

## **Исследование текстов на тибетском языке: корпусный подход (на основе 600 публикаций за 40 лет)**

*Дола*<sup>1</sup>

<sup>1</sup> Северо-Западный университет национальностей (д. 1, Шибей Шинхун, 730030, Ланьчжоу, Ганьсу, КНР)  
доктор филологических наук, профессор

© КалмНЦ РАН, 2021

© Дола, 2021

**Аннотация.** *Введение.* Данная статья является результатом исследовательского проекта по созданию корпуса тибетского языка на основе материала 600 тибетологических работ, опубликованных за последние 40 лет. Языковой корпус данной статьи составляют 642 статьи на тибетском языке из 25 тибетских журналов, опубликованных с 1978 по 2018 гг. *Цель.* В данной работе осуществляется классификация слов, деление на части речи, используются статистические данные по литературным памятникам, проводятся измерительные исследования литературных памятников, словарного состава и т. д. *Результаты.* В результате исследования можно выяснить основные характеристики китайской тибетологии за последние 40 лет, главные направления исследований и тенденций развития.

**Ключевые слова:** библиометрия тибетских литературных памятников, научная статья на тибетском языке, измерение лексического состава

**Благодарность.** Данная работа проведена в рамках исследовательского проекта по философии и социальным наукам Министерства образования (номер проекта: 13 ZD028), Национального проекта по общественным наукам (16ZDA166), Национального фонда естественных наук Китая (проект № 61866034), ключевых дисциплин Северо-Западного университета национальностей и создания языковых ресурсов (регистрационный номер инновационного исследования: 1001050733). Материалы статьи апробированы на Международной научной онлайн-конференции

«Монголоведение в начале XXI в.: современное состояние и перспективы развития – II», проведенной при финансовой поддержке РФФИ (проект № 20-09-22004) и частичной поддержке гранта Правительства РФ (№ 075-15-2019-1879).

**Для цитирования:** Дола. Исследование текстов на тибетском языке: корпусный подход (на основе 600 публикаций за 40 лет) // Бюллетень Калмыцкого научного центра РАН. 2021. № 1. С. 102–120. DOI: 10.22162/2587-6503-2021-1-17-102-120

## Exploring of Tibetan-Language Texts: The Corpus Approach (A Case Study of 600 Publications Issued over 40 Years)

*Dola*<sup>1</sup>

<sup>1</sup> Northwest University for Nationalities (1, Xibei Xinxun, Lanzhou, Gansu, 730030, People's Republic of China)  
Dr. Sc. (Philology), Professor

© KalmSC RAS, 2021

© Dola, 2021

**Abstract. Introduction.** This article is the result of a research project to create a corpus of the Tibetan language based on the measurement of the lexical composition of 600 Tibetological works published over the past 40 years. The language corpus of this article consists of 642 articles in Tibetan from 25 Tibetan journals published from 1978 to 2018. **Goals.** This work introduces a classification words, division into parts of speech, uses statistical data based on the literary monuments, measures syllabic and lexical compositions in literary works, vocabulary and syllabic compositions, etc. **Results.** The paper makes it possible to follow the situation in Chinese Tibetology over the past 40 years, the main directions of research and development trends.

**Keywords:** bibliometry of Tibetan literary monuments, scientific article in the Tibetan language, measurement of lexical composition

**Acknowledgements.** This work is the subject of a major research project on philosophy and social sciences of the Ministry of Education of the People's

Republic of China (project no. 13 ZD028), National Social Sciences Project (16ZDA166), National Natural Sciences Foundation of China (project no. 61866034), key disciplines of Northwest University for Nationalities, and the creation of language resources (Innovation research registration no. 1001050733). The article was presented at the International scientific online conference ‘Mongolian Studies at the Beginning of the 21<sup>st</sup> century: Current State and Development Prospects – II’ funded by RFBR (project no. 20-09-22004) and partially granted by Government of Russia (no. 075-15-2019-1879).

