

УДК 519.66

UDC 519.66

05.00.00 Технические науки

Technical Sciences

**СОЦИОКУЛЬТУРНЫЕ ТЕКСТЫ КАК  
ОБЪЕКТ МАТЕМАТИЧЕСКОЙ  
СТАТИСТИКИ В СФЕРЕ ВЫСШЕГО  
ОБРАЗОВАНИЯ**

**SOCIO-CULTURAL TEXTS AS AN OBJECT OF  
MATHEMATICAL STATISTICS IN THE FIELD  
OF HIGHER EDUCATION**

Алиева Заира Магомедовна  
аспирант кафедры Информационных технологий и  
электронных библиотек  
*Московского Государственного Института  
Культуры, Москва, Россия*  
[Zaira1991@yandex.ru](mailto:Zaira1991@yandex.ru)

Aliyeva Zaira Magomedovna  
postgraduate student, the Department of Information  
technologies and electronic libraries  
*Moscow State Institute Of Culture, Moscow, Russia*  
[Zaira1991@yandex.ru](mailto:Zaira1991@yandex.ru)

В статье анализируются вопросы анализа социокультурных текстов методами математической статистики применительно к сфере высшего образования. Раскрывается сущность социокультурной парадигмы и важности социокультурных факторов в научно-исследовательской деятельности в высших учебных заведениях. Рассмотрены примеры использования информационных технологий, основанных на методах математической статистики для анализа социокультурных текстов. Сделан вывод о том, что социокультурный компонент в обучении способствует реализации социокультурной компетенции, а применение математических и статистических методов при анализе социокультурных текстов способствует получению содержательных выводов за счет возможности анализа больших массивов информации и учета значительного количества факторов

The article analyzes the issues of social and cultural analysis of texts by methods of mathematical statistics applied to the field of higher education. It presents the essence of the socio-cultural paradigm and the importance of sociocultural factors in research activity in higher education institutions. The examples of use of information technologies, based on methods of mathematical statistics for the analysis of social and cultural texts were shown. It is concluded that socio-cultural component in education contributes to the realization of socio-cultural competence, and application of mathematical and statistical methods in the analysis of socio-cultural texts helps to obtain meaningful conclusions due to the possibility of analyzing large amounts of information and larger number of factors

Ключевые слова: СОЦИОКУЛЬТУРНЫЙ ТЕКСТ, СОЦИОКУЛЬТУРНЫЙ КОМПОНЕНТ, МЕТОД, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, АНАЛИЗ, ВЕРОЯТНОСТЬ

Keywords: SOCIO-CULTURAL TEXT, SOCIOCULTURAL COMPONENT, METHOD, MATHEMATICAL STATISTICS, ANALYSIS, PROBABILITY

В последнее время в исследованиях растет доля работ, декларирующих принадлежность к системному подходу, а также опирающихся на общую теорию систем. В связи с этим наблюдается рост экспорта адаптированных теорий и методик из сферы математики и семиотики в сферу социокультурных исследований.

При этом социокультурный подход представлен методологическим подходом на базе системного подхода, суть которого состоит в рассмотрении общества как единства культуры и социальности [1, С. 27].

Применительно к системе высшего образования необходимо отметить, что управление научно-исследовательской деятельностью в университетах в современных социокультурных условиях необходимо рассматривать не только как субъект управления, который обеспечивает формирование компетентности обучающихся в научно-исследовательской деятельности, но также и затрагивать аспекты более высокого, качественного состояния, при котором актуализируются политические, экономические и культурные факторы общественного развития, являющиеся непосредственными стимулами совершенствования личности.

Важнейшим фактором повышения результативности управления научно-исследовательской деятельностью в университетах в социокультурных условиях являются структурно-функциональная модель, включающая взаимосвязанные элементы: цели; принципы, функции и методы управления; структурные компоненты научно-исследовательской деятельности; технологию и результат [2, С. 84].

