

УДК-328(470+571)

DOI: 10.17072/2218-1067-2020-4-138-152

«ДЕПУТАТСКАЯ КАРТОГРАФИЯ» НА МАТЕРИАЛЕ СТЕНОГРАММ ЗАСЕДАНИЙ ГД РФ

Константин Белоусов

Пермский государственный национальный исследовательский университет

Константин Рябинин

Пермский государственный национальный исследовательский университет

Иван Лабутин

Пермский государственный национальный исследовательский университет

Константин Сулимов

Пермский государственный национальный исследовательский университет

Аннотация

Публичные выступления депутатов Государственной думы РФ представляют продуктивный эмпирический материал, требующий применения междисциплинарного подхода к его анализу. Общей целью исследования, в рамках которого выполнена данная статья, является определение вариативности и пространства эквилибриума образа будущего России в представлениях депутатского корпуса. Эмпирической базой являются стенограммы заседаний ГД РФ: весь массив данных с 1994 г. по середину 2020 г. включает 324 тыс. реплик (27 млн словоупотреблений) от 2 773 депутатов и других лиц. В данной статье представлен общий дизайн исследования дискурса депутатов ГД с использованием информационной системы "Семограф", библиотеки SlovNet для анализа текстов на естественном языке на основе методов глубинного обучения, платформы визуальной аналитики SciVi, модуля визуализации карт на основе библиотеки Leaflet, геокодирования объектов на основе картографического провайдера OpenStreetMap и других. Применение подхода показано на примере установления институциональной обусловленности символического депутатского картирования регионов Приволжского федерального округа электоральной связью депутата с регионом («одномандатники» vs «списочники»).

Ключевые слова: депутаты; образ страны; Государственная дума; регионы; машинная обработка; данные; визуализация.

Публичные выступления депутатов в парламентах являются одним из самых заметных видов их деятельности, но при этом остается относительно малоизученным (Proksch et al., 2012), даже несмотря на значительные успехи последних десятилетий в развитии машинной обработки больших объемов текстовой информации. Их размер демонстрируют собранные нами данные по стенограммам выступлений депутатов Государственной думы Российской Федерации¹. Весь массив данных с 1994 г. по середину 2020 г. включает 324 тысячи реплик (27 млн словоупотреблений) от 2 773 депутатов и других лиц. Очевидно, что такие объемы могут быть обработаны только машинным способом, но при этом возникает проблема издержек на «осмысленную» обработку, то есть требуется выбор оптимальной схемы соотношения машинной обработки и экспертного участия, без которого все равно не обойтись (Quinn et al., 2010). Кроме того, требуется решать «проблему» языка или точнее – словоупотребления: многозначность терминов, неполная формализуемость, стилевое и жанровое своеобразие использования, наконец, различия между естественными языками, что препятствует выработке универсальных решений, хотя подходы существуют. Наконец, надо решать содержательные и методологические проблемы, связанные с ответом на вопрос – а что, собственно, мы хотим и можем узнать с помощью всего лишь, так или иначе, подсчета слов? Например, относительно качественно можно выделять тематические линии обсуждений, дискуссий, политической повестки. Труднее дает-

© Белоусов К. И., Лабутин И. А., Рябинин К. В., Сулимов К. А., 2020

¹ База данных стенограмм заседаний Государственной думы. [электронный ресурс]. URL: <http://transcript.duma.gov.ru/> (дата обращения: 02.10.2020).

ся «превратить подсчет слов в оценки «выраженных предпочтений» или «заявленных позиций» (Lauderdale et al., 2016: 392). Трудно вычленил собственно споры и различия в позициях, которые, конечно, являются в некотором смысле сутью политики.

Для того, чтобы решать эти проблемы, необходимо принимать в расчет институциональные условия, в которых разворачивается деятельность депутатов, то есть, в том числе, характер политического режима, особенности избирательной системы, партийных и фракционных механизмов, требования процедурно-регламентных правил самой легислатуры и другие; то есть, при работе со стенограммами пленарных заседаний Государственной думы необходимо учитывать ее место в рамках существующего политического режима и роль в принятии политических решений. Современные исследования показывают, сама по себе недемократичность режима не ведет с необходимостью к марионетности парламента, он может принимать реальное участие в формулировании политики при условии наличия среди элит разных представлений о возможных целях и исходах политики (Gandhi et al., 2020). Применительно к Государственной думе, исследования показывают, что она действует по известной модели “rubber stamp” в случаях, когда речь идет о политически важных инициативах, но также может быть «полем битвы элит» между различными группами интересов (Noble and Schulmann, 2018). Общим условием такой возможности является то, что авторитарная централизация имеет свои пределы, то есть не является в случае России полной и всеобъемлющей (Petrov and Nazrulaeva, 2018).

Из этого следует, что публичная речь в стенах Государственной думы может иметь какой-то смысл. Более того, несмотря на известную фразу одного бывшего председателя Думы, дискуссии в ней все-таки бывают, хотя они мало напоминают то, что принято называть дебатами. Однако какой именно смысл? В некотором стандартном представлении выделяются две возможные причины, по которым законодательные органы уделяют время дебатам или хотя бы дискуссиям (Proksch et al., 2012). Во-первых, они могут быть значимы по причине потенциального влияния на результаты политики – в этой логике депутаты пытаются убедить коллег в превосходстве своей позиции. Во-вторых, дебаты могут функционировать как публичная площадка, на которой депутаты разъясняют результаты голосования и политические позиции своим и другим партиям, и избирателям. В случае Государственной думы мы не можем априорно принять эти позиции – и в отношении значимости аргументации для принятия решений, тем более, что эффективность такой аргументации весьма сомнительна и в других парламентах (см.: Ibid.) и даже в отношении публичного представления политических позиций. Точнее говоря, последнее явно имеет место, но чьи собственно позиции представляются и для кого? Где грань между позицией, ориентированной на внешнего потребителя – публику, избирателя, и позицией, являющейся выражением лояльности режиму?

Но в чем точно можно быть уверенным, так это в том, что в публичной речи депутатов проявляется их представление о мире вокруг них. Теоретико-методологической основой нашего подхода являются концепции политического сообщества и политической идентичности. Политическое сообщество основывается на коренящихся в общей культуре политических идентичностях и интернализованных членами сообщества коллективных представлениях – они и проявляются в речах депутатов. То есть речь идет о макрополитической идентичности, которая базируется на общих, разделяемых членами политического сообщества представлениях (своего рода «картине мира»). Один из ключевых компонентов политического сообщества, то есть представлений о том, кто есть «мы» – коллективно разделяемое представление о стране в целом – образ страны: «знаковая модель, опосредующая представления о национальной общности и ее членах через доступные обыденному сознанию понятия и суждения» (Семененко, 2007: 64).

