



УДК 811

DOI 10.52575/2712-7451-2022-41-2-339-350

## Определение выраженности интернет-диалектов субъектов Российской Федерации

Миронова Д.М., Окунева Е.А., Распопин Д.И.

Юго-Западный государственный университет,  
Россия, 305040, г. Курск, ул. 50 лет Октября, 94  
E-mail: mir-lina@yandex.ru

**Аннотация.** Несмотря на повсеместное распространение школьного образования, стандартизацию языка и наличие литературной нормы, в лексике языка интернет-сетей в различных регионах присутствует достаточно различий, чтобы успешно определять региональную принадлежность текста. До сих пор в недостаточной степени раскрыт вопрос о существовании диалектов письменного языка сети. Целью данного исследования являлся поиск и описание диалектов языка интернета. В результате удалось установить существование различимых диалектных групп, доказать влияние географического фактора на формирование и распределение этих групп, а также сгруппировать субъекты Российской Федерации в группы по диалектному признаку. Новизна исследования заключается в применении методов машинного обучения и статистического анализа больших данных для выполнения поставленных целей. Результаты работы способствуют развитию представлений о языке сети как о неоднородной структуре, что может помочь развить более комплексный взгляд на функционирование русского языка и сформировать новые перспективы его изучения.

**Ключевые слова:** компьютерная лингвистика, машинное обучение, регионализмы, язык интернет-сети, диалект, карта

**Для цитирования:** Миронова Д.М., Окунева Е.А., Распопин Д.И. 2022. Определение выраженности интернет-диалектов субъектов Российской Федерации. Вопросы журналистики, педагогики и языкознания, 41(2): 339–350. DOI: 10.52575/2712-7451-2022-41-2-339-350

---

## Revealing of the Internet Dialects Manifestation of the Regions of the Russian Federation

Diana M. Mironova, Efrosinya A. Okuneva, Daniil I. Raspopin

Southwest State University,  
94 50 Let Oktyabrya St, Kursk 305040, Russian Federation  
E-mail: mir-lina@yandex.ru

**Abstract.** Despite the ubiquity of school education, the standardization of the language and the presence of a literary norm, there are enough differences in the vocabulary of the language of the Internet network in different regions to successfully determine the regional affiliation of the text. So far, the question of the existence of dialects of the written language of the web has been insufficiently disclosed. The purpose of this study was to find and describe the dialects of the Internet language. The result has been to establish the existence of distinguishable dialect groups, to prove the influence of the geographical factor on the formation and distribution of these groups, and to group the subjects of the Russian Federation into dialect groups. The novelty of the research lies in the application of methods of machine learning and statistical analysis of big data in order to achieve the goals set. The results of the work contribute to the development of ideas of the language network as a heterogeneous structure, which may help to develop a more comprehensive view of the functioning of the Russian language and form new perspectives for its study.



**Keywords:** computer linguistics, machine learning, regionalisms, Internet language, dialect, map

**For citation:** Mironova D.M., Okuneva E.A., Raspopin D.I. 2022. Revealing of the Internet Dialects Manifestation of the Regions of the Russian Federation. Issues in Journalism, Education, Linguistics, 41(2): 339–350 (in Russian). DOI: 10.52575/2712-7451-2022-41-2-339-350

## Введение

Данная статья является логическим продолжением работы «Применение технологий машинного обучения для выявления, анализа и классификации диалектов в социальных сетях» [Минакова и др., 2021]. В предыдущей статье была выдвинута и подтверждена гипотеза о существовании явных и выявляемых межрегиональных различий в лексике языка интернет-сети. Было установлено, что несмотря на повсеместное распространение школьного образования, стандартизацию языка и наличие литературной нормы, в письменной речи различных регионов присутствует достаточно различий, чтобы успешно определять региональную принадлежность текста при помощи методов машинного обучения. Распознаваемая разница не могла быть следствием опоры моделей машинного обучения на названия улиц, городов, популярных торговых центров и других уникальных для региона объектов, равно как не может объясняться случайностью, о чем свидетельствует достигаемая при определенных условиях 80 % точность классификации.

Для обозначения различимой разницы в интернет-языке регионов нами был использован термин «интернет-диалект».

Необходимо подчеркнуть важность географического аспекта интернет-диалектов. Первоначально их существование было следствием влияния особенностей устной разговорной речи региона (классических диалектов) на письменную речь жителей этого региона ввиду высокой скорости интернет-коммуникации, которая позволяла трансформировать мысль в письменный текст за предельно малое время [Лутовинова, 2008; Нелюбина и др., 2014; Мирзоева, 2019]. В таких условиях логично было бы ожидать схожести письменного высказывания с устным, что подразумевает также и возможное наличие в письменном высказывании диалектного следа региона автора.

В предыдущей работе было установлено, что одним из свойств интернет-диалекта является его выраженность.

Выраженность интернет-диалекта – это относительный показатель распознаваемости текстов определенного региона. Важным для нашего исследования свойством выраженности интернет-диалекта является ее математическая измеримость, поскольку это свойство было установлено из анализа статистических данных, полученных после применения моделей машинного обучения для классификации текстов разных регионов.

После проверки работоспособности технологий и методик на небольшой выборке из 8 городов было решено применить схожие методы в масштабах всей страны. Разумеется, такой шаг требовал получения значительного массива данных и его последующей обработки.

## Объект и методология исследования

Для определения статистических характеристик выборки текстов, конкретизации и обобщения результатов в исследовании применялся метод статистического анализа. Индивидуальные фрагменты выборки сопоставлялись друг с другом с помощью сравнительного анализа. Идентификация и классификация регионов по языковым особенностям осуществлялась с привлечением методов машинного обучения. Также на каждом этапе исследования использовался описательный метод, представленный приемами наблюдения, обобщения, классификации и интерпретации данных.