**For citation:** Dola. Exploring of Tibetan-Language Texts: The Corpus Approach (A Case Study of 600 Publications Issued over 40 Years). *Bulletin of the Kalmyk Scientific Center of the Russian Academy of Sciences*. 2020. No. 1. Pp. 102–120. DOI: 10.22162/2587-6503-2021-1-17-102-120

## Введение

Библиометрический метод — это использование компьютерной лингвистики, математики и статистики для количественного анализа информации. Это комплексная система знаний, которая объединяет корпусную лингвистику, математику, статистику и источниковедение. В основном она применяется для анализа объема используемой литературы и лексики, количества авторов и т. д. Исследования в области измерения объема используемой литературы подчеркивают важность количественного анализа в области гуманитарных и социальных наук.

За последние 40 лет реформ открытости количество ученых-тибетологов непрерывно росло, научные результаты удвоились, область исследований постоянно расширялась, наблюдается прогресс — как с точки зрения количества, так и качества. Библиометрический метод позволяет провести анализ научных результатов в тибетологии за последние 40 лет, подвести итоги прошлого и заглянуть в будущее, чтобы предоставить справочную информацию о нынешнем состоянии исследований и наметить будущее развитие в данной области. В связи с этим в данной статье проводится количественный анализ более 600 тибетских научных статей за последние 40 лет для того, чтобы проследить динамику иссле-

дований с разных точек зрения. Поскольку тибетское академическое сообщество еще не создало базу данных и поисковую базу данных по литературным индексам CNKI или CSSCI, некоторые важные показатели, такие как количество цитирований и влияние факторов литературы, не могут быть исследованы из-за недостатка данных. Поэтому в дополнение к существующей информации о литературных памятниках эта статья фокусируется на содержании источников и рассматривает его с лексической точки зрения, дополнительно осуществляя анализ слогового состава.

### **1. Исследовательский образец тибетских научных трудов**

В 2016 г. «Инновационная команда по тибетским исследованиям» Северо-Западного университета национальностей начала приводить в порядок основные китайские научные работы по тибетологии с 1978 г, когда существовала потребность в тибетских исследованиях и создании магистерских программ. К октябрю 2018 г. были собраны 10 томов, которые были переданы в издательство «Тибетология Китая». Удачное совпадение, что по случаю 40-й годовщины тибетских исследований в 2018 г. было запланировано провести библиометрический анализ 40-летней истории тибетских исследований в Китае. Это собрание используется в качестве исследовательской выборки для конструирования корпусов, классификации слов<sup>1</sup>, деления на части речи<sup>2</sup>, построения лингвистического корпуса.

Выбор языкового корпуса исследования должен соответствовать научным стандартам, и относительно сбалансированный корпус может лучше отражать объекты, подлежащие исследованию. Все 10 томов — это 25 китайских журналов, официально выпущенных в Китае и за его пределами. В результате интервью,

---

<sup>1</sup> Классификация произведена в соответствии с национальным стандартом «Обработка информации при помощи стандарта сегментации слов тибетского языка» (GB/T 36452–2018).

<sup>2</sup> Помечены в соответствии с национальным стандартом «Обработка информации при помощи критериев классификации частей речи тибетского языка» (GB/T 36337–2018).

консультаций и небольших симпозиумов с соответствующими экспертами были определены принципы отбора документов. В соответствии с ними были отобраны в общей сложности 642 репрезентативных документа из 25 китайских журналов на тибетском языке в качестве образца для исследования корпуса — с периодом в 40 лет с 1978 по 2018 гг. Все 25 журналов являются тибетскими журналами, открыто изданными внутри Китая.

В перечень журналов включены следующие: 1) «Тибетология Китая»; 2) «Тибетские исследования»; 3) «Журнал Тибетского университета»; 4) «Журнал Северо-Западного университета национальностей»; 5) «Восхождение»; 6) «Журнал Цинхайского университета национальностей»; 7) «Исследования тибетского искусства»; 8) «Народное искусство»; 9) «Чжанчаэр»; 10) «Bang Jin Mei Duo»; 11) «Образование Цинхая»; 12) «Культура снежных районов»; 13) «Журнал Цинхайского педагогического университета»; 14) «Тибетское национальное образование»; 15) «Тибетский буддизм»; 16) «Дасай»; 17) «Тибетская литература и искусство»; 18) «Работа на национальном языке»; 19) «Нация»; 20) «Тибетское образование»; 21) «Исследование Андо»; 22) «Журнал Национального педагогического института Ганьсу»; 23) «Мастер»; 24) «Искра знаний»; 25) «Китайский Тибет».