Социокультурный компонент в обучении способствует реализации социокультурной компетенции, которая представлена авторами в качестве целостной системы представлений о ключевых национальных традициях, обычаях и религиях государства. Большинство исследователей разделяют эту точку зрения, при этом конкретизируя и расширяя сам термин социокультурного компонента. Так, Л.Н. Полушина рассматривает социокультурный компонент с двух позиций:

- 1) общественные традиции;
- 2) социальные ритуалы [3, С. 99].

Е. Шварко в своем исследовании пришел к выводу, что ключевой целью обучения социокультурному компоненту является изучение

истории, географии, литературы, искусства, но что более важно - образа жизни народа, с более полной картиной общественных традиций [3, С. 100].

Однако анализ социокультурных текстов в высшем образовании наиболее полно можно провести при использовании не только качественных, но и количественных методов. Так, применение математических и статистических методов способствует получению содержательных выводов за счет возможности анализа больших массивов информации и учета значительного количества факторов. Фактически без применения математического аппарата трудно обойтись при решении любой практически значимой задачи анализа социально-гуманитарных явлений. Математика предоставляет возможность пользоваться интеллектуальными достижениями, которые накопило человечество для изучения конкретных объектов.

Использование математической статистики при исследовании социокультурных текстов позволяет предотвратить рассматриваемую возможность абстрагирования от реальности. Кроме того, использование математической статистики способствует получению качественных содержательных выводов за счет анализа больших массивов информации и учета значительного количества факторов [5, С. 189].

Кроме того, использование математических и статистических методов способствует получению содержательных выводов за счет возможности анализа больших массивов информации и учета значительного количества факторов. Фактически без применения математического аппарата трудно обойтись при решении любой практически значимой задачи анализа социально-гуманитарных явлений. Математика предоставляет возможность пользоваться интеллектуальными достижениями, которые накопило человечество для изучения конкретных объектов [2].

Текст представляет собой образующую единое целое последовательность знаков. Согласно мнению М. Хайдеггера, «естественнонаучному эксперименту соответствует в историко-гуманитарных науках критика источников», то есть текстов разной природы [7, с. 98]. При этом отсутствие текста, не позволяет выделить объект исследования для социально-гуманитарных наук. Это наглядно проявляется в таких гуманитарных науках, как лингвистика, филология, искусствоведение и так далее. Для гуманитарных наук, по мнению М.М. Бахтина, их предмет исследования, то есть «дух (и свой, и чужой) не может быть представлен как вещь (прямой объект естественных наук), а только в знаковом выражении, реализации в текстах и себя самого, и другого» [5, с. 84]. Потому для ученых-гуманитариев необходимо глубокое и тонкое понимание текста, а также количественная интерпретация с использованием математико-статистических методов.

Рассмотрим пример использования метода математической статистики для анализа социокультурных текстов. Предлагаемый метод основан на учете статистики употребления пар элементов, следующих друг за другом в тексте, то есть на формальной математической модели следующих букв текста при реализации цепи Маркова. По рассматриваемым текстам авторов, которые достоверно им создавались, вычисляется матрица переходных частот для каждой пары элементов. Она служит в качестве оценки матрицы вероятности перехода от одного элемента к другому. Для каждого автора строится матрица переходных частот и оценивается вероятность.

Автором текста считают того, для кого вычисленная оценка больше. Исходный корпус текстов после предварительной обработки представлен следующими вариантами:

а) пары букв в их естественных последовательностях в тексте – в словах и пробелах между ними;

б) пары букв в последовательностях букв в приведенных формах слов;

в) пары обобщенных грамматических классов слов в их последовательностях в предложениях социокультурного текста.

К таким классам слов относятся части речи и некоторые условные категории, представленные терминами «конца предложения», «сокращений» и др.;

г) пары менее обобщенных классов слов, включающих семантико-грамматические разряды, в состав которых входят одушевленные и неодушевленные существительные, качественные, относительные, притяжательные прилагательные и т. д.

