Опубликовано: Общая и прикладная ценология как приятие понимания фундаментальности природного закона видового разнообразия особей сообществ третьей научной картины мира материальной и идеальной реальностей. Практические исследования. Обобщающие материалы по общей и прикладной ценологии. Труды XXII встречи-семинара ценологов (Москва, НИУ МЭИ, 16.11.2018) / Ценологические исследования. Вып. 59. СПб, КСИ-ПРИНТ, 2019. С. 108-121.

усечённые ципфоподобные распределения И ЧАСТОТНЫЕ СЛОВАРИ КОНЦЕПТОВ

Линь Цзиньфэн, С.Л. Пущин, Т.Г. Петров, Д.М. Семёнова, С.В. Чебанов

[*linjinfeng1990@163.com*](mailto:linjinfeng1990@163.com) , [*z1q813@mail.ru*](https://e.mail.ru/compose/?mailto=mailto%3az1q813@mail.ru) , *tomas\_petrov@rambler.ru , dasha.glc@gmail.com ,* [*s.chebanov@gmail.com*](mailto:s.chebanov@gmail.com)

Выявлен своеобразный тип распределений с толстыми хвостами, которые обрублены, описывающий распределение в тексте групп тематически близких концептов. При этом магическое число Миллера 7±2 концептов покрывает 80% встречаемости этих концептов в тексте, а пойнтер-точка распределений токенов (лексем или других средств выражения), выражающих этот набор концептов, является зоной концентрации лексем, наиболее полно представляющих этот набор.

Ключевые слова: концепт, квантитативная концептология, распределение концептов, распределения с толстыми хвостами, обрубленные толстые хвосты, пойнтер-точка

New peculiar type of distributions with thick tails, which are chopped off, describing the distribution in the text of groups of thematically similar concepts is revealed. At the same time, the Miller magic number of 7 ± 2 concepts covers 80% of the occurrence of these concepts in the text, and the pointer point of distribution of the tokens (lexemes, other means of expression) expressing of this set of concepts is the area of ​​concentration of the tokens most fully representing this set.

Keywords: concept, quantitative conceptology, distribution of concepts, distribution with thick tails, chopped off thick tails, pointer-point

Уже более трети века, как в центре интереса лингвистики и смежных дисциплин (культурологии, этнологии, когнитивистики и т.д.) находятся концепты языка, противопоставляемые понятиям логики. Особенно популярным оказалось изучение концептов в постсоветской лингвистике.

Выдающиеся отечественные лингвисты дают следующие трактовки концепта:

Концепт – "это как бы сгусток культуры в сознании человека; то, в виде чего культура входит в ментальный мир человека. …концепт – это то, посредством чего человек … входит в культуру, а в некоторых случаях и влияет на неё" (Степанов, 2004, с. 43).

Концепт – "индивидуальные представления, которым в некоторых чертах и признаках даётся общая значимость. Концепт есть мысленное образование, которое замещает нам в процессе мысли неопределённое множество предметов одного и того же рода. Концепт есть образование ума" (Аскольдов, 1997, с. 269).

"Рассматривая, как воспринимаются слово, значение и концепт, мы не должны исключать человека ... , потенции концепта тем шире и богаче, чем шире и богаче культурный опыт человека ..., и чем меньше культурный опыт человека, тем беднее не только его язык, но и его «концептосфера»" (Лихачёв, 1993, с. 320).

"Концепт … отражает не **просто существенные признаки <понятия>, а все те признаки, которые в данном языковом** коллективе заполняют знание сущности. Следовательно, термин "концепт" обладает более объёмным содержанием по сравнению с «понятием»" (Пименова, 2006, с. 49–50).

Выделяются концепты на основании комплексного лексико-семантико-психолого-историко-культурологического анализа, опирающегося в конечном счёте в большой мере на экспертные оценки исследователя.

Между словами, понятиями и концептами существуют следующие соотношения:

1 слово – N концептов (м.б. N=0,1…) – M понятий (м.б. М=0)

1 концепт – L слов (м.б. L=0) – P понятий (м.б. Р=0)

1 понятие – К слов (К=1 – идеальный термин) – S концептов (часто S=1)

Проиллюстрировать это можно на примере слова русского языка *ягода*, ботанического понятия ягода, и русского концепта [ЯгодА] (прописные буквы в квадратных скобках – стандартный способ обозначения концептов - Табл. 1):

Табл. 1. Соотношение слова, понятия и концепта.

|  |  |  |
| --- | --- | --- |
| понятие | слово | концепт |
| Многосемянный плод с кожистым [эпикарпием](https://ru.wikipedia.org/wiki/%D0%AD%D0%BF%D0%B8%D0%BA%D0%B0%D1%80%D0%BF%D0%B8%D0%B9) и сочными [мезокарпием](https://ru.wikipedia.org/w/index.php?title=%D0%9C%D0%B5%D0%B7%D0%BE%D0%BA%D0%B0%D1%80%D0%BF%D0%B8%D0%B9&action=edit&redlink=1) и [эндокарпием](https://ru.wikipedia.org/wiki/%D0%AD%D0%BD%D0%B4%D0%BE%D0%BA%D0%B0%D1%80%D0%BF%D0%B8%D0%B9) | Ягода | Мелкие сладкие плоды |
| Ягода | Крыжовник, Смородина, Черника, Брусника | [Ягода] |
| Сложная костянка | Малина, Ежевика, Костяника | [Ягода] |
| Сложная ложная костянка | Клубника, Земляника | [Ягода] |
| Ложная ягода | Шиповник, роза | [Плод] |
| Костянка | Вишня, Черешня | [Ягода] |
| Яблоко | Рябина, Боярышник | [Ягода] |
| Ягода | Помидор, Огурец, Баклажан | [Овощ] |
| Ягода | Киви, Банан | [Фрукт] |
| Ягода | Арбуз, Дыня, Тыква | [Бахчевые] |

Концепты слагают языковую картину мира (ЯКМ). Ядром ЯКМ является пословичная картина мира (ПКМ), концепты которой выступают как константы (Степанов, 2004), свойственные той или иной языковой культуре, неменяющиеся в обозримый период истории. В связи с этим изучение ПКМ вызывает пристальный интерес.