Следует отметить, что со временем некоторые из этих журналов были переименованы, и здесь приводится текущее название, но в корпус включены все научные статьи из журналов, которые изменили свое название.

С точки зрения количества большинство статей представлено из журналов «Тибетология Китая» и «Тибетские исследования» (154 статьи соответственно), что свидетельствует о их влиянии в области тибетологических исследований. За ними следуют «Журнал Тибетского университета» и «Журнал Северо-Западного университета национальностей» (45 статей и 44 статьи соответственно). Следом идут журналы «Восхождение» и «Журнал Цинхайского университета национальностей», из каждого по 37 статей.

С точки зрения классификации контента основными 10 категориями являются история, люди, источники, литература, народные

обычай, культура, религия, искусство, язык и перевод. Распределение 10 категорий представлено ниже:

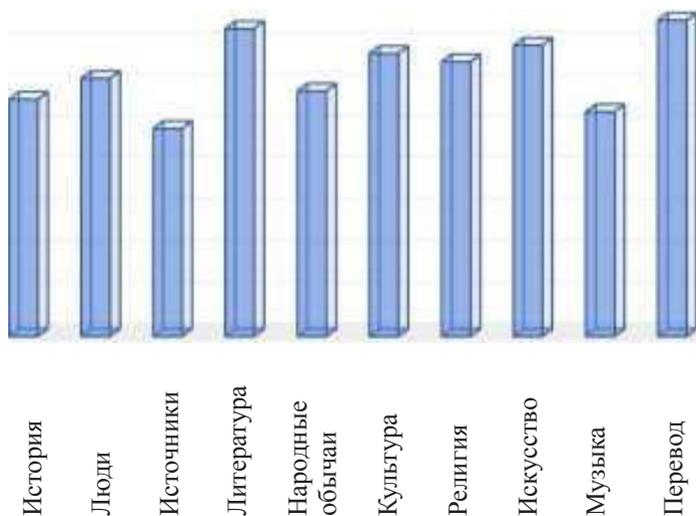


Рис. 1. Классификация контента и количество статей

Согласно выборке статей, основные категории разделены на 51 подкатегорию, то есть 51 конкретную область тибетских исследований, а именно:

1) категория «История»: история, древняя история, исследования периода правления в Тибете Цэнпо, сложные и важные моменты истории Тибета, эпоха раздробленности, краеведческие исследования и т. д.;

2) категория «Люди»: основные исторические личности;

3) категория «Источники»: теория источниковедения, надписи на камне, письменные памятники дуньхуанских пещерных храмов, древние письма и др.;

4) категория «Литература»: теория литературы, исследования классификации литературных произведений, исследования истории литературы, поэзия и теория поэзии, современная поэзия и поэты, исследования рассказов, народная литература;

5) категория «Народные обычаи»: свадебные обряды, погребальные обряды, обряды жертвоприношений, анализ снов, обычаи, связанные с домашними животными, народные верования, народные привычки, народные развлечения и др.;

6) категория «Культура»: топонимы, фамилии, пять отраслей знания (филология, ремёсла и астрономия, медицина, логика, философия), региональная культура, материальная культура, одежда, традиционное естествознание, эпос «Гэсэр» и др.;

7) категория «Религия»: теория буддизма, история буддизма, храмовая культура;

8) категория «Искусство»: тибетская живопись на холсте, ремесленное искусство, искусство гравюры, устное народное искусство, искусство цянов и другие виды исполнительского искусства, комплексные исследования;

9) категория «Язык»: всестороннее изучение языка;

10) категория «Перевод»: общая теория, детализация теории, история перевода, анализ практических вопросов перевода и др. 5 областей. Наличие данной подкатегории также указывает на то, что языковой корпус богат по содержанию и охватывает широкий диапазон, все аспекты тибетских исследований и имеет определенную типичность.

В целом распределение контента относительно равномерно, но особенно активно проводились исследования в области теории буддизма, истории буддизма, тибетской живописи на холсте, литературы, теории перевода, анализа проблем перевода и др., каждая подкатегория представлена более чем 20 статьями.