Для сбора необходимого массива текстов в предыдущей статье было предложено использовать самые популярные тематические сообщества регионов в социальной сети

«ВКонтакте». В настоящей работе методика была изменена с целью предельно эффективного извлечения из социальной сети регионального текстового материала. Теперь группы выбирались не по критерию соответствия той или иной тематике, а по популярности. Так, для каждого региона собирались комментарии от 4 до 8 самых популярных городских сообществ. Если по тем или иным причинам в области, крае, республике или автономном округе помимо крупного административного центра присутствовали и другие крупные города, то загружались также комментарии, полученные из этих городов. [Leng, 2012; Kulkarni et al., 2015; Булатова, Билялова, 2021].

Ввиду различного уровня интернет-активности субъектов Российской Федерации не во всех субъектах удалось собрать минимально необходимое для обучения моделей и тестирования количество текстов. По завершении загрузки было собрано более 68 миллионов текстов из 68 субъектов страны.

Алгоритм обработки текстовых данных также был изменен с целью предельного сближения текстов, с которыми будут работать модели машинного обучения, с естественными текстами из сети. Для этого было решено отказаться от постановки слов в начальную форму. Теперь из каждого текста исключались знаки препинания, устранялись лишние пробелы между словами, в начале и конце предложения, все символы переводились в нижний регистр.

Дополнительным этапом стал поиск и устранение дубликатов. Если одинаковые фразы встречались в нескольких регионах, они удалялись из всех текстов с целью уменьшения шумовой нагрузки на модель во время обучения.

### Проведение эксперимента

Первоначально была предпринята попытка в точности повторить эксперимент предыдущих исследований. Для этого при помощи средств из Python-библиотеки `scikit-learn` был построен классификатор «каждый-против-каждого», а в дальнейшем – и несколько других вариаций классификаторов, таких как классификатор линейных поддерживающих векторов, модель экстремального градиентного бустинга (как в классическом варианте, так и в `LGBM` реализации от Microsoft и в «`catboost`» реализации от компании Яндекс. Обоснование выбора классификатора и подробное описание его работы дано в нашей предыдущей статье [Минакова и др., 2021, с. 211].

Как и ожидалось, тенденция падения точности при увеличении количества классов классификации проявила себя, позволив достичь лишь средней точности в 1,5–3,5 %. Тем не менее даже при такой точности классификации среди регионов обнаружили экстремально явные по степени различимости регионы. Таковыми оказались Республика Северная Осетия, Республика Дагестан, Кабардино-Балкарская Республика, Рязанская область. В то время как точность классификаций прочих регионов колебалась около озвученного среднего значения, точность классификаций этих регионов составляла на менее 30–40 %, что как минимум в 20 раз превосходит среднее значение точности.

Далее была предпринята попытка искусственно свести задачу построения множественной классификации к получению бинарной. Для этого все регионы были распределены в две группы так, чтобы количество обучающих данных в двух группах было приблизительно одинаковым.

Однако уже после первых проверок стало ясно, что для предельно эффективного использования такого метода необходимо иметь представление о диалектных группах интернет-диалектов. Таким образом, мы были вынуждены скорректировать задачу. Теперь, в отличие от первого эксперимента, конечной целью было не построение модели, способной распознать принадлежность текста по всей стране, а построение классификаторов для каждой уникальной пары регионов с целью сбора данных о выраженности их интернет-диалектов и потенциальных способах группировки этих регионов в диалектные группы.

Для достижения этой цели вновь был применен классификатор «каждый-против-каждого», на этот раз запрограммированный вручную для получения дополнительных статистических данных. Базовой единицей этого классификатора стал мультиномиальный наивный Байесовский классификатор из Python-библиотеки scikit-learn [Manning et al., 2008; Кондрашева, Тинина, 2019; Буриева, 2021].

Алгоритм работы классификатора и сбора статистических данных заключался в следующем. Сначала выбирались тексты первого и второго регионов. Далее массивы разбивались на набор тренировочных и тестовых данных в пропорции 90:10 %. Тестовые данные – это уникальные данные, на которых модель не обучалась. Через них достигается корректность тестирования.

После обучения модели на вход ей подавались тестовые тексты первого региона. Процент текстов, принадлежность которых к первому региону была успешно определена, сохранялась в таблицу. То же самое происходило и для тестовых текстов второго региона.

После повторения процедуры на всех уникальных парах регионов была получена таблица, в которой содержатся данные об успешности классификации текстов каждого региона, когда альтернативой классификации был некий другой регион.

На основании этих данных для каждого региона был вычислен коэффициент выраженности его интернет-диалекта. Было решено определить коэффициент выраженности как сумму процентов успешных классификаций текстов этого региона во всех парах с участием этого региона.

Особенности, замеченные в первом эксперименте, отразились и здесь. Наибольшей выраженностью обладали Республика Северная Осетия, Республика Дагестан, Республика Кабардино-Балкария, Рязанская область.

Было решено графически изобразить выраженность интернет-диалектов регионов на карте (рис. 1).

