengine.co.uk/skell/>

Thomas 2016 — Thomas J. *Discovering English with Sketch Engine. A corpus-based approach to language exploration*, VERSALITE, 2016.