Сложные и важные моменты истории Тибета, краеведческие исследования, письменные памятники дуньхуанских пещерных храмов, погребальные обряды, практические вопросы перевода, материальная культура и т. д. являются важными областями исследований, каждая подкатегория представлена более чем 15 статьями.

Следует отметить, что категории «Люди» и «Язык» не имеют подкатегорий и представлены 62 статьями и 61 статьей соответственно. Однако в этих областях можно выделить много подка-

тегорий. Например, традиционная грамматика, лингвистика, морфология, синтаксис и т. д., здесь мы не использовали подобную дополнительную классификацию.

Контент категории «Люди» относительно прост — об исторических личностях, классификация исходит из исторического периода, поэтому наблюдается упрощение самого контента, исследование исторических личностей также является популярным.

Если рассматривать с точки зрения количества авторов, то 642 статьи в 10 томах написаны 452 разными авторами, 27 соавторов, всего 479 человек, на главного автора в среднем приходится 1,42 статьи, частота повторения низкая, что указывает на то, что авторский охват относительно высокий. 642 тибетских научных труда, тщательно отобранных, подготовлены 479 людьми, которые занимаются тибетологическими исследованиями, и эта часть авторского состава является основной. Так, корпус охватывает почти всех известных тибетологов в тибетских научных кругах, таких как Цайданьсяжун (Ценден Церинг), Дунгалосанчиле (Дугла Лобсанг), Маоэргайсанмудань, Доши (Гаши), Цябайцыданьпинцо, Сайцанлусанхуадань, Лабапинцо (Лама Цо), Наньканобу (Намка Норбу), Ночжаноуцзян, Чэньцинъян (Тен Церинг), Басанвандуй (Басанг Ванду), Цзаньла Аванцочэн, Мацзинью, Гаожуй (Гачо), Гацантомэй (Гацан Тогме), Каганчжасицайжан, Цягаданьчжэн и др.

Были отобраны их работы, в среднем больше 2 статей на человека, в том числе 8 статей Цябайцыданьпинцо, 6 статей Маоэргайсанмуданя, 5 статей Дунгалосанчиле, по 4 статьи Цайданьсяжуна, Доши, Мацзинью. Конечно, в области тибетских исследований имеется множество работ молодых ученых и ученых средних лет, а также работ молодых докторов тибетологии и докторантов.

Если рассматривать материал с точки зрения временной классификации, 642 статьи с 1978 по 2015 гг. делятся так: 1 статья 1970-х гг., 105 статей 1980-х гг., 145 статей 1990-х гг. и 205 статей 2000-х гг., 10 статей 2010-х гг., что свидетельствует о развитии тибетологии в Китае (см. рис. 2).

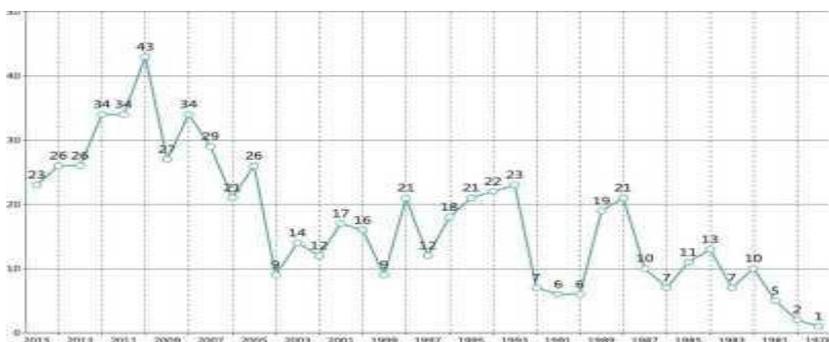


Рис. 2. Год публикации и количество статей

Среди отобранных работ больше всего статей 2010 г. — 43 статьи. Мы также обнаружили период с количеством написания статей более 20, всего 16 временных отрезков, что в определенном смысле отражает качество и влияние опубликованных работ той эпохи, или пиковый период высококачественных научных работ. Данные 16 годов и количество статей представлены на рис. 5:

Говоря о 1980-х гг., следует отметить, что только в один 1988 г. количество статей превысило 20, в 1990-х гг. было уже 4 таких года, в 2000-х гг. — 5, в 2010-х гг. — 6. Увеличение количества статей проходило равномерно.