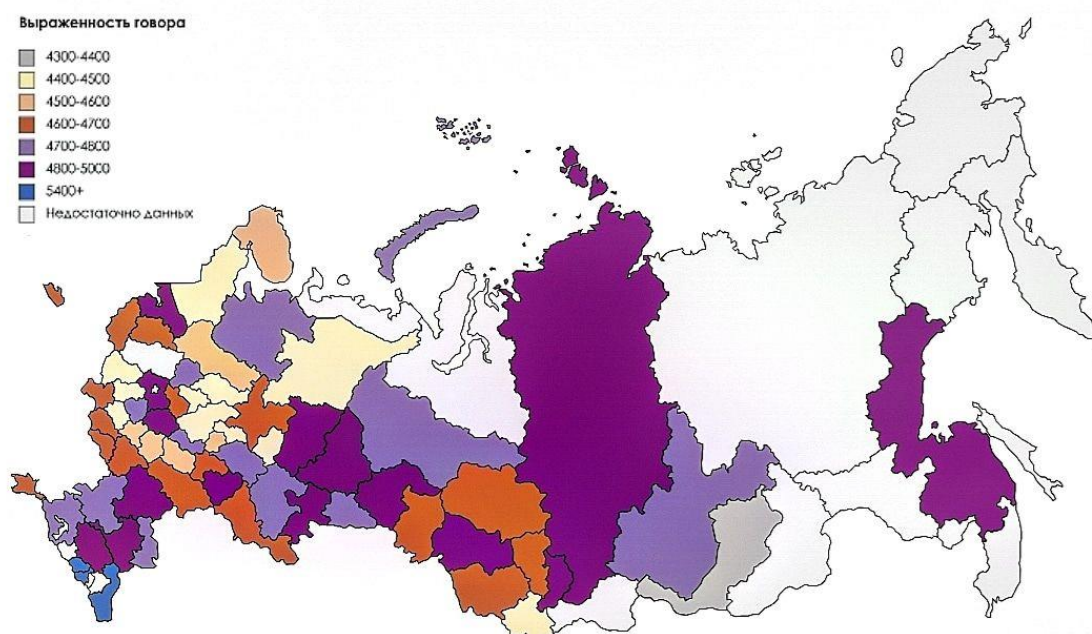


Рис. 1. Выраженность интернет-диалектов субъектов Российской Федерации  
Fig. 1. Differentiability of Russian regions' internet dialects

На карте находят выражение несколько интересных тенденций:

1. В центральной части России, в непосредственной близости от Московской области, выраженность диалектов достаточно слаба. Тенденция резко меняется при движении

на юг и на запад: Белгородская, Курская, Воронежская области – регионы-носители исторически подтвержденных диалектов [Филин, 1965; Лексический атлас..., 2014].

2. Высокой выраженностью обладают Псковская область и Новгородская область. Как и в случае с Курской, Белгородской и Воронежской областями, эти два региона обладают исторически подтвержденными диалектами [Филин, 1965; Лексический атлас..., 2014].

3. Выраженным и исторически подтвержденным диалектом обладает Ленинградская область [Филин, 1965; Лексический атлас..., 2014].

4. Высокой выраженностью обладают области проживания различных народностей России – Республика Татарстан, Республика Башкортостан, Пермская область. Подобная тенденция высокой выраженности наблюдается для большинства регионов, в которых русский язык используется наряду с иным национальным языком [Алексеев, 2013; Бадрях, Ширнэн, 2015; Голев, 2019].

5. Можно говорить о выраженности сибирских областей, Алтайского края, Новосибирской, Омской, Томской, Кемеровская область. Крупные сибирские города обладают достаточно выраженными интернет-диалектами.

6. Выраженным диалектом обладает Хабаровский и Приморский край. Расположение на дальнем востоке России позволило этим регионам развить достаточно выраженный интернет-диалект.

7. Регионы на юге России обладают весьма выраженными интернет-диалектами. Предельно яркими являются Кавказские диалекты, несколько уступают им в выраженности Крым, Краснодарский Край, Севастополь.

Данная карта отражает важнейшую для интернет-диалектов тенденцию – регионы с существующими, лингвистически выделенными территориальными диалектами обладают также ярко выраженными интернет-диалектами. Так как интернет-диалект – это следствие влияния диалекта устной речи на письменную, отсутствие такой корреляции на карте свидетельствовало бы о ее ошибочности.

Полученную информацию необходимо учесть для дальнейшей классификации интернет-диалектов в диалектные группы. Успешной оказывается та группировка, которая соответствует географическому и диалектологическому критериям.

Таким образом, логично предполагать наличие неких паттернов группировки интернет-диалектов, паттернов, напоминающих группировку диалектов устной речи. Тем не менее в той же степени ошибочно было бы ожидать полного повторения группировки устных диалектов, так как, в отличие от устного общения, в интернет-коммуникациях никак не учитывается расстояние между собеседниками, так что определенные речевые выражения, характерные для одного региона, вполне могут оказаться в другом.

Несмотря на то, что в одной диалектной группе могут оказаться регионы, не примыкающие друг к другу географически, у разрыва между регионами все-равно есть определенный предел. Язык отражает реалии и культурный контекст региона, поэтому совершенно ошибочно было бы соединять в одной группе южный курортный Краснодарский край и северный Ямало-Ненецкий автономный округ. Реалии двух регионов слишком сильно отличаются друг от друга.

Зафиксировав озвученные факторы и критерии, нами была предпринята попытка сгруппировать регионы по их интернет-диалектологической принадлежности.

### **Методология группировки**

Первоначально для группировки необходимо иметь сведения о том, какие регионы наиболее тесно связаны с точки зрения их диалектов.

Получить подобные сведения из имеющихся сугубо математических данных можно было лишь одним способом. Так как для каждого региона имеется ряд данных об успеш-

ных классификациях в парах с другими регионами, можно вычислить средний процент успешных классификаций и работать уже не в абсолютных, а в относительных величинах.

Для того чтобы убедиться, что среднее значение действительно адекватно отражает среднюю меру успешных классификаций региона, к вычислению среднего значения дополнительно вычислялась медиана ряда [Лемешко, 1997].

