

RU

Проблемы исследования словообразовательного потенциала: автоматизированное конструирование гипотетических дериватов

Васильева Е. В., Ташлыкова М. Б.

Аннотация. Цель настоящей работы – обосновать использование автоматизированных методов для изучения деривационного потенциала русской лексики и продемонстрировать принципы разработки алгоритма, позволяющего моделировать производные существительные со значением «носитель психического признака». Научная новизна исследования состоит в разработке нового подхода к отбору дериватов: выявление некодифицированных производных осуществляется не случайным образом, а системно – с помощью автоматизированного конструирования гипотетических дериватов и последующей проверки наличия/отсутствия каждого из них в интернет-пространстве. В ходе работы определяется перечень прилагательных, обозначающих психические характеристики человека, и суффиксов, способных участвовать в образовании имен лиц. Затем разрабатывается скрипт на языке Python, автоматически моделирующий производные на основе данного материала. Кроме того, осуществляется фильтрация полученных производных, исключающая некорректные сочетания морфем. Результатом исследования является перечень гипотетических дериватов и готовый алгоритм, который может быть адаптирован для конструирования производных единиц, образуемых по другим моделям с иным набором производящих основ и аффиксов.

EN

Problems in the study of derivational potential: automated construction of hypothetical derivatives

E. V. Vasileva, M. B. Tashlykova

Abstract. The aim of the present study is to justify the use of automated methods for investigating the derivational potential of Russian lexicon and to demonstrate the principles of developing an algorithm capable of modeling derived nouns with the meaning of “a bearer of a mental attribute”. The scientific novelty of the research lies in the development of a new approach to the selection of derivatives: the identification of uncodified derivatives is carried out not randomly but systematically – through automated construction of hypothetical derivatives followed by verification of their presence or absence in the internet space. During the study, a list of adjectives denoting mental characteristics of a person and suffixes capable of participating in the formation of personal nouns is determined. Subsequently, a script in Python is developed that automatically models derivatives based on this material. Additionally, the filtering of the generated derivatives is performed to exclude incorrect morpheme combinations. The result of the research is a list of hypothetical derivatives and a ready algorithm that can be adapted for constructing derived units formed by other models with different sets of stems and affixes.

Введение

Выяснение причин «расхождения между потенциями системы словообразования и тем, как их реализует норма языка» (Земская, 1973, с. 204), остается одной из актуальных задач деривационных исследований.

Одним из ярких примеров подобных расхождений являются лакуны в составе словообразовательных парадигм (далее – СП) имен прилагательных, обозначающих психические характеристики человека. Производные, называющие лицо по признаку, представлены далеко не во всех парадигмах, ср.: *жадина* ← *жадный*, * ← *щедрый*, *весельчак* ← *веселый*, * ← *грустный*, *скромняга* ← *скромный*, * ← *дерзкий*, *строптивец* ← *строптивый*, * ← *кроткий* и т. п.

Факты такого рода неоднократно привлекали внимание исследователей, изучающих конкретные и типовые СП или словообразовательный потенциал частей речи (Лыткина, 2001; Белякова, 2007; Голев, Фаломкина, 2021; Замятина, 2023; Ло, 2024).

Заметим, что подходы «от парадигмы» и «от производящего слова», различаясь тем, что находится в фокусе внимания лингвиста (результаты деривационного процесса в первом случае или его источник – во втором), содержательно не могут не быть близки, поскольку так или иначе сосредоточены на выявлении тех факторов, которые стимулируют или, наоборот, сдерживают реализацию словообразовательных потенциалов. Ясно, что отсутствие ограничений обеспечивает возможность создания производного слова и, соответственно, отсутствие лакуны в СП; наличие ограничений порождает существование пустых клеток в ее составе.

В ходе изучения данной проблематики были получены результаты, нашедшие отражение не только в научной литературе, но и в вузовских учебниках по словообразованию. Был, в частности, составлен перечень ограничений на словопроизводство, в числе которых ограничения формальные, лексические, семантические, словообразовательные, стилистические (Земская, 1973).

Важно, однако, подчеркнуть, что все эти закономерности выявлены и осмыслены на основании анализа материала, зафиксированного в традиционных лексикографических источниках:

- Словарь русского языка: в 4 т. / гл. ред. А. П. Евгеньева. М., 1981-1984.
- Тихонов А. Н. Словообразовательный словарь русского языка: в 2 т. М.: АСТ; Астрель, 2008.

Так, отсутствие производных со значением «носитель признака (лицо)» от прилагательных типа *вороватый*, *хамоватый* в Словообразовательном словаре А. Н. Тихонова, а также неимение упоминаний о сочетаемости основ на *-оват(ый)* с формантами, образующими отадективные наименования лица, в Русской грамматике (1982) (РГ-80) позволяют заключить, что реализация деривационного потенциала прилагательных на *-оват(ый)* блокируется формальными ограничениями (Васильева, 2013; 2016).