Одна из задач нашего исследования – на материале всего корпуса стенограмм Государственной думы определить образ страны в представлениях депутатского корпуса. Этот образ требует представлений не только об идеологических смыслах, но и о пространстве, в котором он существует – о границах, структуре, составных элементах и связях пространства. Соответственно, необходимо выявить картину, конкретизирующую видение пространства и представлений о частях, из которых состоит страна, создать образно-символический ряд. Но при этом мы исходим из того, что выступления депутатов обусловлены, и конечный замысел исследования состоит в определении комбинации политических, институциональных и иных факторов, влияющих как на содержание образа, так и на степень согласия по его поводу.

В данной статье сначала будут представлены наши основные методологические и технологические подходы к работе со всем корпусом текстов стенограмм пленарных заседаний Государственной думы. Затем мы покажем его работоспособность на примере одного локального сюжета – «ре-

гионального картирования» депутатов от регионов Приволжского федерального округа (далее – ПФО).

Как было сказано выше, имеют значение не только идеологический смысл, связанный с образом страны в публичных выступлениях депутатов, но и их представление о структуре пространства, в котором этот смысл только и становится смыслом, то есть получает то или иное содержание. Важный сегмент пространственной структуры страны составляют ее регионы. В понятие региона можно вкладывать разное содержание, но в данном случае мы ограничимся политико-административными регионами, то есть субъектами РФ. Депутаты в своих выступлениях по самым разным поводам регулярно упоминают разные регионы (их названия или иные «сущности», однозначно связываемые с тем или иным субъектом РФ). При этом можно уверенно ожидать, что частота упоминания разных регионов будет различаться в целом и у отдельных депутатов. Представляется, что это имеет важный символический смысл. Речь, разумеется, не идет о том, что если конкретный депутат или вся совокупность не упоминают тот или иной регион, то он для них не существует. Ясно, что символическая реальность страны может и почти наверняка отличается от ее физической реальности. Символическое существование и значимость региона определяются его проявленностью в символическом пространстве и интенсивностью этой проявленности. В операциональном отношении это может быть, до какой-то степени, «схвачено» через частоту упоминания региона, в данном случае – в публичной речи депутатов.

Однако важно не только зафиксировать символическую карту регионов России, ее интерпретация должна соотноситься с контекстными условиями, которые могут влиять на место региона на этой карте. Очевидно, что конкретных причин упоминания региона конкретным депутатом может быть очень много, но представляется, что в первую очередь необходимо выделить один универсальный фактор, то есть действующий в отношении всего депутатского корпуса, причем, безотносительно персональных, политико-идеологических или иных характеристик конкретных депутатов. Таким фактором выступает электоральная связь депутата с регионом, то есть, говоря конкретнее, избирался ли он по одномандатному округу в конкретном регионе или по партийному списку (федеральная часть или региональная группа, включающая несколько регионов). Мы хотим проверить, будут ли «одномандатники», в силу различий в совокупности стимулов, чаще упоминать «свой» регион нежели «списочники», а также будут ли они чаще упоминать «свой» регион в сравнении с другими регионами, как минимум ПФО.

Методология исследования

В качестве исходных данных были использованы стенограммы выступлений депутатов Государственной думы Российской Федерации, находящиеся в свободном доступе на официальном интернет-портале Государственной думы. Массив данных включает 324 тыс. реплик от 2 773 депутатов; имеется также информация о региональной и партийной принадлежности депутатов (для тех созывов, где она была доступна). Исследование осуществлялось в несколько этапов; дизайн исследования представлен на рис. 1.

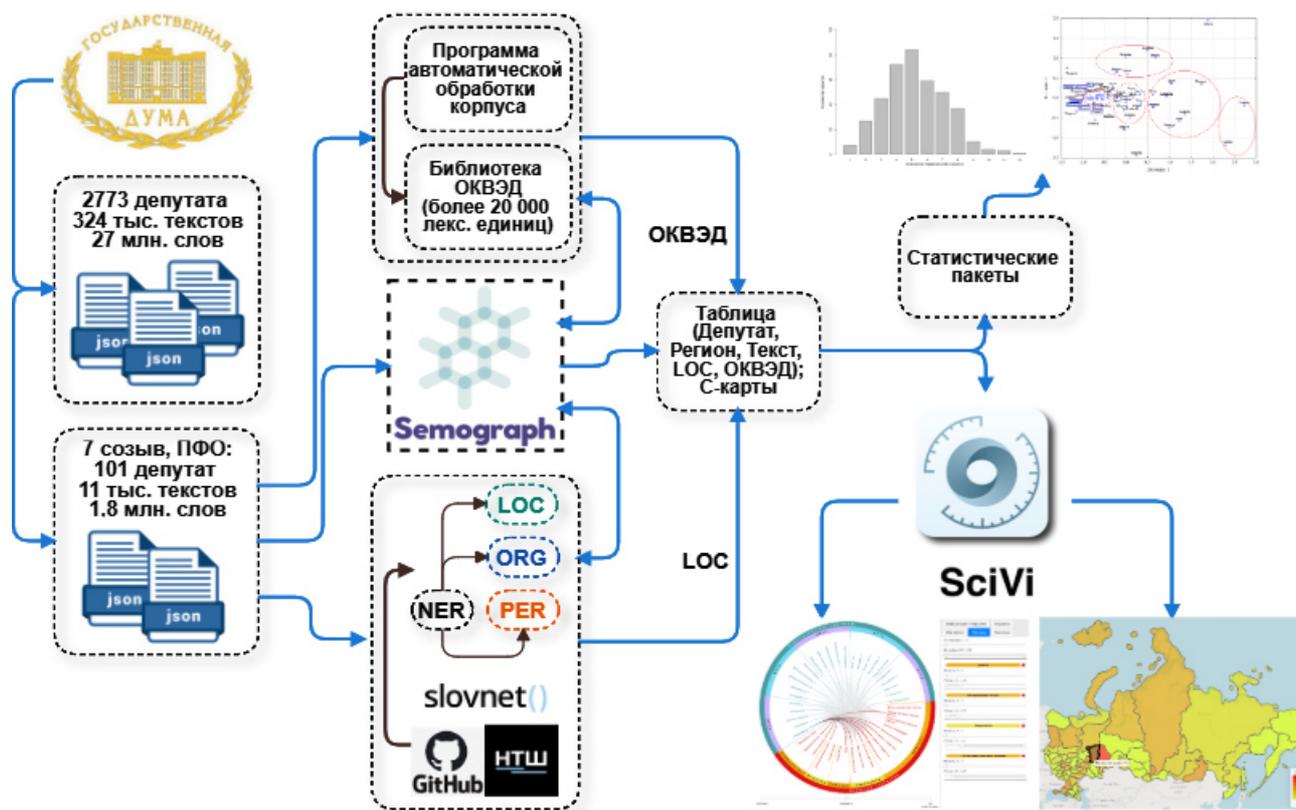


Рис. 1 Общий дизайн исследования дискурса депутатов ГД

Источник: собственная разработка авторов.