Демонстрация результатов применения методики на графике успешных классификаций Санкт-Петербурга представлена ниже (рис. 2).

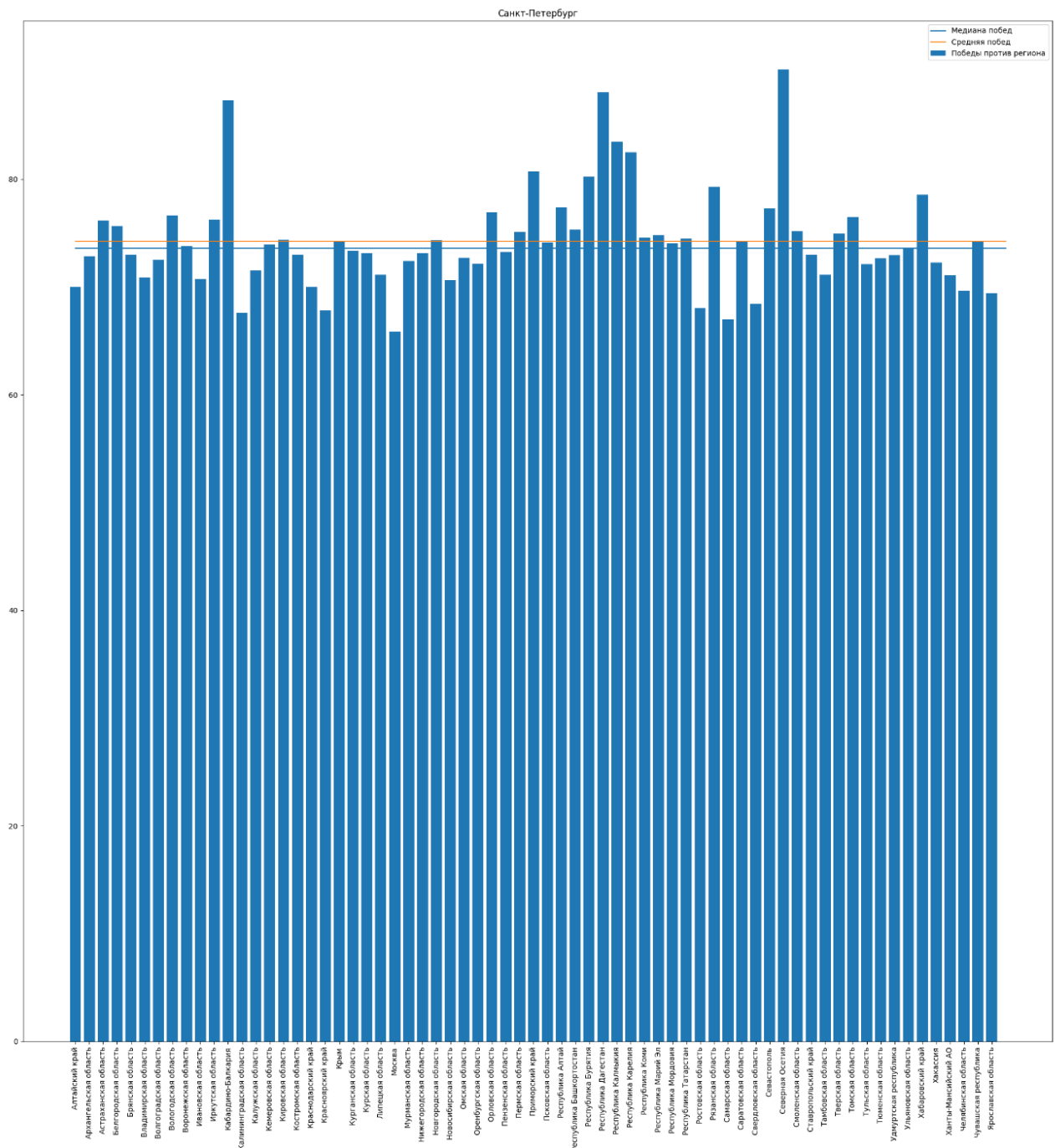


Рис. 2. Результаты применения методики на графике успешных классификаций Санкт-Петербурга  
 Fig. 2. Results of the method's application shown on the chart of successful Saint Petersburg dialect classification

Благодаря наличию двух черт на графике можно условно разбить регионы на две категории: те, что оказались под чертой, и те, что оказались над ней.

Регионы, оказавшиеся над чертой, – это регионы, в паре с которыми целевой регион классифицировал свои тексты успешнее, чем в среднем. Иными словами, по определенным причинам эти два региона контрастируют друг с другом. Данный фактор является весомой причиной не включать эти два региона в одну группу, и этот фактор тем весомее, чем сильнее выступает регион над линией среднего.

Далее наш интерес был сосредоточен на регионах, оказавшихся под чертой. Чем ниже опускается под чертой регион, тем хуже, по тем или иным причинам, удалось целевому региону классифицировать свои тексты в паре с этим регионом. Если на графике региона целевой регион так же находится под чертой, опускаясь при этом сопоставимо низко, то это позволяет утверждать существование некоей закономерности. Возможно, она свидетельствует о том, что письменный язык двух регионов похож между собой, что является поводом поместить их в одну интернет-диалектную группу.

Так как за графиками стоят фактические математические величины, то и тесноту связи интернет-диалектов регионов можно измерить математически.

На основании этого вывода была предпринята попытка произвести группировку регионов в интернет-диалектные группы путем вычисления тесноты связи между регионами и внутренней тесноты связи группы. Таким образом, существовало две функции, которых было необходимо максимизировать для признания группировки успешной.