Аналогичным образом можно предположить наличие лексических ограничений в тех случаях, когда потенциальные отадективные дериваты конкурируют с уже существующими отглагольными именами лиц. Так, производные от *хвастливый* и *молчаливый* со значением «носитель признака (лицо)» не образуются, поскольку соответствующие места уже заняты словами *хвастун* (от *хвастать*) и *молчун* (от *молчать*).

Новые возможности изучения динамики языкового развития, появившиеся в условиях высокотехнологичной коммуникации и наличия online доступа к большим массивам данных, делают актуальной задачу проверить и подтвердить, опровергнуть или уточнить сложившиеся теоретические представления.

Уже самый предварительный анализ материала показывает, что перечисленные выше ограничения не являются абсолютными. Приведем несколько показательных примеров.

Прилагательное *вороватый*, судя по данным, полученным с использованием Яндекс Поиска, может образовывать существительное *вороватик*:

(1) *В Эстонии экономика лучше, потому что там у власти не стоят латышские вороватики* (Наш город: новости Даугавпилса. Комментарии за 11.01.2013. <https://gorod.lv/lenta/11.01.2013/kommentarii?mesto=&page=2>).

Слова *хвастливый* и *молчаливый* могут выступать производящими для дериватов *хвастливец* и *молчаливец* соответственно:

(2) *Гриффиндор – самовлюблённые хвастливицы, Когтевран – зануды, Пуффендуй – безвольные куклы, Слизерин – нарциссичные садисты* (Трикки: клуб любителей тестов. 4Ролевая игра по ГП4: битва за Хогвартс. 2021. <https://trikky.ru/forum/topic/rolevaya-igra-po-gp-bitva-za-hogwarts?part=437>).

(3) *Интересный ЖЖ у него какой, у молчаливца*) (Традиционный пятничный эльфлинг // LiveJournal. 01.11.2013. <https://marat-ahtjamov.livejournal.com/725763.html?ysclid=maxb5cwug1142607123>).

Подобные примеры показывают, что изучение деривационного потенциала частей речи не может основываться только на лексикографических данных, поскольку «на фоне активной экспансии компьютерно-опосредованной коммуникации состав и функциональность практически всех живых языков мира изменяются <...> и динамику языкового развития, которая раньше не спеша описывалась постфактум, сегодня... необходимо выявлять и квалифицировать в режиме реального времени» (Астапкина, Баркович, 2025, с. 98). В свете сказанного отбор единиц для анализа становится ключевой исследовательской задачей. Неполнота эмпирической базы создает риск неточных теоретических интерпретаций, искажая представления о границах реализации деривационных возможностей слова, о феномене лакунарности, о факторах, способствующих участию/неучастию лексических единиц в процессах словопроизводства.

Лингвисты по-разному решают проблему ограниченности материала, представленного в словообразовательных словарях.

Одни исследователи обращаются к дополнительным лексикографическим источникам (Ло, 2024; Евсеева, Чжао, 2023). Другие расширяют эмпирическую базу, используя данные Национального корпуса русского языка (Замятина, 2023), средств массовой информации (Николина, Рацибурская, Фатхутдинова, 2020), форумов, социальных сетей и иных интернет-платформ пользовательской коммуникации (Радбиль, Ясаи, Палоши, 2022; Ли, 2020).

В условиях дефицита актуальных лексикографических данных для отбора дериватов привлечение интернет-ресурсов становится важным инструментом исследования современных словообразовательных тенденций: «Сама возможность поиска в Интернете неузуальных неодериватов представляется актуальной и информативной, поскольку на разнообразных по жанру интернет-сайтах, вовлеченных браузерами в поисковый процесс, впервые в истории коммуникации масштабно фиксируется новейший «речевой материал» (Петрухина, Дедова, 2019, с. 153-154).

Этот тезис не вызывает сомнений; однако здесь возникает важный вопрос, связанный с технологией поиска. Если исследователя интересуют неузуальные единицы определенного типа (как, например, нас интересуют номинации лица по психическому признаку), как следует организовать их поиск, чтобы быть уверенным

в полноте отбора? Сложность задачи усугубляется тем, что отадеквативные имена лиц образуются в русском языке с помощью 28 суффиксов (словообразовательных формантов), многие из которых представлены набором алломорфов (РГ-80), например: *-ag(a)*, *-ец-* / *-авец-*, *-ик-*, *-к(a)* / *-ушк(a)* / *-ашк(a)* / *анк(a)* / *инк(a)* / *-овк(a)* и т. д.