Исходя из сказанного выше, можно увидеть, что масштаб исследований был достаточно широкий, охватывались огромные области изучаемого объекта, рамки исследований были обширны, присутствовала определенная типичность, между областями исследований присутствовал определенный баланс.

Приведенная выше информация о языковом корпусе письменных документов отражает тенденции и динамику исследований в тибетологии за последние 40 лет реформ открытости.

## 2. Количественный анализ лексического состава

Лексические измерения заключаются в количественном определении различных фактов с использованием таких статистиче-

ских данных, как кратность и частота, совокупная кратность и совокупная частота, а также информационная энтропия. Лексические измерения — это процесс от количественного наблюдения до качественного анализа с последующим выводом на основе позитивизма, при помощи языковой инженерии и статистических данных, при использовании дедуктивного метода.

Количественный метод измерения лексики противоположен традиционному качественному методу, но в совокупности они дополняют друг друга. Лексические измерения все больше играют важную роль в области социальных наук, имеют свои уникальные методы интерпретации в изучении содержания, семантическом анализе, анализе литературного контента, исследовании ключевых слов и так далее.

### **3. Общий анализ 10 томов литературных памятников**

В 642 научных статьях, рассмотренных в данной работе, в общей сложности 162 100 разных слов, которые встречаются 2 824 300 раз. Словарный запас научных работ настолько широк, что в свою очередь демонстрирует тематическое разнообразие типетских исследований.

Согласно общему принципу статистического анализа лингвистического корпуса, слово, встречающееся более 5 раз, является высокочастотным, которое обычно в области изучения языка относится к общеупотребительной лексике. Данная статья содержит 24 302 слова, которые встречаются более 5 раз, что составляет 93,3 % от общего лексического запаса. Количество общеупотребительных слов большое, что отражает широту и разные области исследования.

Было обнаружено 30 436 антропонимов, в общей сложности встречаются 102 501 раз. Среди них 21 809 — это имена, которые встречаются 49 077 раз, 4 306 географических названия — 18 883 раза, 419 названий учреждений — 1 403 раза, 2 630 специальных терминов — 10 073 раза и т. д.

Что касается классификации слов, насчитывается 88 классов слов. Класс слов включает в себя две категории: знаменательные

слова и служебные слова, и он основан на национальном стандарте «Классификация тибетских слов для обработки информации» (GB/T 36337–2018). Среди них существует 45 классов, которые встречаются более чем 10 000 раз, в общей сложности повторяющиеся 27 255 212 раз, что составляет 96,5 % от общего количества слов.

Классы слов и отдельных граммем, встречающиеся более чем 10 000 раз, и их частотность представлены в табл. 1.

*Таблица 1. Частотность классов слов и некоторых граммем*

№	Часть речи	Обозначение	Кратность обнаружения
1	Простое имя существительное	nn	692 222
2	Родительный падеж	gi	306 417
3	Переходный глагол	vt	231 583
4	Творительный падеж	is	142 819
5	Указательное местоимение	rz	116 621
6	Непереходный глагол	vi	103 295
7	Соединительный союз	cn	70 495
8	Простое имя прилагательное	as	70 360
9	Существительные темпоративы	nt	66 825
10	Соединительно-разъединительные союзы	cd	64 664
11	Собственное имя человека	nr	49 077
12	Вспомогательный глагол	vu	46 068
13	Активный залог	bo	45 401
14	Существительные локативы	nf	43 568
15	Глагол-связка	vj	43 454
16	Обобщающий падеж 同体格	ld	42 595
17	Имя числительное	mj	41 473
18	Начальный падеж	jg	32 379
19	Союз ссылки	ci	31 207
20	Семантически связанный союз	fg	29 691