На первом этапе была собрана база стенограмм заседаний Государственной думы Российской Федерации. Отдельно были собраны профили депутатов ГД, представленные в открытом доступе¹, в том числе на старой версии сайта². Было установлено, что до 7-го созыва в профилях депутатов информация была представлена непоследовательно, часто отсутствовал регион, с которым аффилирован депутат. В данной работе используются материалы выступлений депутатов 7-го созыва ГД, представляющих регионы ПФО. Общий объем материала свыше 11 тыс. выступлений/реплик, состоящих из 1,8 млн словоупотреблений (подробнее см. табл. 1). Все собранные данные для удобства работы с контекстами (поиска, фильтрации) были внесены в информационную систему "Семограф"³.

Таблица 1

Общая характеристика анализируемых данных

	Депутатов	Реплик	Словоупотреблений
Всего	2773 чел.	324 тыс.	27 млн.
7 созыв	407 чел.	41 тыс.	3,7 млн.
7 созыв, ПФО	101 чел.	11 тыс.	1,8 млн

Источник: собственные расчеты авторов по собранной базе стенограмм заседаний ГД.

¹ Состав Государственной думы [электронный ресурс]. URL: <http://duma.gov.ru/duma/deputies> (дата обращения: 23.05.2020).

² Состав Государственной думы [электронный ресурс]. URL: <http://old.duma.gov.ru/about/history/convocations/1/> (дата обращения: 23.05.2020).

³ Информационная система "Семограф" [электронный ресурс]. URL: <https://semograph.org/> (дата обращения: 23.10.2020).

На втором этапе все тексты реплик каждого депутата были обработаны с помощью библиотеки SlovNet¹, реализующей извлечение именованных сущностей из текстов на естественном языке на основе методов глубинного обучения. В текстах реплик были выделены следующие категории сущностей: личность (PER), организация (ORG), географический объект (LOC). Основным интересом для нас в этой работе представляют именно географические объекты (названия городов, стран, регионов и т.п.).

Одновременно с помощью словаря ОКВЭД₂, составленного ранее на основе анализа текстов научных публикаций по экономике, каждой реплике автоматически присвоены коды ОКВЭД на основании вхождения понятий из словаря в текст реплики депутата с учетом количества этих вхождений. В работе использовался тот же метод, что применялся ранее для индексации прикладных научных исследований (см. Баранов и др., 2015). В результате предварительной автоматической обработки каждый текст корпуса выступлений был представлен в следующем виде (в качестве примера взято выступление депутата от Пермского края А. Б. Василенко):

```
"deputy" : "Василенко Александр Борисович",
"header" : "Заседание № 105",
"date" : "27.03.2018",
"convening" : 7,
"question" : "О проекте федерального закона № 378747-7 \"Об особенностях исчисления платы за негативное воздействие на окружающую среду на территориях Республики Крым и города федерального значения Севастополя\".",
"stage" : "Рассмотрение законопроекта в первом чтении",
"number" : 1,
"speech" : "ВАСИЛЕНКО А. Б., фракция \"ЕДИНАЯ РОССИЯ\". Уважаемый Иван Иванович, коллеги! Действительно, вот по вопросам и выступлениям видно, что мы понемногу становимся заложниками той экономической ситуации, которая сегодня сложилась в Крыму <...> Действительно, здесь звучало требование о том, что мы должны получить четкие планы развития этих предприятий, того, как будут строиться очистные сооружения, как они будут функционировать на деле, и вот с учетом выполнения этого условия фракция \"ЕДИНАЯ РОССИЯ\" предлагает поддержать законопроект в первом чтении, но он, его выполнение, я так сказал бы, подлежит очень четкому контролю.
"entities" : [
  {
    "text" : "ЕДИНАЯ РОССИЯ",
    "type" : "ORG",
    "start" : 1087,
    "stop" : 1100
  },
  {
    "text" : "Иван Иванович",
    "type" : "PER",
    "start" : 52,
    "stop" : 65
  },
  {
    "text" : "Крым",
    "type" : "LOC",
    "start" : 368,
    "stop" : 373
  }
],
"okved" : {
  "37. Сбор и обработка сточных вод" : 1
}
},
```

¹ SlovNet [online]. Available at: <https://github.com/natasha/slovnet> (Accessed: 12th August 2020).

В результате машинной обработки массива текстов выступлений каждому депутату было поставлено в соответствие значение региональной принадлежности (взято из профиля на сайте ГД) и набор сущностей, выделенных из текстов его выступлений. В данной работе рассматриваются только географические объекты, выделенные из текстов выступлений.

Третий этап анализа данных состоял в геокодировании выделенных географических объектов. Например, выделенную именованную сущность LOC "Мотовилихе" нужно привести к субъекту РФ "Пермский край". Для этого выполнялась нормализация текста именованной сущности с использованием библиотек проекта *Natasha*¹, а затем данные передавались геокодеру *OpenStreetMap*². Результатом работы геокодера стало отнесение всех геообъектов к субъектам РФ. Полученные данные дополнительно просматривались несколькими экспертами; были исправлены ошибки автоматического геокодирования. В случае неоднозначной интерпретации результатов использовались первичные данные стенограмм, внесенные в ИС "Семограф".

Таким образом, результатом работы стала таблица данных, представляющая для каждого субъекта Российской Федерации все упоминания других субъектов РФ.

Четвертый этап состоял в визуализации полученных данных; для этого была использована платформа *SciVi* (*Ryabinin and Churpina, 2017*). Программная архитектура платформы *SciVi* организована таким образом, что внесение новой и корректировка существующей функциональности в области предоставляемых платформой средств визуальной аналитики осуществляется исключительно пополнением/модификацией онтологий в её базе знаний, без модификации программного кода ядра *SciVi*.

Ранее платформа *SciVi* активно применялась в задачах анализа данных высокой связности, для чего в её состав были включены средства визуализации различных видов графов (*Рябинин и др., 2018*). В настоящем исследовании возникла необходимость анализа данных, имеющих привязки к регионам России, в связи с чем потребовалось добавить в *SciVi* поддержку визуализации географических карт. Модуль визуализации карт был реализован на основе библиотеки *Leaflet*³, с использованием картографического провайдера *OpenStreetMap*.