Соответственно, если мы хотим уточнить, существуют ли в реальной речевой практике производные имена лиц, например, от прилагательного *безмозглый*, мы должны сформировать пул запросов – *безмозглик*, *безмозглец*, *безмозглыш*, *безмозгляк* и др. – и проверить каждое слово, используя поисковые системы интернета.

Таким образом, проблема создания исчерпывающей эмпирической базы для дальнейшего анализа может быть решена за счет конструирования как можно более полного перечня гипотетических (потенциальных) производных.

Подход, основанный на моделировании возможных словообразовательных структур, используется в работах Н. Б. Лебедевой (2010), Н. Д. Голева, И. П. Фаломкиной (2021), И. В. Ивлиевой (2020). Авторы вручную формируют дериваты на основании знаний о сочетаемости мотивирующих основ и аффиксов, участвующих в словопроизводстве. Однако если моделирование производных требует использования более сотни производящих основ и нескольких десятков словообразовательных формантов, ручной метод оказывается неэффективным. Таким образом, актуальность настоящего исследования определяется необходимостью создания автоматизированного метода, позволяющего работать с большим массивом данных.

При создании алгоритма мы опираемся на базовые принципы формального описания морфологических процессов: концепцию двухуровневой морфологии (Koskenniemi, 1983) и методы конечных автоматов (Beesley, Karttunen, 2003). В отличие от более сложных моделей, наш подход носит тривиальный характер: мы создаем максимально простой алгоритм, который позволяет автоматически соединять основы прилагательных с подходящими суффиксами и их алломорфами.

Для реализации предложенного подхода и достижения цели исследования необходимо решить следующие задачи:

- 1) составить перечень мотивирующих единиц и суффиксов, используемых для образования интересующих нас дериватов;
- 2) разработать алгоритм автоматического соединения основ прилагательных с соответствующими суффиксами;
- 3) отфильтровать данные с опорой на условия сочетаемости фонем.

Исследование предваряется критическим анализом научной литературы, позволяющей определить круг производящих, – в данном случае прилагательных, обозначающих психические характеристики человека, – а также установить набор словообразовательных формантов (суффиксов и их алломорфов), используемых для образования отадеквативных имен лиц. С помощью методов классификации и систематизации собранные прилагательные группируются по типу основы. Это позволяет заложить формальные основания для дальнейшего конструирования дериватов.

Автоматическое соединение основ прилагательных с соответствующими суффиксами осуществляется с использованием методов моделирования и алгоритмизации: строится теоретическая схема деривационного процесса и разрабатывается последовательность формальных операций, необходимых для автоматического порождения производных единиц.

Для фильтрации полученных слов используется метод сопоставительного анализа, заключающийся в сравнении финалей смоделированных дериватов с финалями зафиксированных в русском языке лексем. Оценивается, какие сочетания основы и суффикса уже существуют в языке, а какие нет.

Теоретическую базу исследования составляют труды Н. Б. Лебедевой (2010), М. Г. Шкуропацкой (2013), Н. Д. Голева, И. П. Фаломкиной (2020; 2021), И. В. Ивлиевой (2020), К. Koskenniemi (1983), К. R. Beesley, L. Karttunen (2003). Кроме того, учитываются сведения по морфотактике, извлеченные из РГ-80 и других источников.

Для отбора прилагательных используются два типа источников: психологические и лингвистические. Среди первых – работы, посвященные систематизации психических явлений (например, труды Б. М. Теплова (1953; 1961), Н. Д. Левитова (1964; 1969)). Лингвистические источники включают:

- Ефремова Т. Ф. Толковый словарь словообразовательных единиц русского языка. М.: Русский язык, 1996.
- Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений / под общ. ред. Н. Ю. Шведовой. М.: Азбуковник, 2003. Т. 3. Имена существительные с абстрактным значением. Бытие. Материя, пространство, время. Связи, отношения, зависимости. Духовный мир, состояние природы, человека. Общество
- Словарь-тезаурус прилагательных русского языка / под общ. ред. Л. Г. Бабенко. Екатеринбург: Изд-во Уральского ун-та, 2012.
- Тематический словарь русского языка около 25000 слов / под ред. В. В. Морковкина. М.: Дрофа, 2010.
- Убийко В. И. Концептосфера внутреннего мира человека в русском языке (комплексный функционально-когнитивный словарь). Уфа, 1998.
- Шрамм А. Н. Очерки по семантике качественных прилагательных (на материале современного русского языка). Л.: Изд-во Ленинградского ун-та, 1979.