21	Отглагольное существительное	nv	29 265
22	Ссылочный союз 具指连词	ce	29 093
23	Отрицательное наречие	df	29 052
№	Часть речи	Обозначение	Частота обнаружения
24	Экзистенциальный глагол	VC	26 962
25	Собирательные счетные слова	qj	25 550
26	Служебное слово сравнения	ub	24 080
27	Наречие предела	dw	22 447
28	Наречие степени	dc	21 427
29	Предикативное прилагательное	ad	20 147
30	Собирательное имя существительное	ne	19 033
31	Топоним	ns	18 483
32	Личное местоимение	rr	18 296
33	Приведенный союз	cv	18 085
34	Комплексный союз	cu	17 320
35	Зависимый соединительный союз	eg	16 182
36	Противительный союз	cz	13 320
37	Временной падеж	lh	12 813
38	Причинный падеж	bg	12 750
39	Видо-временное вспомогательное слово	us	12 184
40	Вопросительное местоимение	ry	11 819
41	Название должностей	nx	11 798
42	Вспомогательное слово, которое встречается в конце предложения	ud	11 780
43	Конечное вспомогательное слово	uz	11 684
44	Порядковое числительное	mx	11 655
45	Собственное имя существительное	nz	10 073
Итого		2 725 512	

Полученное количество данных достаточно большое, и нельзя считать, что все аспекты данных отражены. Поэтому в данной статье мы изучим первые 50 существительных, чтобы взглянуть на фокус тибетских исследований за последние 40 лет. Например: тибетская нация, история, человек, письменность, регион, нация, культура, характеристика, закон, речь, ученый, искусство, предложение, литература, язык, общество, тибетский язык, поэзия, понятие, живопись, натура, древние книги, дух, жизнь, Индия, обычай, литературный памятник, бон, язык, теория, славословие, статуя или тело, содержание, религия, мысль, пример, биография, Дуньхуан, тибетский язык, мощь, книга, тема, учение, персонаж, мир, перевод, обычай, знание, политика, Будда и др.

Среди них слово «тибетская нация» встретилось более 12 000 раз, «история» — более 6 000 раз, «человек» — более 4 000 раз, «письменность», «регион», «нация», «культура», «характеристика» и т. д. — 3 000 раз, остальные — от 1 000 до 3 000 раз.

Эти 50 имен существительных, в общем, появились 104 189 раз, что составляет 3,69 % от общего словарного запаса, это отражает то, что «тибетская нация», «культура», «регион», «письменность», «искусство» и так далее являются главными темами исследований. Высокая кратность и частота употребления 50 существительных представлена в табл. 2.

Таблица 2. Частота употребления высокочастотных существительных

№	Существительные в переводе с тибетского языка на русский язык	Частота	Частотность	Итого, частота	Итого, частотность
1	Тибетская нация	12 671	0.44863 %	12 671	0.44863 %
2	История 历史	6 252	0.22136 %	18 923	0.66998 %
3	Человек 人	4 289	0.15186 %	23 212	0.82184 %
4	Письменность 文字	3 801	0.13458 %	27 013	0.95642 %
5	Регион	3 589	0.12707 %	30 602	1.08349 %
6	Нация	3 492	0.12364 %	34 094	1.20712 %

7	Культура	3 393	0.12013 %	37 487	1.32726 %
8	Характеристика	3 086	0.10926 %	40 573	1.43652 %
9	Закон	2 904	0.10282 %	43 477	1.53934 %
10	Речь	2 904	0.10282 %	46 381	1.64216 %
11	Ученый	2 360	0.08356 %	48741	1.72571 %
12	Искусство 艺术	2 238	0.07924 %	50 979	1.80495 %
13	Предложение	2 142	0.07584 %	53 121	1.88079 %
14	Литература	2 066	0.07315 %	55 187	1.95394 %
15	Язык	1 996	0.07067 %	57 183	2.02461 %
16	Общество	1 924	0.06812 %	59 107	2.09273 %
17	Тибетский язык	1 871	0.06624 %	60 978	2.15897 %
18	Поэзия	1 786	0.06323 %	62 764	2.22221 %
19	Понятие	1 750	0.06196 %	64 514	2.28417 %
20	Живопись	1 731	0.06129 %	66 245	2.34546 %
21	Натура	1 589	0.05626 %	67 834	2.40172 %
22	Древние книги	1 550	0.05488 %	69 384	2.45659 %
23	Дух	1 548	0.05481 %	70 932	2.51140 %
24	Жизнь	1 516	0.05368 %	72 448	2.56508 %
25	Индия	1 514	0.05360 %	73 962	2.61868 %
26	Обычай	1 447	0.05123 %	75 409	2.66991 %
27	Литературный памятник	1 437	0.05088 %	76 846	2.72079 %
28	Бон	1 400	0.04957 %	78 246	2.77036 %
29	Язык	1 345	0.04762 %	79 591	2.81798 %
30	Теория	1 335	0.04727 %	80 926	2.86525 %
31	Славословие	1 335	0.04727 %	82 261	2.91251 %
32	Статуя, тело	1333	0.04720 %	83 594	2.95971 %
33	Содержание	1 316	0.04659 %	84 910	3.00630 %
34	Религия	1 300	0.04603 %	86 210	3.05233 %
35	Мысль	1 272	0.04504 %	87 482	3.09737 %
36	Пример	1 268	0.04489 %	88 750	3.14226 %
37	Биография	1 200	0.04249 %	89 950	3.18475 %
38	г. Дуньхуан	1 176	0.04164 %	91 126	3.22639 %
39	Тибетский язык	1 172	0.04150 %	92 298	3.26788 %