В процессе обработки отбрасываются все слова, для которых не удастся автоматически определить грамматический класс, а также знаки препинания, слова с заглавной буквы. Каждый символ кодируется числом. При этом производится перекрестная проверка метода по материалам 385 текстов 82 авторов. В качестве показателя точности метода представлен процент правильно определенных произведений. Для вариантов получено следующее количество точных определений: а - 73 %, для б – 62 %, для в – 61 %.

На материале варианта г получили худшие результаты – 4 %. В работе рассматривается, как последовательность символов текста не может обладать свойствами простой цепи Маркова. При этом гипотеза, выдвинутая в начале исследования, опровергнута. Тем не менее, анализ результатов экспериментов позволяет сделать вывод, что использование нескольких подряд идущих в тексте букв может дать более точные результаты, чем использование одиночных грамматических классов слов. Потому предполагается, что в буквенных парных структурах не всегда точно отображаются структуры морфем словоформ текста, в частности, префиксальных, корневых, суффиксальных и флективных. Таким образом,

довольно большой объем словообразовательной информации о структуре слов окажется отображенным в статистике парной встречаемости букв. Это, в свою очередь, послужит фактором высокого уровня эффективности использования этой статистики для анализ социокультурного текста [6, С. 127].

Таким образом, необходимо отметить, что существующий программный инструментарий обработки социокультурных текстов методами математической статистики весьма неудовлетворителен, в первую очередь ввиду недостаточного лингвистического обеспечения, а также неразвитости выходных интерфейсов. В существующих программах зачастую отсутствует возможность поиска и статистической обработки совместной повторяемости слов со свободным порядком, который характерен для индивидуального стиля автора.

Для решения выше обозначенных проблем предлагается методика статистического анализа социокультурных текстов с целью количественной оценки степени зависимости слов, которые образуют лексические комплексы, а также позволяющая определять меру их значимости. Разработанный по этой методике программный инструментарий позволяет решать проблемы выявления как очевидных, так и скрытых связей слов.

Для того чтобы оценить, насколько значимыми являются исследуемые комбинации, предлагается использовать статистические коэффициенты MI и t-score. Мера MI (англ. - mutual information) – представлена коэффициентом взаимной информации и используется для сравнения зависимых контекстно-связанных частот с независимыми, к примеру, когда слова в лексическом комплексе появляются в тексте совершенно случайно.

Значение данной меры вычисляется по формуле

$$MI(k, c) = 2 \frac{f(k, c)}{N} \log \frac{f(k, c)}{f(k) f(c)}, \quad (1)$$

где  $k$  – ключевое слово (keyword);

$c$  – коллокат (collocate);

$f(k, c)$  – частота встречаемости ключевого слова  $k$  в паре с коллокатом  $c$ ;

$f(k), f(c)$  – абсолютные (независимые) частоты ключевого слова  $k$  и слова  $c$  в тексте;

$N$  – общее число словоупотреблений в тексте.

Если значение  $MI(k, c)$  больше установленного значения, тогда такое сочетание слов считается статистически значимым. С точки зрения теории вероятности  $MI$  представлена как способ проверки независимости появления слов в тексте. Так, если слова полностью независимы, тогда вероятность их совместного появления равна произведению вероятностей каждого из них.

Мера  $MI$  зависит от размера корпуса. При этом чем больше исследуемый корпус, тем выше получаемые по нему значения  $MI$ .

Это свойство отражает большую степень доверия к данным, которые получают на материале большего корпуса. Такая «дискриминация» по размеру корпуса затрудняет сравнение показателей, полученных на разных корпусах. Другой мерой, которая используется в исследовании, стала мера  $t$ -score. Она учитывает частоту совместной встречаемости ключевого слова и его коллоката и позволяет определить степень неслучайности силы ассоциации (связанности) между коллокатами [8, С. 65].

Используя примеры, взятые из наших полевых исследований, мы рассмотрели два пути, в которых формы восприятия выстроены через статистическую практику с учетом «социальной структуры повседневной деятельности». При этом, можно сделать вывод, что, во-первых, в работе статистиков рассмотрение социокультурных текстов опирается не только на числовые значения. Действительно, при исследовании статистики обнаруживают переплетение культурно-технического мышления с

числовыми данными. Необходимо понимать, что статистики должны работать «этнографически», а также математически не только понять свои данные, но также понимать, когда они используют эти данные в своем анализе.