Практическая значимость работы состоит в возможности применения ее результатов в автоматизированном отборе дериватов с использованием поисковых систем интернета. Кроме того, результаты исследования могут быть основой для разработки аналогичных автоматизированных инструментов. Алгоритм может быть адаптирован для конструирования производных единиц, образуемых по другим моделям с иным набором производящих основ и аффиксов.

Обсуждение и результаты

Лексико-семантическая группа прилагательных, обозначающих психические характеристики человека, включает 901 единицу. Их основы представлены 24 структурными типами, среди которых непроизводные основы (*безалаберный, бесшабашный, веселый, взбалмошный, хитрый* и т. д.), основы с суффиксом *-н-* (*беззаботный, дурашный, желчный* и т. д.), с суффиксом *-лив-* (*бережливый, боязливый, брюзгливый* и т. д.) и др. (подробнее см. Рисунок 1).

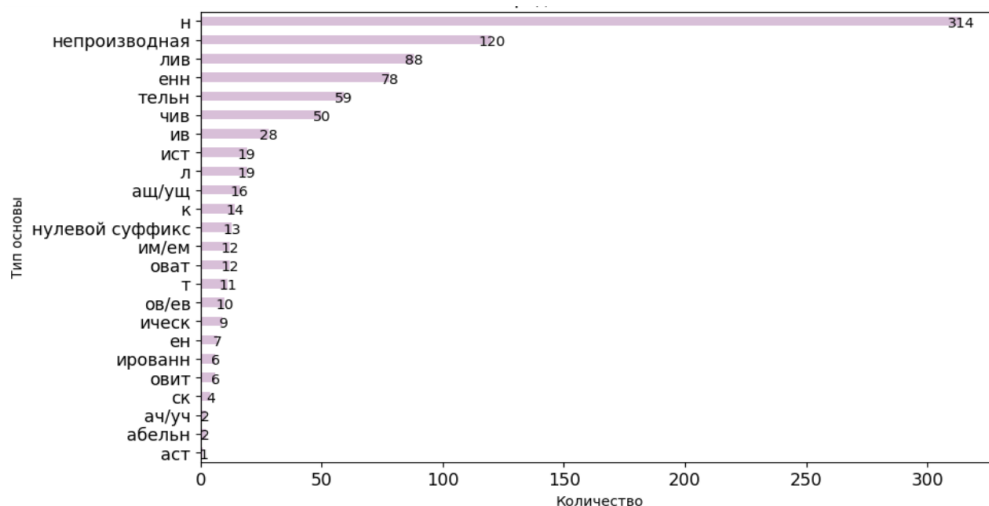


Рисунок 1. Распределение типов адъективных основ

Как отмечалось выше, существительные со значением «носитель признака (лицо)» образуются от прилагательных с помощью 28 суффиксов (РГ-80). Каждый из этих словообразовательных формантов имеет свои правила сочетаемости (немаловажно, что эти правила сформулированы в «докомпьютерную» эпоху, без учета больших массивов данных). Считается, например, что суффикс *-ак-* присоединяется преимущественно к непроизводной основе прилагательного, реже – к производной основе с суффиксами *-н-* и *-л-* (Ефремова, 1996; РГ-80): *чужой + -ак* → *чужак*, *смелый + -ак* → *смельчак* и т. п. Из этой формулировки следует, что формант не сочетается с основами на *-лив-*, *-чив-*, *-тельн-* и др., поскольку они не указаны в списке возможных. Однако, имея в виду цель и задачи настоящего исследования, мы считаем допустимым пренебречь отмеченными ограничениями и составить перечень гипотетических дериватов, образуемых от каждой основы с помощью любого из 28 имеющихся суффиксов. Это делается не вручную, а с помощью написанного на языке Python скрипта.

Разработанный алгоритм предназначен для генерации возможных дериватов со значением «носитель признака (лицо)» от прилагательных, обозначающих психические характеристики человека.

Данный процесс включает несколько этапов. На входе алгоритм принимает два списка: *adjective_list*, который содержит прилагательные, и *suffixes*, включающий суффиксы, участвующие в образовании отадъективных имен лиц.

Перед добавлением словообразовательных формантов необходимо выделить основу прилагательного. Для этого используется функция *remove_endings()*, удаляющая окончания производящих единиц *-ый*, *-ой*, *-ий*. В итоге мы получаем список адъективных основ: *агрессивн-*, *бессердечн-*, *хвастлив-*, *щедр-* и т. д.

После выделения основы, на следующем этапе, алгоритм перебирает все возможные суффиксы из списка и формирует новые слова путем конкатенации основы и словообразовательного форманта. Полученные дериваты записываются в новый список *new_words*: *агрессивнага*, *агрессивняга*, *агрессивнак*, *агрессивняк*, *агрессивнчак*, *агрессивнач*, *агрессивнец*, *агрессивнавец* и т. д. (см. Таблицу 1).