40	Мощь	1 169	0.04139 %	93 467	3.30927 %
41	Книга	1 128	0.03994 %	94 595	3.34921 %
42	Тема	1 096	0.03880 %	95 691	3.38801 %
43	Учение	1 080	0.03824 %	96 771	3.42625 %
44	Персонаж	1 069	0.03785 %	97 840	3.46410 %
45	Мир	1 067	0.03778 %	98 907	3.50188 %
46	Перевод	1 066	0.03774 %	99 973	3.53962 %
47	Обычай	1 065	0.03771 %	101 038	3.57733 %
48	Знание	1 063	0.03764 %	102 101	3.61497 %
49	Политика	1 051	0.03721 %	103 152	3.65218 %
50	Будда	1 037	0.03672 %	104 189	3.68889 %

Мы также исследовали 500 имен существительных, самая высокая кратность среди которых составляла 12 671 раз, а самая низкая кратность — 191. Они встречались 355 500 раз, что составляет 12,59 % от общего словарного запаса.

#### **4. Анализ классификации 10 томов литературных памятников**

Как упоминалось выше, 642 статьи разделены на 10 категорий, общее количество слов в 10 категориях составляет 293 900, а количество упоминаний — 2 824 300 слов, что эквивалентно в среднем 9,61 разу на слово.

В 10 категориях было 31 215 слов, касающихся литературных памятников, которые встречаются 266 855 раз; 31 071 слово из культуры — 265 062 раза; 27 447 слов из литературы — 312 525 раз; 34 357 слов, касающихся персонажа, — 291 722 раза; 31 724 слова из искусства — 304 211 раз; 31 759 слов из истории — 301 452 раза; 24 491 лингвистический термин — 251 705 раз; 27 516 слов из обычаев — 247 022 раза; 30 229 религиозных слов — 304 107 раз; 24 091 слов из перевода — 279 666 раз.

Обзор выглядит следующим образом: литературные памятники, культура, литература, персонаж, искусство, история, язык, обычай, религия, перевод.

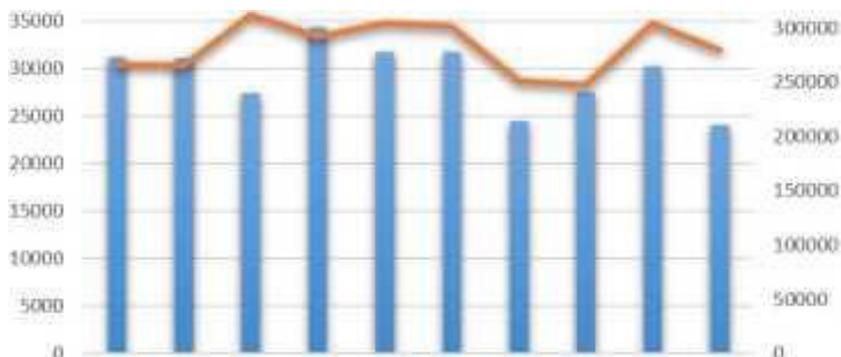


Рис. 3. Классификация слов и их частота

В соответствии с классификацией 10 основных категорий литературных памятников, первые 20 высокочастотных существительных также являются ключевыми для анализа слов в литературе. Как установлено в ходе комплексного исследования, ключевые слова существительных в каждой категории не сильно меняются, но высокочастотные существительные из категории «литературные памятники» имеют различный акцент, и тенденция отнесения их к этой категории более очевидна, например, слова из категории «литературные памятники»: литературный памятник, текст, каталог; категория «культура»: культура, древние книги, нация, общество и т. д. Категория «литература»: литература, поэзия, искусство, автор, мысль, жизнь и т. д. Все эти слова обладают очевидными лексическими особенностями отнесения их к той или иной категории. Интересно, что среди 10 категорий термин «тибетская нация» занимает первое место. Это отражает то, что основным объектом или темой тибетских исследований является «тибетская нация».