В качестве входных данных для добавленного в *SciVi* картографического модуля выступают контуры 85 субъектов Российской Федерации, взятые из *OpenStreetMap* и представленные в формате *GeoJSON*⁴, а также полученная ранее таблица упоминаний этих субъектов в репликах депутатов Государственной думы.

Таблица устроена следующим образом: строки и столбцы озаглавлены названиями субъектов РФ, а на пересечении каждого *i*-го столбца и каждой *j*-й строки находится число устных упоминаний депутатами, представляющими *j*-й субъект РФ, геоконцептов, относящихся к *i*-му субъекту федерации. Число столбцов равно числу строк, и на главной диагонали располагается количество упоминаний депутатами субъектов своего избирательного округа.

На рисунке 2 приведена схема алгоритма отображения данных на карте. Эта схема (называемая диаграммой потоков данных) строится пользователем в специальном визуальном редакторе⁵ в составе платформы *SciVi* (с возможностью сохранения и последующей загрузки) из блоков, доступных в палитре инструментов. Диаграмма потоков данных задаёт конкретные преобразования данных, предшествующие их визуализации, а также определяет назначение таких интерактивных средств, как, например, выделение отдельных объектов кликом мыши. Подход на основе построения диаграмм потоков данных значительно увеличивает гибкость и аналитические возможности платформы, позволяя быстро перебрать разные варианты фильтрации данных и их графического изображения с целью поиска наиболее адекватных решаемой задаче визуальных средств, а также с целью проверки различных гипотез, выдвигаемых в процессе анализа.

¹ *Natasha* [online]. Available at: <https://github.com/natasha/natasha> (accessed: 12th August 2020).

² *OpenStreetMap* [электронный ресурс]. URL: <https://www.openstreetmap.org/> (дата обращения: 23.10.2020).

³ *Leaflet* [online]. Available at: <https://leafletjs.com> (accessed: 12th August 2020).

⁴ *GeoJSON* [online]. Available at: <https://geojson.org> (accessed: 12th August 2020).

⁵ Система визуализации *SciVi* [электронный ресурс]. URL: <https://scivi.semograph.com/> (дата обращения: 23.10.2020).

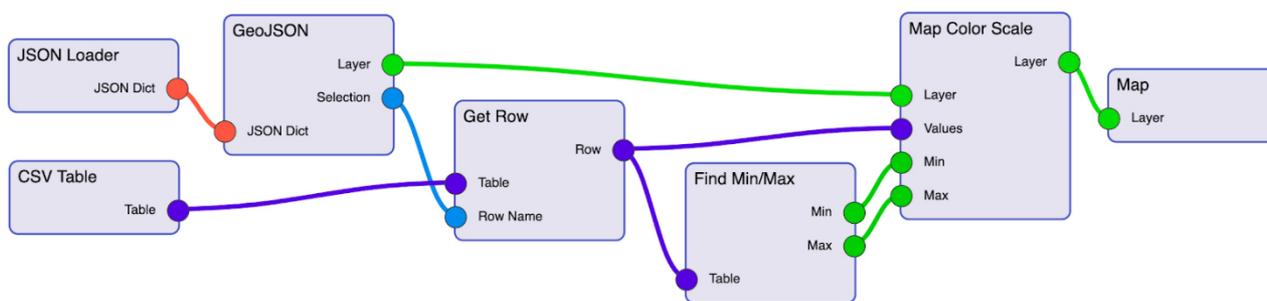


Рис. 2. Диаграмма потоков данных в среде платформы визуальной аналитики SciVi

Прямоугольные блоки на диаграмме потоков данных обозначают шаги загрузки, преобразования и визуализации данных; линии обозначают связи по данным; цвета линий маркируют типы передаваемых данных.

Диаграмма потоков данных, показанная на рис. 2, задаёт вариант визуализации, при котором пользователь может «кликнуть» на интересующий его регион на карте, и все отображаемые регионы окрасятся в цвета, зависящие от количества упоминаний этих регионов депутатами, представляющими выбранный регион (см. рис. 3). При наведении курсора на каждый из регионов выводится всплывающая подсказка с названием этого региона и числом его упоминаний. Таким образом, для каждого субъекта федерации можно быстро оценить, о каких регионах больше/меньше говорят представляющие его депутаты.

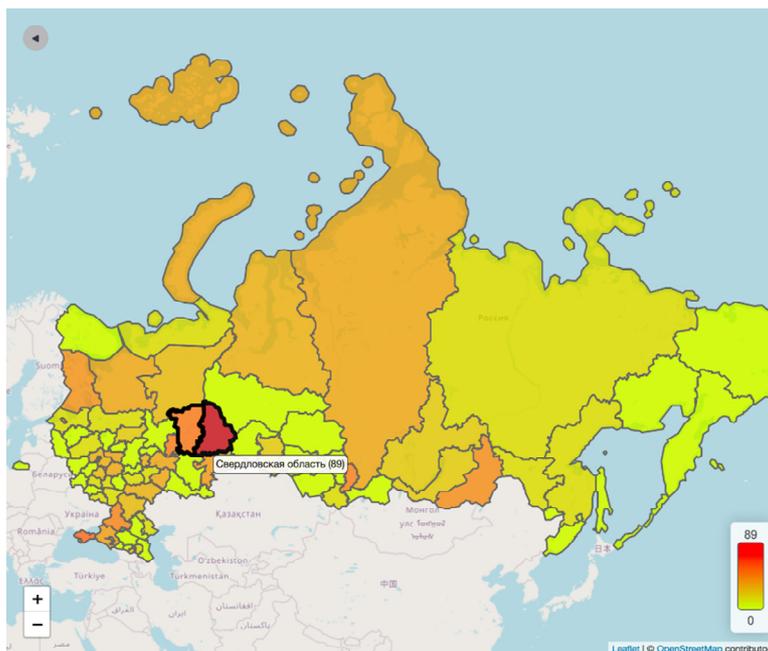


Рис. 3. Визуализация числа упоминаний разных регионов России депутатами, представляющими Пермский край

Источник: Система визуализации SciVi (<https://scivi.semograph.com/>).

Использование диаграмм потоков данных в качестве средства настройки визуализации позволяет очень просто и быстро решить и обратную задачу – узнать для выбранного субъекта федерации, депутаты от каких субъектов говорят о нём в большей/меньшей степени. Для этого в диаграмму добавляется единственный дополнительный блок, задающий транспонирование входной таблицы (см. рис. 4).