Таблица 1. Фрагмент датасета со сконструированными дериватами

	adjective	final_suffix	root	suffix	new_word
0	агрессивный	н	агрессивн	ага	агрессивнага
1	агрессивный	н	агрессивн	яга	агрессивняга
2	агрессивный	н	агрессивн	ак	агрессивнак
3	агрессивный	н	агрессивн	як	агрессивняк
4	агрессивный	н	агрессивн	чак	агрессивнчак

Кроме того, мы учитываем случаи, когда финаль производящей основы может отсекаться: *безумный* – *безумец*, *остроумный* – *остроумец*, *холодный* – *холодец* (РГ-80, с. 166); аналогичным образом «перед суффиксом *-ист-* отсекаются финали основ мотивирующих слов *-н-*, *-ск-* (в положении после согласной; исключение:

ручнист без усечения), *-ичн-* и *-ическ*» (РГ-80, с. 167). Предварительный анализ материала показывает, что некодифицированные дериваты с такими усечениями возможны. Мы можем образовать существительные *агрессивец, агрессивист* (наряду с *агрессивнец, агрессивнист*) и обнаружить примеры их употребления:

(1) Ну, то есть не у вас, а у **агрессивцев** при виде людей триггер срабатывает, они бегут по прямой и попадают в ловушку (Иевлев П. Сказки пустошей. Часть третья. «Свет в конце тоннеля». 2023. https://semiurg.ru/wasteland_tales/wasteland_tales3_04/).

(2) *Агрессивисты, подразделяющиеся на такие категории, как «паровые катки», или «танки» (грубые бесцеремонные люди, действующие напором и яростью)* <...> (Ключников С. Ю. Как вести себя с трудными людьми // Элитариум. Центр дополнительного образования. 20.05.2016. <https://www.elitarium.ru/ljudi-obshhenie-chuvstva-trudnosti-konflikt-svoystvo-chuvstvitelnost-psiologija/?ysclid=m9a4rgyhua180580026>).

С учетом сказанного в работе формируются дополнительные варианты адъективных основ с усечениями, которые соединяются с суффиксами *-ец-* и *-ист-*: *агрессивец, агрессивист, амбициозец, амбициозист* и т. п.

На следующем этапе осуществляется фильтрация сконструированных дериватов по определенным условиям (фонетическим особенностям). Такой порядок работы – сначала конструирование, затем очистка данных – обусловлен рядом причин.

Во-первых, этот подход обеспечивает гибкость и масштабируемость алгоритма. Генерация всех возможных сочетаний основ и суффиксов без предварительных ограничений позволяет легко модифицировать алгоритм в будущем и адаптировать его под другие исследовательские задачи. Если условия фильтрации изменятся или потребуются протестировать новые модели словообразования, не нужно будет заново пересматривать сам процесс генерации, достаточно будет скорректировать лишь правила отбора.

Во-вторых, разделение генерации и фильтрации делает алгоритм более структурированным. Один этап отвечает за создание полного перечня возможных форм, а второй – за их отбор, что делает код более читаемым и удобным для модификации. Такой подход соответствует общему принципу обработки данных: сначала формируется полный массив возможных комбинаций, а затем он очищается от неподходящих вариантов.

Наконец, предварительная генерация всех возможных форм дает возможность анализа промежуточных данных. Это позволяет выявить типичные ошибки, понять, какие комбинации морфем могут встречаться в языке, а также скорректировать правила фильтрации, добавляя новые ограничения или, наоборот, расширяя список допустимых форм.

Одним из таких методов фильтрации является проверка фонетической совместимости основы и суффикса. Например, мы исключаем те дериваты, где основа и суффикс имеют одинаковые фонемы на стыке, что нарушает фонетические правила. Это достигается с помощью функции `phonetically_valid()`, которая проверяет, не совпадают ли последние буквы основы и первые буквы словообразовательного форманта.

Такой подход позволяет исключить те варианты, которые могут быть трудными для произношения, например, формы вроде *чуткка* или *робкка*, где на стыке появляются две *к*. Только дериваты, проходящие этот фонетический фильтр, включаются в итоговый список.

Второе, на что следует обратить внимание при фильтрации, – это то, что не все сочетания звуков, которые могут появиться на стыке основы и форманта гипотетического производного, возможны в русском языке. Эту проблему можно решить с помощью Грамматического словаря русского языка А. А. Зализняка (www.gramdict.ru). Так, упомянутое выше слово *агрессивнага* следует исключить из датасета на том основании, что сочетание *-нага* имеется только в слове *донага*, где последняя *а* является ударной, что означает, что слов с сочетанием *-нага* в русском языке не встречается.