Во время проверки результатов группировки был обнаружен главный недостаток такой методики. Такие факторы, как подавление регионом с сильно выраженным интернет-диалектом региона с более слабо выраженным интернет-диалектом и случайные корреляции нарушили ход группировки и привели к появлению в одной группе регионов, находящихся на большом удалении друг от друга. Так, Алтайский край по математической классификации находился в одной группе с Москвой, а Новосибирская область входила в состав диалектной группы, включавшей также Белгородскую, Курскую и Воронежскую области.

Тем не менее в математической классификации были обнаружены интересные результаты. Так, Крым и Севастополь оказались в одной диалектной группе, Республика Северная Осетия, Республика Дагестан и Кабардино-Балкарская Республика были изолированы от остальных в отдельную группу. При ближайшем рассмотрении даже в неподходящих по установленным критериям группах наблюдались следы сосуществования двух разных диалектных групп. Примером этого являлась группа, в которой Курская, Белгородская, Воронежская области сосуществовали с Новосибирской, Томской, Омской, Кемеровской областями. Очевидно, что две независимые друг от друга группы оказались вместе только посредством обнаруженной в процессе математической группировки случайной корреляции между несколькими городами.

Следующим шагом стала попытка группировки вручную, но с опорой на математическую метрику.

В результате детального анализа было выявлено существование особо сильных диалектных центров, которые, даже не будучи связанными с регионом и его группой, обнаруживали свое присутствие в списке наиболее подходящих кандидатов для группировки.

Опираясь на математическую метрику, а также принимая во внимание критерии успешной группировки, нами была составлена следующая карта групп интернет-диалектов России (рис. 3).

При составлении этой карты были сделаны следующие наблюдения:

1. Самым устойчивым по внутренней тесноте связи, географическому и диалектному критериям оказался регион, помеченный на карте розовым цветом и включающий в состав Ленинградскую, Псковскую, Новгородскую, Вологодскую, Архангельскую, Мурманскую, Кировская области.

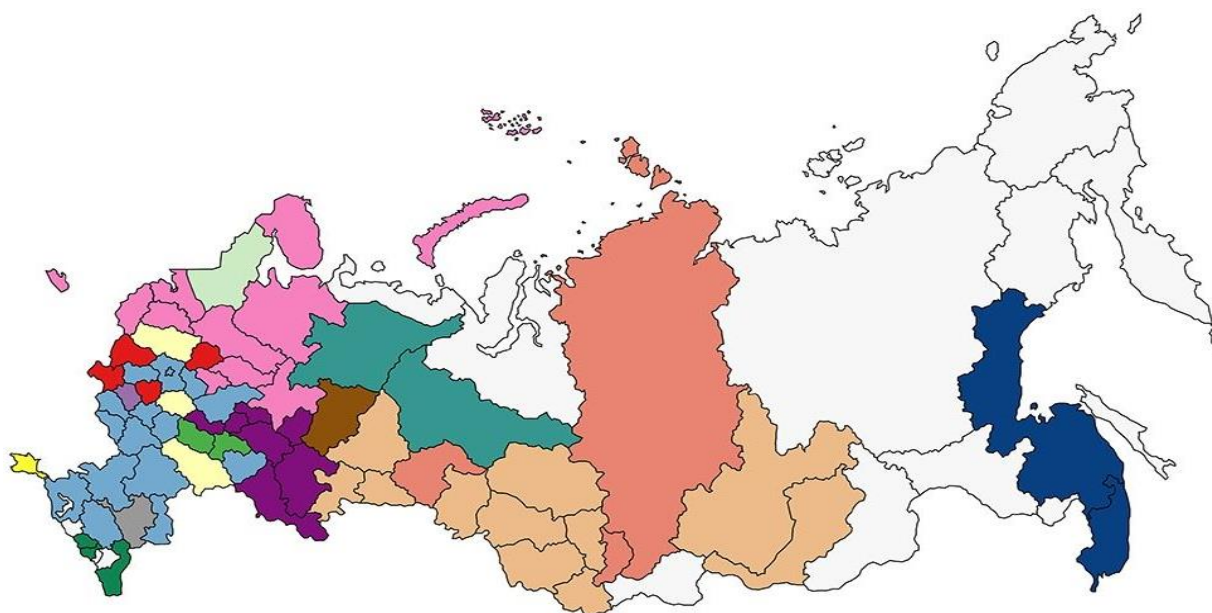


Рис. 3. Карта интернет-диалектов России  
Fig. 3. Russian internet dialect map

2. Пермская область (помеченная на карте коричневым цветом) оказалась в центре пересечения минимум трех диалектных групп. Такое местоположение было отражено в математических метриках Пермской области, она испытывала влияние и с севера, и с юга, и из Сибири.

3. Регионы, использующие русский совместно с другим языком, преимущественно помечены на карте фиолетовым цветом. Это говорит о том, что такие регионы, как Республика Татарстан, Республика Чувашия, Республика Марий Эл, Оренбургская область, Республика Башкортостан, достаточно близки друг другу своими интернет-диалектами.

4. Приморский и Хабаровский края географически изолировали себя в отдельную группу.

5. На карте просматриваются условные очертания Южно-Русского диалекта, который начинается из Курской области и идет на Юг до Ростовской области и Краснодарского края.

6. В регионах, находящихся на пересечении диалектных групп, почти всегда прослеживалось влияние и присутствие нескольких диалектных групп сразу. Это говорит о плавности перехода от одной диалектной группы к другой.

Несмотря на выявленные особенности, признать такую группировку удачной невозможно по нескольким причинам.