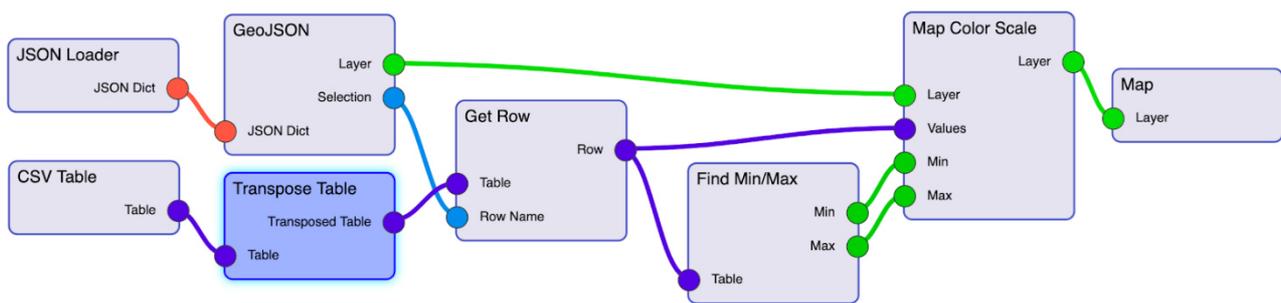


Рис. 4. Модифицированная диаграмма потоков данных в среде платформы визуальной аналитики SciVi

На рисунке 5 представлена, для примера, карта упоминаний Пермского края в выступлениях депутатов ГД. На рисунке видно, что Пермский край редко упоминается в выступлениях не только депутатов, представляющих другие субъекты ПФО, но и самими представителями Пермского края (15 случаев). Сопоставление с рис. 3 свидетельствует о большем интересе, проявленном депутатами от Пермского края к другим регионам, в первую очередь, к Свердловской области (89 упоминаний), которая намного «популярнее» у «пермских» депутатов любого региона ПФО. Использовать интерактивный граф можно по ссылке [<https://scivi.semograph.com/?preset=PFOTranspose.json>].

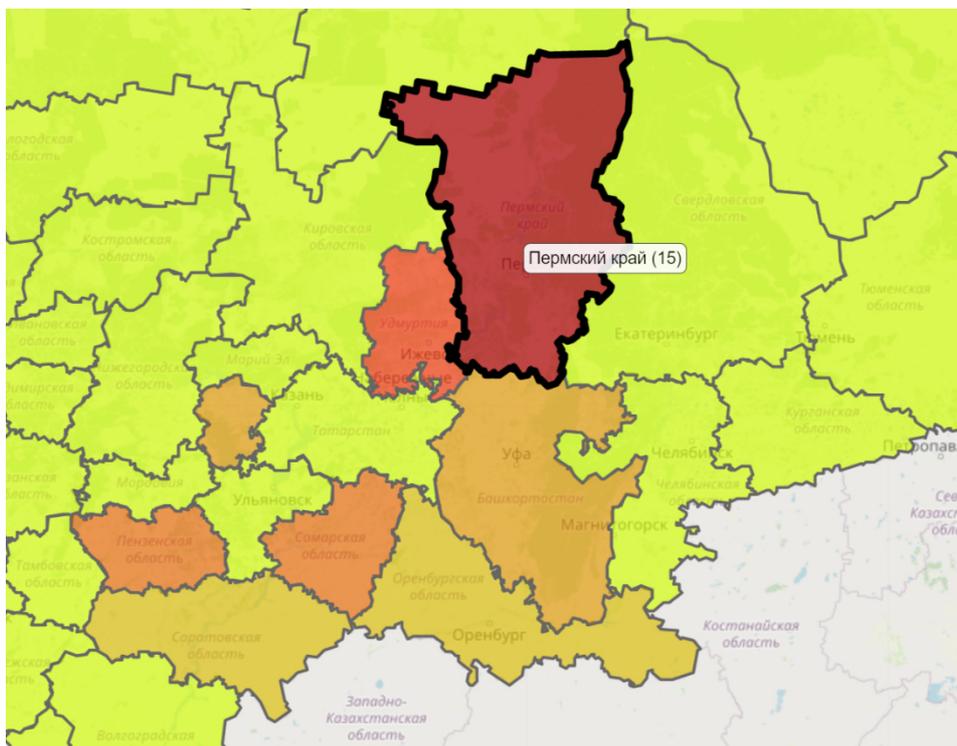


Рис. 5. Визуализация числа упоминаний Пермского края депутатами, представляющими субъекты ПФО

Источник: Система визуализации SciVi (<https://scivi.semograph.com/>).

«Региональное картирование» депутатов от регионов ПФО

Таким образом, количественный анализ и визуализация результатов позволили поставить проблему упоминаний субъектов РФ в выступлениях депутатов. Причины, по которым депутаты не упоминают в речи регионы, которые они представляют, нуждаются в отдельном исследовании с опорой на контекстуальный анализ. На данном этапе мы можем высказать предположение, сформированное в рамках количественного подхода: вероятно, «одномандатники» в большей степени обращаются к реалиям, соотношенным с географией своего региона, чем «списочники». Для проверки гипотезы

тезы все депутаты были разделены на три группы в логике связи с одним или несколькими регионами и институциональным характером этой связи: 1) представляющие один регион ПФО, 2) представляющие несколько регионов ПФО; 3) представляющие один регион в ПФО и входящие в списочную часть других регионов за пределами ПФО (либо представляющие один регион ПФО и все регионы РФ).

Всего депутатов, аффилированных тем или иным образом с регионами ПФО на середину 2020 г., – 101 чел. (в сентябре 2020 г. проходили довыборы, но победители в выборке отсутствуют). В выборку попали только те депутаты, которые выступали в 7-м созыве (т.е. их выступления присутствуют в информационной системе Госдумы). Всего таких оказалось 93 чел. Из них 38 – одномандатники, 35 – депутаты, которые избирались в составе региональных групп партийных списков, большая часть которых покрывала регионы ПФО (эти депутаты «представляют» в ПФО от двух до четырех регионов каждый), а также 20 чел., избравшихся по спискам, но внутри ПФО имеющих привязку только к одному региону (то есть от региональных групп, включавших только один регион в ПФО, или от федеральной части). В таблице 2 представлены данные по типам и количеству депутатов в разрезе регионов.