Иначе говоря, из созданного на первом шаге перечня гипотетических производных нужно удалить такие, в которых представлены фонетические сочетания, нехарактерные для русского языка в целом. Для этого необходимо решить, финаль какого объема подлежит проверке.

Предварительный анализ материала показывает, что оптимальным вариантом является ограничение в соответствии с длиной суффикса плюс одна буква основы, непосредственно ему предшествующая: *-н- + -ага, -н- + -ак, -н- + -авец* и т. п. Для этого сначала рассчитывается длина словообразовательного форманта (см. столбец `suffix_length` в Таблице 2), затем добавляется столбец `last_letters`, в котором для каждой строки извлекаются последние *n+1* символы из деривата, где *n* – длина суффикса.

Таблица 2. Фрагмент датасета, подготовленного для фильтрации

	adjective	final_suffix	root	suffix	new_word	suffix_length	last_letters
0	агрессивный	н	агрессивн	ага	агрессивнага	3	нага
1	агрессивный	н	агрессивн	яга	агрессивняга	3	няга
2	агрессивный	н	агрессивн	ак	агрессивнак	2	нак
3	агрессивный	н	агрессивн	як	агрессивняк	2	няк

Все финали из столбца `last_letters` автоматически извлекаются из датасета и помещаются в отдельный список – `letters_list`. Далее проводится автоматизированная фильтрация дериватов, основанная на результатах поиска сайта Грамматического словаря русского языка А. А. Зализняка – www.gramdict.ru. В процессе работы используется список сочетаний, к каждому из которых в соответствии с логикой работы сайта добавляется символ * в начале, что позволяет осуществить поиск всех слов, содержащих данные фрагменты.

Затем для каждой финали формируется поисковый запрос, который отправляется на сайт: [f*{x}' for x in letters_list], где x – это финаль, которую мы ищем. Итог запроса – все лексемы, найденные на страницах результатов. Для извлечения данных используется библиотека Selenium, которая позволяет получать динамически загружаемые страницы и извлекать необходимые элементы (лексемы) из HTML-кода.

```

<div class="body-content">
  <div>
    <div class="infinite-scroll-component " style="height: auto; overflow: hidden;">
      <div class="results-table short-results">
        <div class="result-entry">
          <div class="lemma copy-field" style="text-align: right;">донага</div>
          <div class="symbol copy-field">н</div>
          <div class="grammar copy-field"></div>
        </div>
      </div>
    </div>
  </div>
</div>

```

Рисунок 2. Фрагмент HTML-кода

Этот HTML-код представляет собой структуру, использующую классы и теги для отображения лексем на интернет-странице. Основной компонент, полезный для парсинга – <div class="lemma copy-field" style="text-align: right;">донага</div>. Это контейнер, в котором хранятся слова, найденные по запросу (см. Рисунок 2).

После завершения загрузки данных для каждого сочетания мы получаем информацию о том, существуют ли соответствующие лексемы на сайте (см. Таблицу 3).

Таблица 3. Фрагмент датасета с результатами парсинга

	last_letters	results
0	*нага	донага
1	*няга бедняга, скромняга, коняга, парняга, дворн...	
2	*овыш	Нет результатов
3	*нач	пернач, зурнач, мохнач
4	*навец	сквернавец
5	*обыш	Нет результатов

Далее мы сравниваем данные двух датасетов по столбцу “last_letters”, содержащему фрагменты дериватов. Для каждого значения “last_letters” из датасета с конструируемыми производными (df_1) выполняется поиск соответствующей строки в датасете с результатами парсинга (df_2). Если для этого фрагмента в df_2 найдены результаты, то в столбец “result_found” в df_1 добавляется значение “True”, если результатов нет – “False” (см. Таблицу 4).

Далее из датасета с производными удаляются все дериваты, которым в столбце “result_found” соответствует значение “False”. Это позволяет исключить нерелевантные данные и уточнить выборку производных.

Таблица 4. Фрагмент датасета, включающего производные и результаты парсинга

	adjective	final_suffix	root	suffix	new_word	suffix_length	last_letters	result_found
0	агрессивный	н	агрессивн	ага	агрессивнага	3	нага	True
1	агрессивный	н	агрессивн	яга	агрессивняга	3	няга	True
2	агрессивный	н	агрессивн	ак	агрессивнак	2	нак	True
3	агрессивный	н	агрессивн	як	агрессивняк	2	няк	True
4	агрессивный	н	агрессивн	чак	агрессивнчак	3	нчак	True

Одним из результатов проделанной работы является выявление того факта, что разноструктурные прилагательные по-разному участвуют в образовании гипотетических имен лиц.