Во-вторых, работа статистиков требует проявления понимания тех, для кого они намереваются использовать статистические инструменты, которые они производят. Проявление понимания является наиболее мощным фактором, поскольку оно встроено в само производство.

Таким образом, необходимо отметить, что первоочередным условием правильного, научно обоснованного использования методов математической статистики при исследовании социокультурных текстов является соответствие ее средств и аппарата реальным отношениям, существующим в объектах исследования. При этом требуется наличие определенных закономерностей и связей, допускающих математическую обработку и выражение на языке современной математики.

### Список литературы

1. Годин А.Е. Научная традиция как социокультурный феномен: На примере Московской философско-математической школы: диссертация ... кандидата философских наук: 24.00.01 Москва, 2005 216 с.: 61 05-9/601
2. Долгова Н.В. Управление научно-исследовательской деятельностью студентов университета в современных социокультурных условиях. Автореферат на поиск. Ученой степени канд. Пед. Наук. Горно-Алтайск. 2007.
3. Никитенко З.Н., Осиянова О.М. К проблеме выделения культурного компонента в содержании обучения английскому языку в школе // Иностранные языки в школе -1993 -№ 3 - С. 5-10.
4. Никитина Ю.А. Вестник Волжского университета им. В.Н. Татищева. 2009. № 1. С. 210-226.
5. Шапко В.Т. Социология молодежи. Учебник [Текст] / Ю.Р. Вишневский. Екатеринбург: ГОУ ВПО УГТУ-УПИ, 2006.
6. Батура Т.В Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2012. Т. 10. № 4. С. 81-94.
7. Michael Mair, University of Liverpool; Christian Greiffenhagen, Loughborough University; W.W. Sharrock, University of Manchester. Statistical Practice: Putting Society on Display. URL: <http://tcs.sagepub.com/content/early/2015/01/06/>
8. Авхимович В.В., Калинина Е.А., Самойлова Т.А. Системы компьютерной математики и их приложения. 2014. № 15. С. 232-236.



### References

1. Godin A.E. Nauchnaja tradicija kak sociokul'turnyj fenomen: Na primere Moskovskoj filosofsko-matematicheskoy shkoly: dissertacija ... kandidata filosofskih nauk: 24.00.01 Moskva, 2005 216 c.: 61 05-9/601
2. Dolgova N.V. Upravlenie nauchno-issledovatel'skoj dejatel'nost'ju studentov universiteta v sovremennyh sociokul'turnyh uslovijah. Avtoreferat na poisk. Uchenoj stepeni kand. Ped. Nauk. Gorno-Altajsk. 2007.
3. Nikitenko Z.N., Osijanova O.M. K probleme vydelenija kul'turnogo komponenta v sodержanii obuchenija anglijskomu jazyku v shkole // Inostrannye jazyki v shkole -1993 -№ 3 - S. 5-10.
4. Nikitina Ju.A. Vestnik Volzhskogo universiteta im. V.N. Tatishheva. 2009. № 1. S. 210-226.
5. Shapko V.T. Sociologija molodezhi. Uchebnik [Tekst] / Ju.R. Vishnevskij. Ekaterinburg: GOU VPO UGTU-UPI, 2006.
6. Batura T.V Vestnik Novosibirskogo gosudarstvennogo universiteta. Serija: Informacionnye tehnologii. 2012. T. 10. № 4. S. 81-94.
7. Michael Mair, University of Liverpool; Christian Greiffenhagen, Loughborough University; W.W. Sharrock, University of Manchester. Statistical Practice: Putting Society on Display. URL: <http://tcs.sagepub.com/content/early/2015/01/06/>
8. Avhimovich V.V., Kalinina E.A., Samojlova T.A. Sistemy komp'juternoj matematiki i ih prilozhenija. 2014. № 15. S. 232-236.