Таблица 2

Регионы ПФО по количеству и разновидностям депутатов в выборке

Регионы	Все депутаты	«Одномандатники»	«Списочники»	«Списочники» – один регион в ПФО
Кировская область	5	1	2	2
Нижегородская область	13	5	7	1
Оренбургская область	12	3	8	1
Пензенская область	11	1	9	1
Пермский край	9	4	3	2
Башкортостан	14	5	2	7
Марий Эл	11	1	10	–
Мордовия	11	1	10	–
Татарстан	11	3	2	6
Самарская область	13	5	8	–
Саратовская область	12	3	9	–
Удмуртия	7	2	5	–
Ульяновская область	11	2	9	–
Чувашия	12	2	10	–

Для каждого из трех списков была сгенерирована таблица, включающая показатели упоминаний своего региона (регионов), других регионов ПФО и других регионов за пределами ПФО. Для каждого депутата вычислялись показатели доли упоминаний своего региона от общего количества обращения в речи к географическим локациям в пределах страны. Далее было сформировано две выборки: «одномандатники» и «списочники». В первую группу были внесены часть «списочников», избранных по федеральному партийному списку, но по конкретному региону ПФО – включение основывается на ожидании, что они могут вести себя (в данном случае в отношении упоминания регионов) скорее, как одномандатники. Из каждой выборки были исключены депутаты с небольшим общим количеством упоминаний регионов (порог отсечения в двух списках составил 30 упоминаний). Сформированные выборки анализировались с помощью U-критерия Манна-Уитни, который позволяет выявлять различия между малыми выборками разного объема (результаты см. табл. 3).

**Результаты сопоставления выборок «одномандатников» и «списочников»
(U-критерий Манна-Уитни)**

Сум.ранг	Сум.ранг	$U_{Эмп}$	$U_{Кр}$	p-уров.	N	N
185	166	46	50	0,05	11	15

Таким образом, результаты показывают различия в количестве обращений к своим регионам в выступлениях «одномандатников» и «списочников» при $p < 0,05$. Повышение уровня статистической значимости исследования возможно в случае общего увеличения выборок (при анализе ситуации на уровне всех субъектов РФ).

В таблице 4 представлены данные по упоминаниям депутатами-«одномандатниками» или «списочниками»¹ своих и других регионов в разрезе регионов, с которыми они аффилированы. Эти данные также показывают существенную и ожидаемую разницу между двумя разновидностями депутатов. «Одномандатники» в среднем по выборке регионов намного чаще, почти вдвое, упоминают именно свой регион. Они также в три раза чаще упоминают свой регион, нежели другие регионы в ПФО. А депутаты, избранные по партийным спискам, и аффилированные с несколькими регионами в ПФО, склонны в среднем чаще упоминать другие регионы в ПФО, нежели «свои». Но при этом можно наблюдать значительную межрегиональную вариацию по частоте упоминаний «своего» региона депутатами. «Одномандатники» Кировской и Пензенской области не делают этого совсем. Тогда как депутаты от Ульяновской области, Чувашии, Татарстана, Саратовской, Нижегородской и Самарской областей упоминают «свой» регион намного чаще своих коллег, избранных по партийным спискам.

Таблица 4

Частота упоминаний регионов в выступлениях «одномандатников» и «списочников»

Регионы	«Одномандатники» Кол-во упоминаний регионов (на одного депутата)			«Списочники» Кол-во упоминаний регионов (на одного депутата)		
	«Свой» регион	Другие регионы ПФО	Регионы вне ПФО	«Свои» регионы	Другие регионы ПФО	Регионы вне ПФО
Ульяновская область	49,5	14	62	1,9	3,4	20,8
Чувашия	28	4	39	9,3	9,4	39,4
Татарстан	24,7	10,3	26,7	4	7	48,5
Саратовская область	19	1	14,3	1,4	1,8	22,7
Нижегородская область	8	0,4	4,2	0,4	7	34,9
Башкортостан	5,6	2,6	26,2	21	1	25,5
Самарская область	4,2	12,2	29,8	1,5	2,5	42,3
Марий Эл	3	0	1	3	22,6	46,7
Удмуртия	2,5	1	1,5	2,4	13,2	36,4
Мордовия	2	2	18	2	16,7	39,4
Пермский край	1	0,3	3	3	7,7	49
Оренбургская область	0,3	1	5,3	0,3	3,8	42,3
Кировская область	0	0	9	28,5	11	43
Пензенская область	0	1	34	0,4	2,8	22,7
Среднее	10,56	3,56	19,57	5,65	7,84	36,67

¹ Для сравнения выбраны только те, которые имеют аффилиацию с несколькими регионами в ПФО.

Значимых исключений из такой логики соотношения всего два – Башкортостан и Кировская область, в которых «списочники» оказываются намного активнее. Например, в случае Кировской области, ее 57 раз упоминает К.И.Черкасов, который избран по региональной группе списка ЛДПР, включавшей, кроме Кировской области, также республики Марий Эл и Коми. Марий Эл депутат упоминает только 8 раз и 82 раза регионы за пределами ПФО, возможно в это число входит упоминание Республики Коми. Объяснить особое внимание депутата к Кировской области можно тем фактом, что некоторое количество лет он работал в ней.

При этом и некоторые другие «списочники» демонстрируют похожую логику упоминаний. Председатель Госдумы Володин, избранный по партийному списку, чаще других среди регионов ПФО упоминает Саратовскую область, откуда он родом и где начинал свою политическую карьеру. Подобное же можно предположить и в отношении В.К.Гартунга, который был избран по региональной группе списка Справедливой России, включающей Челябинскую, Оренбургскую и Самарскую области. Он один раз за созыв упомянул Оренбургскую область, ни разу – Самарскую область и всего четыре раза – другие регионы ПФО. Но у него есть 183 упоминания регионов за пределами ПФО, при этом сам он из Челябинской области, которая не входит в ПФО и данные по ней не попали в выборку, но можно предполагать, что часть этих внешних для ПФО упоминаний приходится на Челябинскую область.

Депутаты, избранные в одномандатных округах, также могут отличаться в своем упоминании регионов от ожидаемой логики. Например, на четырех таких депутатов, аффилированных с Пермским краем, приходится всего четыре упоминания этого субъекта РФ. Другие регионы ПФО они, правда, упоминают еще реже. Депутат Хинштейн, который избран в одномандатном округе Самарской области, упоминает ее намного чаще других регионов, за исключением Нижегородской, от которой избирался раньше.