Во-первых, на карте присутствуют группы, не соответствующие ни одному из выделенных русскими лингвистами диалектов [Филин, 1965].

Во-вторых, для подтверждения успешности группировки необходимо отдельно провести лингвистический анализ каждой группы, проверить схожесть словарного запаса, региональных клише и в целом исследовать письменную речь членов группы.

В-третьих, на карте присутствуют несколько изолированных регионов (Карелия, Калмыкия, Орел), которые, по всей логике, должны быть распределены в близлежащие группы, однако этого не позволяет сделать отсутствие математической корреляции между этими регионами и близлежащими группами [Лексический атлас..., 2014; Петрухина, Дедова, 2019].



В-четвертых, математическая корреляция лишь частично подкрепляет группировку. В некоторых случаях части групп оказываются все еще нестабильны, что приводит к увеличению математической успешности группировки тогда, когда регионы распределяются в неверные, с точки зрения двух оставшихся критериев, группы.

Такие ограничения математического метода объясняются большим масштабом классификации, приводящим к появлению большого числа случайных корреляций. На данный момент методы машинного обучения хорошо зарекомендовали себя в группировке текстов в небольшое число хорошо различимых классов [Классификация текстов..., 2017; Хазеев, 2019]. В свою очередь, локальные особенности языка зачастую выражаются на уровнях фонетики и словообразования, в лексике, морфологии и сочетаемости слова, поэтому изучение этих особенностей с помощью компьютерных технологий требует внимания к единицам, меньшим, чем тексты.

### Заключение

Выводы, полученные в работе, подтверждают гипотезы, выдвинутые нами в предыдущей статье. Удалось установить факт влияния особенностей устной речи на письменную, что позволяет оспорить популярную точку зрения о размытии языковых особенностей в интернет-пространстве [Игнатенко, 2009; Иванова, 2011].

Способы, которыми в языке достигается выражение особенностей региональной устной речи необходимо изучить подробнее, это означает, что выводы, полученные в работе, открывают целый ряд вопросов для изучения лингвистами.

Технологии, примененные в процессе исследования, доказывают возможность создавать практически ничем не ограниченные по объему выборки текстов из сети для статистического и других исследований русского языка.

Факт достижения точности классификации до 80 % при определенных условиях позволяет с уверенностью судить о наличии статистически значимых закономерностей в исследованных нами данных. Увеличение размеров выборки и числа регионов позволяет установить локальный характер расхождений текстов из разных регионов, что означает, что способность успешно классифицировать регион происхождения текста напрямую коррелирует с наличием в тексте диалектных или редких слов.

Для дальнейшего уточнения полученных данных существует возможность включить в выборку текстов не только комментарии в социальных сетях. В таком случае размер выборки вырастет с использованных нами 68 миллионов текстов до нескольких сотен миллионов, что обеспечит еще более высокую степень статистической достоверности полученных результатов.

Полученные выводы позволяют судить о неоднородности языка сети.

Немаловажным выводом является так же то, что помимо современного влияния письменного языка сети на устную речь в той же степени верен и факт обратного влияния. Помимо этого, выводы, полученные в работе, позволяют утверждать, что наличие единой литературной нормы и массовое обучение правописанию отнюдь не искоренили диалекты из письменной обиходной речи обычных людей.

### Список источников

- Классификация текстов с помощью сверточных нейронных сетей. 2017. Хабр. URL: <https://habr.com/ru/post/332078/> (дата обращения: 10 сентября 2021).
- Филин Ф.П. 1965. Словарь русских народных говоров. Вып. 1: А. М., Л., Наука, 303 с. URL: <https://iling.spb.ru/dictionaries/srng/01.pdf> (дата обращения: 10 февраля 2021).
- Multinomial Naïve Bayes classifier documentation page, scikit-learn python API. 2007-2022. Scikit-learn. org. Available at: [https://scikit-learn.org/stable/modules/generated/sklearn.naive\\_bayes.MultinomialNB.html](https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html) (accessed: 25 June 2021).