Максимальное количество лексем (60 единиц) конструируется для непроизводных прилагательных: *храбрый, ммурый, хитрый, добрый, щедрый* и др.; минимальное (21 единица) – для прилагательных с суффиксами *-ащ-* / *-ящ-* и *-уц-* / *-юц-*: *мыслящий, завидующий, загребущий, заискивающий* и др.

Наиболее частотные суффиксы, используемые в моделировании производных лексем, – это *-к(а)*, *-ец-*, *-ак* (см. топ-15 суффиксов на Рисунке 3).

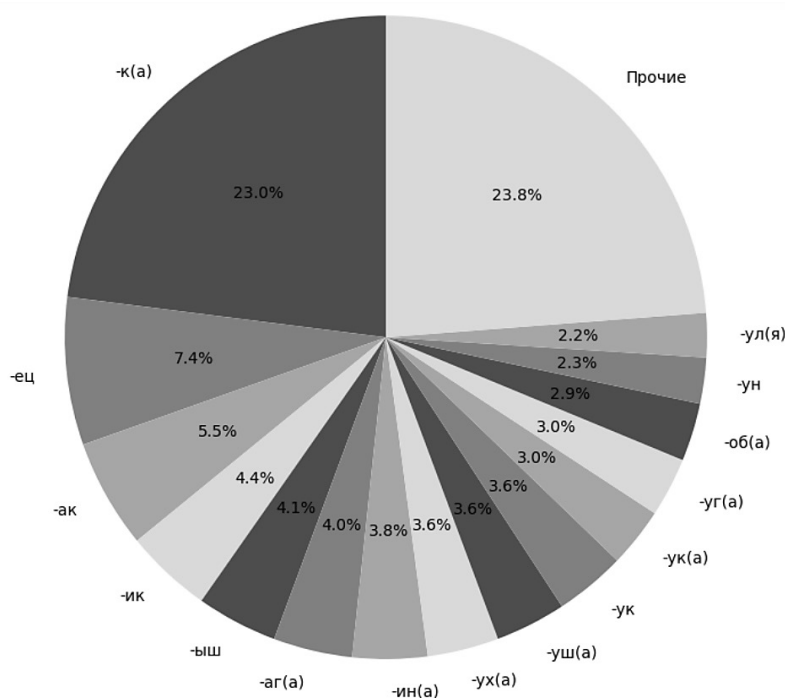


Рисунок 3. Топ-15 суффиксов, участвующих в образовании гипотетических производных

Таким образом, из 57681 изначально сконструированных производных после фильтрации остаются 41915 дериватов (около 73%), которые в дальнейшем можно использовать для формирования поисковых запросов в интернете.

Заключение

Настоящее исследование представляет собой важный шаг в разработке автоматизированных методов для конструирования гипотетических дериватов. В ходе работы создается алгоритм, позволяющий моделировать производные единицы с использованием современных технологий автоматизации. Для повышения точности результатов и гарантии их корректности в скрипт на языке Python интегрировано обращение к Грамматическому словарю русского языка А. А. Зализняка, что значительно повышает практическую ценность работы.

Предложенный метод открывает новые возможности для дальнейших разработок в области автоматизации поиска дериватов и создания аналогичных инструментов, адаптированных к другим словообразовательным моделям. Его применение может стать основой для разработки других программных средств анализа и конструирования различных типов производных единиц в русском языке.

Данный алгоритм повышает эффективность использования интернет-ресурсов для поиска производных в условиях дефицита лексикографической информации.

На следующем этапе оказывается возможным ответить на целый ряд других вопросов: используются ли данные производные в реальной речевой практике, какова их частотность, в каких функциональных стилях они востребованы и почему.

В перспективе это позволит уточнить представление о степени заполненности словообразовательных парадигм в определенный период развития языка и тем самым, с одной стороны, уточнить представления о словообразовательном потенциале единиц различных лексико-семантических групп, с другой – разработать методы описания деривационных процессов в их динамике.

Источники | References

1. Астапкина Е. С., Баркович А. А. Динамика русскоязычного интернет-дискурса: особенности словарной репрезентации лексики // *Russian Language Studies*. 2025. Т. 23. № 1.
2. Белякова Г. В. Деривационный потенциал словообразовательной категории суффиксальных локативных существительных в современном русском языке // *Известия Российского государственного педагогического университета им. А. И. Герцена*. 2007. № 8.
3. Васильева Е. В. Деривационный потенциал прилагательных на -ОВАТ- со значением подобия в сфере производства имен лиц // *Современная филология: теория и практика: материалы XIV международной научно-практической конференции*. М.: Научно-информационный издательский центр «Институт стратегического исследования», 2013.