Но эти региональные и индивидуальные отклонения от ожидаемого распределения упоминаний регионов не меняют общую картину. Совокупность депутатов, избранных в одномандатных округах, демонстрирует, по крайней мере на этой выборке, общую логику соотношения упоминания «своих» и других регионов как в ПФО, так и за его пределами. Коэффициент корреляции в первом случае составляет 0,683, во втором 0,755 (корреляция значима на уровне 0,01, 2-сторонняя). В отношении депутатов, избранных по партийным спискам и имеющих множественную региональную аффилиацию, такие значимые связи отсутствуют. Однако необходимо отметить, что такую картину можно наблюдать только, если рассматривать депутатов именно в региональном ракурсе. Если же смотреть на всю совокупность депутатов, то «списочники» демонстрируют даже более сильные связи между упоминанием своих номинальных регионов и других регионов, т.е. отличия возникают именно в силу региональных комбинаций разных депутатов, но эта тема выходит за пределы нашего интереса.

Полученные данные позволяют строить сводные карты разметки символического регионального пространства, которые могут включать три перспективы взгляда:

- 1) «мы» – самопредставление, то есть упоминания своего региона,
- 2) «они» – взгляд на «других», то есть упоминания других регионов,
- 3) «они о нас» – взгляд «других», то есть упоминание конкретного региона депутатами от других регионов.

Интерес представляют все три возможных сочетаний этих перспектив. Предварительное понимание возможного соотношения, например, самопредставления и взгляда других дает диаграмма, представленная на рис. 6. Предварительность определяется возможным несовпадением пространства ПФО с пространством, которое действительно значимо для самих депутатов – выше отмечалось, что депутаты Пермского края намного чаще любого региона ПФО упоминают Свердловскую область, которая к ПФО не относится. На диаграммах сведены данные по депутатам, аффилированным с одним регионом внутри ПФО, то есть и избранных по одномандатным округам и по спискам, но не имеющих других «своих» регионов внутри округа.

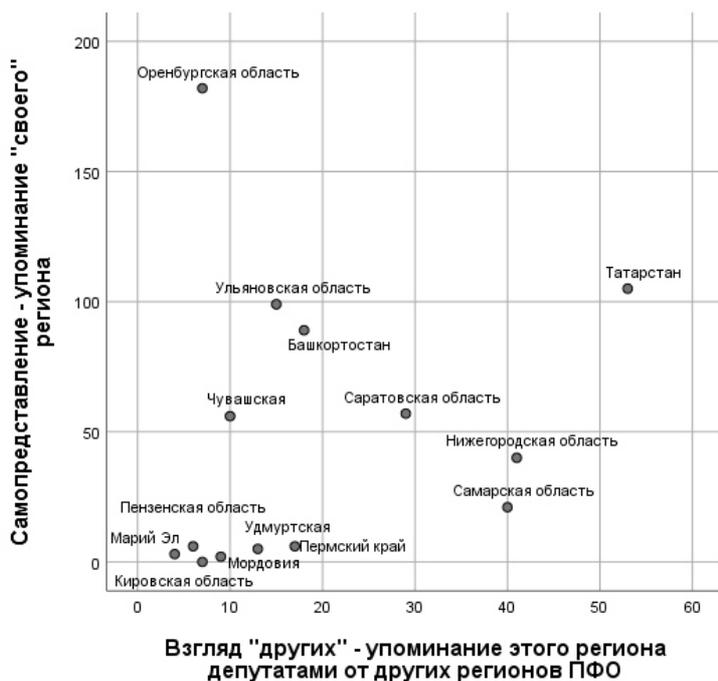


Рис. 6. Соотношение «мы» и взгляда «других» в упоминаниях депутатов «своих» и «других» регионов ПФО

Можно видеть, например, «выброс» самопредставления Оренбургской области, который резко контрастирует с вниманием к этому региону со стороны других депутатов. И напротив, большое внимание к «мы» у депутатов, связанных с Татарстаном, хорошо соотносится с вниманием к нему депутатов из других регионов. Также очевидна очень слабая выраженность «мы» у целой группы регионов, которая отражена на диаграмме в левом углу, и в целом это совпадает с их относительной «непопулярностью» у других. Но, повторимся, явно требуется расширить используемые данные путем привлечения информации по другим регионам, чтобы получить более релевантную картину.

* * *

Представленный методологический и технологический подход к анализу дискурса депутатов Государственной думы позволяет наметить способы реконструкции их картины и модели мира. Апробация количественного метода на материале упоминаний регионов в выступлениях депутатов, электорально аффилированных с регионами ПФО, позволила подтвердить ожидания о связи между типом представителя региона («одномандатник» или «списочник») и связью с ним в рамках создаваемого символического пространства. Также получены данные для построения карт разметки, в данном случае, символического регионального пространства, в публичных речах депутатов.

Развитие метода мы видим как в плоскости более широкого охвата регионов (вплоть до всех регионов страны), так и в учете контекстуального фактора упоминаний географических региональных реалий. Это позволит, в частности, выявить причины (мотивацию) номинаций регионов в публичной политической речи. Другое направление смыслового анализа, связанное с геотегированием, – тематизация региональных проблем, хорошо представляемая с помощью выявленных ОКВЭДов.

И, наконец, развитие метода, как видно из проведенной работы, связано с расширением средств визуальной аналитики. Благодаря тому, что платформа визуальной аналитики SciVi управляется расширяемой онтологической базой знаний, её функциональность может быть быстро пополнена новыми аналитическими средствами, требуемыми для геопространственного анализа. Один из таких конкретных инструментов – отображение кругового графа рядом с картой и демонстрация на нем связанных с выбранным регионом данных по цитированиям и другим аспектам (например, сгенерированным на основании отнесения к ОКВЭД экономическим профилям регионов в представлениях депутатов). Визуальная аналитика нуждается и в инструментах сопоставления двух выбранных регионов, например, отображения взаимных упоминаний регионов стрелками на карте, или отображения сходств/различий по другим значимым параметрам на круговом графе. Эти и другие инструмен-

ты особенно актуальны при работе с многопараметрическими данными (в нашем случае, это регионы, депутаты, сущности (локации, организации, персоны), ОКВЭДы и др.), так как позволяют «на лету» выявлять связи между объектами изучаемого фрагмента действительности. Кроме того, средства пространственной аналитики обладают собственными возможностями, учитывающими, в том числе, и расположение регионов, наличие общей границы и другое.

Таким образом синтез методов машинного обучения, лингвистической семантики и визуальной аналитики при работе с большими объемами текстовых данных – стенограммами заседаний ГД может служить новым источником для политологических исследований картины мира, формируемой официальным политическим дискурсом.

Дополнительные материалы

Интерактивный граф упоминаний субъектом ПФО других субъектов РФ см. <https://scivi.semograph.com/?preset=PFO.js>

Интерактивный граф упоминаний субъекта ПФО другими субъектами ПФО: см. <https://scivi.semograph.com/?preset=PFOtranspose.json>

При переходе по ссылкам загружается окно потока данных; для просмотра интерактивных графов нужно перейти по ссылке VISUALIZE, расположенной в правом верхнем углу окна.