## Список литературы

- Алексеев М.Е. 2013. Языки народов России и постсоветского пространства: проблемы изучения и перспективы развития. Родной язык, 1: 8–17.
- Бадрах А., Ширнэн Ц. 2015. К вопросу о функционировании русского языка в интернет среде. Вестник Новосибирского государственного педагогического университета, 5 (5): 134–141. URL: <http://sciforedu.ru/article/1573> (дата обращения: 8 февраля 2021). DOI: <http://dx.doi.org/10.15293/2226-3365.1505.12>.
- Булатова И.З., Билялова А.А. 2021. Особенности интернет-коммуникации в системе национального русского языка. Международный научно-исследовательский журнал, 3-3(105): 107–110. URL: <https://research-journal.org/wp-content/uploads/2021/03/3-105-3.pdf> (дата обращения: 9 августа 2021). DOI: 10.23670/IRJ.2021.105.3.080.
- Буриева М. 2021. О симплификации языка в интернет-пространстве. Вестник Московского государственного лингвистического университета. Гуманитарные науки, 5(847): 22–33. URL: [http://www.vestnik-mslu.ru/Vest/5\\_847\\_H.pdf](http://www.vestnik-mslu.ru/Vest/5_847_H.pdf) (дата обращения: 19 августа 2021). DOI: 10.52070/2542-2197\_2021\_5\_847\_22
- Голев Н.Д. 2019. Дискурсивный словарь диалектной лексики новейшего времени (на материалах рунета): инновационный лексикографический проект. Вопросы лексикографии, 16: 113–137. DOI: 10.17223/22274200/16/7.
- Иванова Т.С. 2011. Речевое поведение Интернет-общения. Вестник Адыгейского государственного университета. Серия 2: Филология и искусствоведение, 3: 132–136.
- Игнатенко И.И. 2009. Языковые и коммуникативные процессы в эпоху глобализации. Наука и школа, 3: 24–27.
- Кондрашева Е.В., Тинина А.О. 2019. Новая лексика языка интернета: способы образования, причины появления. Эпоха науки, 18: 134–139. URL: [http://eraofscience.com/EofS/Vypyski2019/18\\_iyun2019/30.pdf](http://eraofscience.com/EofS/Vypyski2019/18_iyun2019/30.pdf) (дата обращения: 19 августа 2021). DOI: 10.24411/2409-3203-2018-11830.
- Лексический атлас русских народных говоров (материалы и исследования). 2014. Под ред. А.С. Герда. Санкт-Петербург, Издательство Нестор-История, 808 с. URL: <https://elibrary.ru/item.asp?id=23245144> (дата обращения: 10 февраля 2021).
- Лемешко Б.Ю. 1997. Робастное оценивание параметров распределений. Заводская лаборатория. Диагностика материалов, 63(5): 43–49. URL: [https://ami.nstu.ru/~headrd/seminar/publik\\_html/Z\\_lab\\_1.htm](https://ami.nstu.ru/~headrd/seminar/publik_html/Z_lab_1.htm) (accessed: 19 августа 2021).
- Лутовинова О.В. 2008. Интернет как новая ‘устно-письменная’ система коммуникации. Известия Российского государственного педагогического университета имени А.И. Герцена, 11(71): 58–65.
- Минакова И.В., Распопин Д.И., Окунева Е.А. 2021. Исследование диалектов в письменной речи социальных сетей. Modern Humanities Success, 3: 211–221. URL: <https://mhs-journal.ru/wp-content/uploads/2021/03/mhs-3.pdf> (дата обращения: 10 февраля 2021).
- Мирзоева Л.Ю. 2019. Словообразование в интернет-сети и проблемы экологии языка. Экология языка и коммуникативная практика, 2(17): 26–34. URL: [http://ecoling.sfu-kras.ru/wp-content/uploads/2019/08/4.-Mirzoeva-L.Yu\\_.pdf](http://ecoling.sfu-kras.ru/wp-content/uploads/2019/08/4.-Mirzoeva-L.Yu_.pdf) (дата обращения: 4 сентября 2021). DOI: 10.17516/2311-3499-055.
- Нелюбина Е.А., Амитрова М.В., Гусарова Ю.В. 2014. Языковые особенности интернет-дискурса. Балтийский гуманитарный журнал, 4(9): 21–23.
- Петрухина Е.В., Дедова О.В. 2019. Интернет как источник лингвистической информации (для изучения динамики русского словообразования) Вестник Томского государственного университета. Филология, 57: 137–159. DOI: 10.17223/19986645/57/8.
- Хазеев Д.Р. 2019. Приложение нейронных сетей к определению стиля текста. Новые информационные технологии в автоматизированных системах, 22. Вып.: Новые информационные технологии в автоматизированных системах. Материалы двадцать второго научно-практического семинара, г. Москва, 19 апреля 2019 г. М., ВШЭ: 121–124.
- Kulkarni V., Perozzi B., Skiena S. 2015. Freshman or Fresher? Quantifying the Geographic Variation of Internet Language. Available at: [arXiv:1510.06786 \[cs.CL\]](https://arxiv.org/abs/1510.06786) (accessed: 25 June 2021). DOI: <https://doi.org/10.48550/arXiv.1510.06786>

- Leng Z. 2012. A study of the features of Internet language. In: 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang, China, 21-23 April 2012. IEEE Xplore: 97-100. Available at: <https://ieeexplore.ieee.org/document/6201466> (accessed: 21 June 2021). DOI: 10.1109/CECNet.2012.6201466.
- Manning C.D., Raghavan P., Schuetze H. 2008. Introduction to Information Retrieval. Cambridge University Press, 506 p. DOI: <https://doi.org/10.1017/CBO9780511809071>.