4. Васильева Е. В. Формальные ограничения, препятствующие образованию отадективных имен лиц // Сибирский филологический журнал. 2016. № 1.
5. Голев Н. Д., Фаломкина И. П. О роли антропоцентрического фактора в реализации деривационного потенциала наименований копытных животных как мотивирующих лексем // Сибирский филологический журнал. 2021. № 3.
6. Голев Н. Д., Фаломкина И. П. Словообразовательный тип в аспекте его лексической реализации (к проблеме лакунарности словообразовательной системы русского языка) // Сибирский филологический журнал. 2020. № 2.
7. Евсева И. В., Чжао М. Лакуны и их типы в лексико-словообразовательном гнезде // Сибирский филологический форум. 2023. № 1 (22).
8. Замятина А. О. Деривационный потенциал колоративов и его реализация в русском языке // Актуальные проблемы лингвистики и литературоведения: сборник материалов X (XXIV) Международной научно-практической конференции молодых учёных. Томск: Национальный исследовательский Томский государственный университет, 2023. Вып. 24.
9. Земская Е. А. Современный русский язык. Словообразование. М.: Просвещение, 1973.
10. Ивлиева И. В. Особенности семантики лакун-модификаций. К идее создания словаря лакун (на материале глаголов звучания) // Интерактивная наука. 2020. № 6 (52).
11. Лебедева Н. Б. Экспериментальное исследование производного слова: соотношение выводимости как словообразовательного признака и узуальности как лексического признака // Новые явления в славянском словообразовании: система и функционирование: труды и мат. Междунар. науч. симпозиума «Славянские языки и культуры в современном мире». М.: Макс Пресс, 2010.
12. Левитов Н. Д. О психических состояниях человека. М.: Просвещение, 1964.
13. Левитов Н. Д. Психология характера. М.: Просвещение, 1969.
14. Ли Л. Бестиарий русского языка: пути заполнения лакун в обозначении самок и детенышей // Известия Южного федерального университета. Филологические науки. 2020. № 3.
15. Ло Ч. Деривационный потенциал наименований животных в современном русском языке // Вестник Минского государственного лингвистического университета. Серия 1. Филология. 2024. № 4 (131).
16. Лыткина О. И. Словообразовательные парадигмы непроективных антонимов в разных частях речи в современном русском языке: дисс. ... к. филол. н. М., 2001.
17. Николина Н. А., Рацибурская Л. В., Фатхутдинова В. Г. Новые явления в сфере деривационных формантов как отражение динамики словообразовательной системы русского языка // Вестник Волгоградского государственного университета. Серия 2. Языкознание. 2020. Т. 19. № 2.
18. Петрухина Е. В., Дедова О. В. Интернет как источник лингвистической информации (для изучения динамики русского словообразования) // Вестник Томского государственного университета. Филология. 2019. № 57.
19. Радбиль Т. Б., Ясаи Л., Палоши И. В. Лингвопрагматический потенциал активных процессов в русском неологическом словообразовании новейшего периода // Научный диалог. 2022. № 11 (1).
20. Русская грамматика: в 2 т. / гл. ред. Н. Ю. Шведова. М.: Наука, 1982. Т. 1.
21. Теплов Б. М. Проблемы индивидуальных различий. М.: Изд-во Акад. пед. наук РСФСР, 1961.
22. Теплов Б. М. Психология. М.: Учпедгиз, 1953.
23. Шкуропацкая М. Г. Методика исследования системной и эмпирической реализации словообразовательного типа в узуальной и потенциальной лексике // Мир науки, культуры, образования. 2013. № 6 (43).
24. Beesley K. R., Karttunen L. Finite-State Morphology: Xerox Tools and Techniques. Stanford: CSLI Publications, 2003.
25. Koskeniemi K. Two-level morphology: a general computational model for word-form recognition and production. Helsinki: University of Helsinki, Department of General Linguistics, 1983.

Информация об авторах | Author information

RU

Васильева Елена Викторовна¹
Ташлыкova Марина Борисовна², к. филол. н., доц.
^{1, 2} Иркутский государственный университет

EN

Elena Viktorovna Vasileva¹
Marina Borisovna Tashlykova², PhD
^{1, 2} Irkutsk State University

¹ elvavi2301@yandex.ru, ² taschlykoff@mail.ru

Информация о статье | About this article

Дата поступления рукописи (received): 23.04.2025; опубликовано online (published online): 29.05.2025.

Ключевые слова (keywords): словообразовательный потенциал прилагательных; отадективные имена лиц; гипотетическое производное; автоматическое конструирование; Python; derivational potential of adjectives; deverbal personal nouns; hypothetical derivative; automated construction.