Финансовая поддержка

Статья подготовлена при финансовой поддержке РФФИ, грант № 20-011-00624 «Будущее России: содержание, смыслы и границы национального согласия».

Библиографический список

- Рябинин, К. В., Белоусов, К. И., Чуприна, С. И., Щebetенко, С. А., Пермьяков, С. С. (2018) 'Средства визуальной аналитики для комплексного исследования результатов многопараметрического описания пользователей социальных интернет-сервисов', *Научная визуализация*. Москва: Национальный исследовательский ядерный университет МИФИ, к. 3, т. 10, 4, сс. 82–99, doi: 10.26583/sv.10.4.07. [Ryabinin, K. V., Belousov, K. I., Chuprina, S. I., Schebetenko, S. A., Permyakov, S. S. (2018) 'Visual analytics tools for a comprehensive study of the results of a multi-parameter description of users of social Internet services' [*Sredstva vizual'noi analitiki dlya kompleksnogo issledovaniya rezul'tatov mnogoparametricheskogo opisaniya pol'zovatelei sotsial'nykh internet-servisov*], *Nauchnaya vizualizatsiya*. Moscow: Natsional'nyi issledovatel'skii yadernyi universitet MIFI, k. 3, t. 10, 4, pp. 82–99, doi: 10.26583/sv.10.4.07 (In Russ.)].
- Семененко, И. С. (2007) 'Социокультурные механизмы формирования и восприятия образа России' в: *Образ России в мире: становление, восприятие, трансформация* / отв. ред. И. С. Семененко. Москва: ИМЭМО, сс. 61–78. [Semenenko, I. S. (2007) 'Socio-cultural mechanisms of formation and perception of the image of Russia' in: *The image of Russia in the world: formation, perception, transformation*, I. S. Semenenko (ed.) [*Sotsiokul'turnye mekhanizmy formirovaniya i vospriyatiya obraza Rossii v: Obraz Rossii v mire: stanovlenie, vospriyatie, transformatsiya*], Moscow: IMEMO, pp. 61–78. (In Russ.)].
- Belousov, K. I., Baranov, D. A., Zelyanskaya, N. L., Karlina, T. V. (2015) 'The use of economic classifiers for the indexing of scientific publications', *Scientific and Technical Information Processing*, vol. 42, 4, pp. 299–305, doi:10.3103/S0147688215040115.
- Gandhi, J., Noble, B., Svulik, M. (2020) 'Legislatures and Legislative Politics Without Democracy', *Comparative Political Studies*, vol. 53(9). pp. 1359–1379, doi: 10.1177/0010414020919930.
- Lauderdale, B., Herzog, A. (2016) 'Measuring Political Positions from Legislative Speech?', *Political Analysis*, 24(3), pp. 374–394.
- Noble, B., Schulmann, E. (2018) 'Not just a rubber stamp: Parliament and lawmaking' in: Treisman D. (ed.) *The New Autocracy: Information, Politics, and Policy in Putin's Russia*. Washington, D.C.: Brookings Institution Press, pp. 49–82.

- Petrov, N., Nazrullaeva, E. (2018) 'Regional Elites and Moscow' in: Treisman D. (ed.) *The New Autocracy: Information, Politics, and Policy in Putin's Russia*. Washington, D.C.: Brookings Institution Press, pp. 109–136.
- Proksch, S.-O., Slapin, J. B. (2012) 'Institutional Foundations of Legislative Speech', *American Journal of Political Science*, vol. 56, no. 3, pp. 520–537.
- Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., Radev, D. R. (2010) 'How to Analyze Political Attention with Minimal Assumptions and Costs', *American Journal of Political Science*, 54(1), pp. 209–228, doi:10.1111/j.1540-5907.2009.00427.x.
- Ryabinin, K., Chuprina, S. (2017) 'High-Level Toolset For Comprehensive Visual Data Analysis and Model Validation', *Procedia Computer Science*, vol. 108, pp. 2090–2099, doi: 10.1016/j.procs.2017.05.050.

Информация об авторах

Белусов Константин Игоревич – доктор филологических наук, профессор кафедры теоретического и прикладного языкознания, Пермский государственный национальный исследовательский университет. E-mail: belousovki@gmail.com (ORCID: 0000-0003-4447-1288. Researcher ID: A-4891-2016).

Лабутин Иван Александрович – аспирант, ассистент кафедры математического обеспечения вычислительных систем, Пермский государственный национальный исследовательский университет. E-mail: barracuda72@bk.ru (ORCID: 0000-0001-6858-1479).

Рябинин Константин Валентинович – кандидат физико-математических наук, доцент, доцент кафедры математического обеспечения вычислительных систем, Пермский государственный национальный исследовательский университет. E-mail: kostya.ryabinin@gmail.com (ORCID: 0000-0002-8353-7641. Researcher ID: N-7890-2016).

Сулимов Константин Андреевич – кандидат политических наук, заведующий кафедрой политических наук, Пермский государственный национальный исследовательский университет. E-mail: k.sulimov@psu.ru (ORCID ID: 0000-0003-4277-3135. Researcher ID: N9455-2018).

Статья принята к печати: 30.10.2020

"DEPUTIES CARTOGRAPHY" ON THE MATERIAL OF THE STENOGRAMS OF THE SITTINGS OF THE STATE DUMA

K. I. Belousov

Perm State University

I. A. Labutin

Perm State University

K. V. Ryabinin

Perm State University

K. A. Sulimov

Perm State University

Abstract

Public speeches of the deputies of the Russian Federation State Duma represent productive empirical material that requires an interdisciplinary approach to its analysis. The goal of the reported study is defining the degree of the concord regarding the future of Russia among deputies and the localization of this agreement in the semantic space (equilibrium point). The empirical base under consideration comprises the transcripts of Russian Federation State Duma sittings. This dataset covers time period from 1994 to mid-2020 and includes

324 thousand phrases (27 million words) from 2773 deputies and other people. The paper presents the general design of the study of the State Duma deputies discourse using the "Semograph" information system, the SlovNet library for natural language text processing based on deep learning, the SciVi visual analytics platform, the map visualization module based on the Leaflet library, the geocoding of geographic objects based on the OpenStreetMap map provider and other tools. The application of the approach showed the importance of the institutional foundations of the symbolic deputies' mapping of the regions of the Volga Federal District (the electoral connection of a deputy with the region – deputies elected by party list or pluralist rule).

Keywords: deputies; country image; The State Duma; regions; machine processing; data; visualization.

Acknowledgments:

The reported study was funded by RFBR, project number 20-011-00624.