## References

- Alekseev M.E. 2013. Yazyki narodov Rossii i postsovetskogo prostranstva: problemy izucheniya i perspektivy razvitiya [Languages of the peoples of Russia and the post-Soviet space: problems of study and development prospects]. *Rodnoy yazyk*, 1: 8–17.
- Badrakh A., Shirnen Ts. 2015. The study of use of russian language in internet environment. *Novosibirsk State Pedagogical University Bulletin*, 5(5): 134–141. Available at: <http://sciforedu.ru/article/1573> (accessed: 8 февраля 2021) (in Russian). DOI: <http://dx.doi.org/10.15293/2226-3365.1505.12>.
- Bulatova I.Z., Bilyalova A.A. 2021. Aspects of internet communication in the system of the national russian language. *International research journal*, 3-3(105): 107–110. Available at: <https://research-journal.org/wp-content/uploads/2021/03/3-105-3.pdf> (accessed: 9 august 2021) (in Russian). DOI: 10.23670/IRJ.2021.105.3.080.
- Burieva M. 2021. On Language Simplification on the Internet. *Vestnik of Moscow State Linguistic University. Humanities*, 5(847): 22–33. Available at: [http://www.vestnik-mslu.ru/Vest/5\\_847\\_H.pdf](http://www.vestnik-mslu.ru/Vest/5_847_H.pdf) (accessed: 19 august 2021) (in Russian). DOI: 10.52070/2542-2197\_2021\_5\_847\_22.
- Golev N.D. 2019. A modern dialect vocabulary discourse dictionary (based on runet materials): an innovative lexicographic project. *Russian Journal of Lexicography*, 16: 113–137 (in Russian). DOI: 10.17223/22274200/16/7.
- Ivanova T.S. language behavior of internet communication. *Bulletin of the Adygea State University. Series 2: Philology and Art History*, 3: 132–136 (in Russian).
- Ignatenko I.I. 2009. Linguistic and communicative processes in global age. *Science and School*, 3: 24–27 (in Russian).
- Khazeev D.R. 2019. Prilozhenie neyronnykh setey k opredeleniyu stilya teksta [Application of Neural Networks to Text Style Determination]. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh*, 22. Is.: *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh* [New information technologies in automated systems]. *Materialy dvadtsat' vtorogo nauchno-prakticheskogo seminara, g. Moskva, 19 aprelya 2019 g. M., Publ. VShE*: 121–124.
- Kondrasheva E.V., Tinina A.O. 2019. Verbal aggression in the language of media: the varieties and causes. *Epokha nauki*, 18: 134–139. Available at: [http://eraofscience.com/EofS/Vypyski2019/18\\_iyun2019/30.pdf](http://eraofscience.com/EofS/Vypyski2019/18_iyun2019/30.pdf) (accessed: 19.07.2021). DOI: 10.24411/2409-3203-2018-11830.
- Kulkarni V., Perozzi B., Skiena S. 2015. Freshman or Fresher? Quantifying the Geographic Variation of Internet Language. Available at: [arXiv:1510.06786 \[cs.CL\]](https://arxiv.org/abs/1510.06786) (accessed: 25 June 2021). DOI: <https://doi.org/10.48550/arXiv.1510.06786>.
- Leksicheskiy atlas russkikh narodnykh govorov (materialy i issledovaniya) [Lexical atlas of Russian folk dialects (materials and research)]. 2014. Ed. A.S. Gerd. Sankt-Peterburg, Publ. Izdatel'stvo Nestor-Istoriya, 808 p. Available at: <https://elibrary.ru/item.asp?Id=23245144> (accessed: 10 fevralya 2021).
- Lemeshko B.Yu. 1997. Robust methods for the estimation and rejection of anomalous measurements. *Industrial Laboratory. Diagnostics of Materials*, 63(5): 43–49. Available at: [https://ami.nstu.ru/~headrd/seminar/publik\\_html/Z\\_lab\\_1.htm](https://ami.nstu.ru/~headrd/seminar/publik_html/Z_lab_1.htm) (accessed: 19.07.2021) (in Russian).
- Leng Z. 2012. A study of the features of Internet language. In: 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang, China, 21-23 April 2012. IEEE Xplore: 97-100. Available at: <https://ieeexplore.ieee.org/document/6201466> (accessed: 21 June 2021). DOI: 10.1109/CECNet.2012.6201466.
- Lutovinova O.V. 2008. Internet as a new “oral-written” communication system. *Scientific journal of Herzen University*, 11(71): 58–65 (in Russian).
- Manning C.D., Raghavan P., Schuetze H. 2008. Introduction to Information Retrieval. Cambridge University Press, 506 p. DOI: <https://doi.org/10.1017/CBO9780511809071>.



- Minakova I.V., Raspopin D.I., Okuneva E.A. 2021. Research of dialects in the written speech of social networks. *Modern Humanities Success*, 3: 211–221. Available at: <https://mhs-journal.ru/wp-content/uploads/2021/03/mhs-3.pdf> (accessed: 10 fevralya 2021) (in Russian).
- Mirzoeva L.Yu. 2019. Word creation in internet network and language ecology. *Ecology of Language and Communicative Practice*, 2(17): 26–34. Available at: [http://ecoling.sfu-kras.ru/wp-content/uploads/2019/08/4.-Mirzoeva-L.Yu\\_.pdf](http://ecoling.sfu-kras.ru/wp-content/uploads/2019/08/4.-Mirzoeva-L.Yu_.pdf) (accessed: 4 сентября 2021). DOI: 10.17516/2311-3499-055.
- Nelyubina E.A., Amitrova M.V., Gusarova Yu.V. 2014. Language features of internet discourse. *Baltic humanitarian journal*, 4 (9): 21–234 (in Russian).
- Petrukhina E.V., Dedova O.V. 2019. The internet as a source of linguistic information (for studying the dynamics of russian word formation). *Tomsk State University Journal of Philology*, 57: 137–159 (in Russian). DOI: 10.17223/19986645/57/8.

**Конфликт интересов:** о потенциальном конфликте интересов не сообщалось.

**Conflict of interest:** no potential conflict of interest related to this article was reported.

Поступила в редакцию 30.12. 2021

Received December 30, 2021

Поступила после рецензирования 15.02.2022

Revised February 15, 2022

Принята к публикации 10.06.2022

Accepted June 10, 2022

#### ИНФОРМАЦИЯ ОБ АВТОРАХ

#### INFORMATION ABOUT THE AUTHORS

**Миронова Диана Михайловна**, доцент кафедры теоретической и прикладной лингвистики, Юго-Западный государственный университет, г. Курск, Россия

**Diana M. Mironova**, Associate Professor at the Department of Theoretical and Applied Linguistics, Southwest State University, Kursk, Russia

**Распопин Даниил Игоревич**, студент кафедры таможенного дела и мировой экономики, Юго-Западный государственный университет, г. Курск, Россия

**Daniil I. Raspopin**, student of the cathedra of customs and global economics of South-West State University, Kursk, Russia

**Окунева Ефросинья Андреевна**, студент кафедры теоретической и прикладной лингвистики, Юго-Западный государственный университет, г. Курск, Россия

**Efrosinya A. Okuneva**, student of the cathedra of theoretical and applied linguistics of South-West State University, Kursk, Russia