Первые 20 высокочастотных существительных из 10 категорий показаны в табл. 4.

Таблица 4. Высокочастотные существительные в зависимости от тематических категорий

№	Литературный памятник	Культура	Литература	Персонаж	Искусство	История	Язык	Обычай	Религия	Перевод
1	Тибетская нация	Тибетская нация	Литература	Тибет	Тибет	Тибетская нация	Письменность	Тибетская нация	Закон	Значение
2	(год)	Культура	Поэзия	История	Живопись	История	Речь	Год	Тибетская нация	Тибетская нация
3	Письменность	Наименование	Тибетская нация	Наша эра	Искусство	Год	Наименование	Дух	Логика	Наименование
4	История	Год	Искусство	Закон	Особенность	Цэнпо	Тибетская нация	Обычай	Регион	Перевод
5	Наша эра	История	История	История	История	Король	Семантика	Культура	Понятие	Подлинник
6	Значение	Человек	Особенность	Ученый	Джянди (музыка)	Наименование	Язык	Церемония	Знание	Предложение
7	Регион	Регион	Нация	Король	Человек	Наша эра	Предложение	Жизнь	Форма	Культура
8	Текст	Нация	Автор	Регион	Форма	Нация	Каменная стела	Эпоха	Структура	Нация
9	Буддийский закон	Бон	Форма	Тело, статуя	Нация	Язык	Дунхуань	Женщина	Точка зрения	Тибетский язык

10	Человек	Общество	Мысль	Нация	Культура	Общество	Послелог	Лу	Натура	Концепция
11	Литературный памятник	Древние книги	Жизнь	История	Эпоха	Дунхуань	Тибетский язык	Год	глубоко-мудрый	Диалект
12	Наименование	Возраст	Язык	Человек	Статуя	Письменность	Характеристики	Общество	Древние книги	Различие
13	Тибетский язык	Особенность	Тема	Духовный учитель	Король	Гробница	Год	Религия	Храм	Тибетский язык
14	Каталог	Ученый	Предложение	Монах-переводчик	Произведение	Мощь	Славословие	Король	Наша эра	Характеристика
15	Соглашение	Наша эра	Общество	Индия	Дух	Литературный памятник	Регион	Сон	Теория	Язык
16	Характеристика	Отношение	Культура	Цэнпо	Танец	Политика	Тибетский язык	Сознание	Человек	Речь
17	Нация	Цивилизация	Поэзия	Летопись	Религия	Слово	Звук	Форма	Предопределение	Китайский язык
18	Издательство	Эпоха	Произведение	Вежливая форма обращения	Традиция	Отношение	Основное слово	Первобытный	Бон	Письменность
19	Политика	Цэнпо	История	Амдо	Индия	Бон	Грамматика	Сноха	Сущность	Перевод
20	Эпоха	Религия	Образ	Литературный памятник	Народ	Ученый	История	Гробница	Будда	Переводчик

### **Заключение**

Тибетология стала ведущей отраслью международной науки. За 40 лет реформ китайская тибетология вступила в новую стадию развития. Масштаб, широта и глубина исследований значительно возросли, количество исследовательских коллективов также стремительно увеличилось. Тибетологи обращаются к исследованиям на китайском языке внутри страны, а также и к исследованиям зарубежных ученых.

### **Литература**

- Dola 2010 — *Dola*. [Cognitive similarities and differences of Chinese and Tibetan language and the appropriate attitude]. Xiwei University Journal (Social Science Edition), 2010. 25 (2). Pp. 82–85. (In Chin.)
- Dola 2015 — *Dola* (ed. Tashigi). [Dictionary of standard syllable frequency of the Tibetan language]. China Social Sciences Press, 2015. (In Chin.)