Однако, концепты, ЯКМ и ПКМ изучаются пока только качественно, в то время как квантитативной концептологии практически не существует. Поэтому исключительный интерес представляет их количественное изучение.

Такое исследование было проведено по эталонным корпусам пословиц, в качестве которых выступили для русских пословиц собрание В. И. Даля «Пословицы русского народа» (1861-1862), а для китайских – 中国谚语资料 – Собрание китайских пословиц Ланчжоуского института искусств (1961, 1962). В этих собраниях были выделены пословицы, в которых фигурируют реалии, соотносимые с концептами [ТЕЛО], [ДУША], [ДУХ] (табл. 2).

Табл. 2. Число пословиц с концептами [ТЕЛО], [ДУША], [ДУХ]

|  |
| --- |
| В Собрании Даля |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Пословиц с концептом | [ТЕЛО] | [ДУША] | [ДУХ] | Всего | | Абс. кол-во | 7611 | 5935 | 1500 | 11584\* | | % от общего кол-ва в собрании (31351) | 24,28 | 18,93 | 4,78 | 36,95 | | % от общего кол-ва с концептами (11584) | 65,70 | 51,23 | 12,95 | 100,00 |   В Собрании китайских пословиц |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Пословиц с концептом | [ТЕЛО] | [ДУША] | [ДУХ] | Всего | | Абс. кол-во | 5147 | 3080 | 1020 | 7252 | | % от общего кол-ва в собрании (32576) | 15,80 | 9,45 | 3,13 | 22,26 | | % от общего кол-ва с концептами (7252) | 70,97 | 42,47 | 14,07 | 100,00 | |

\*7611+5935+1500>11584; 24,28+18,93+4,78>36,95; 65,70+51,23+12,95>100,00 и аналогично в китайских из-за того, что одна пословица может содержать более одного концепта.

Если для каждого из языков взять лексику, представляющую все три концепта (включая и описания их частей), то (после исключения первых наиболее высокочастотных классов – ср. Фуфаев, 2009) получается обычное ранговое распределение гиперболического вида (рис. 1) или в полулогарифмических координатах (без исключения первых классов, рис. 2). Во втором случае имеется практически линейная зависимость, нарушаемая на первых рангах сверхвысокочастотными классами (исключенными из графика рис. 1) и ступеньками в области больших рангов за счет встречаемости нескольких или многих равночастотных классов.

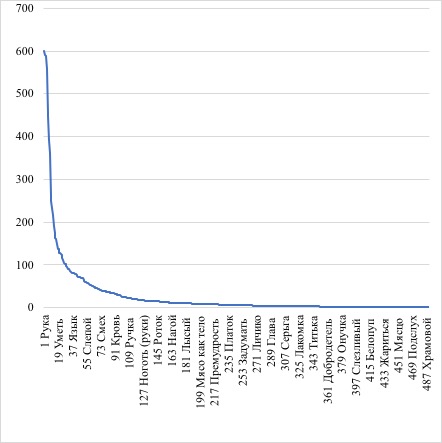


Рис. 1. Частотно-алфавитное распределение лексем описания составности человека в русских пословицах (за исключением первых четырёх наиболее частотных)

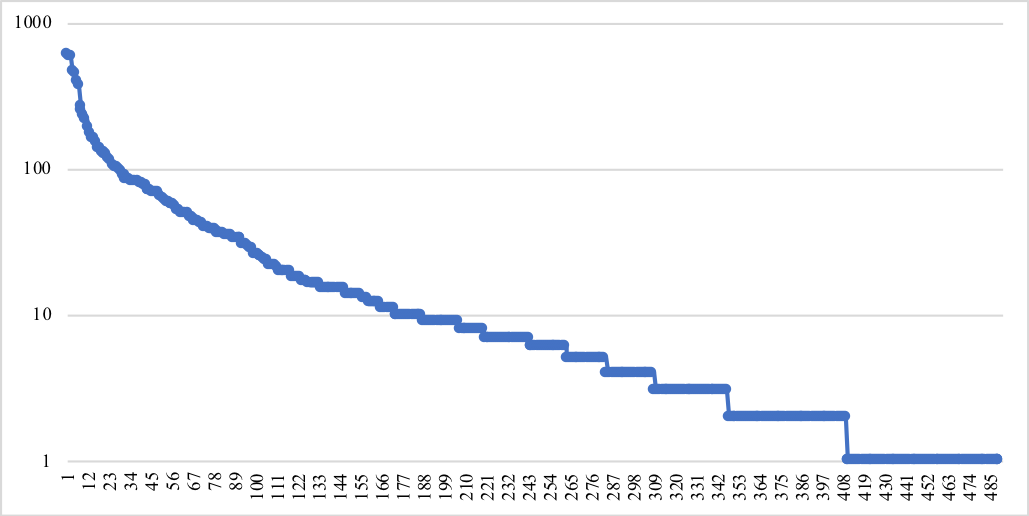


Рис. 2. Частотно-алфавитное распределение лексем описания составности человека в русских пословицах (в полулогарифических координатах)

На основе анализа семантики токенов[[1]](#footnote-1) (лексем, словосочетаний, описательных конструкций, косвенных наименований и т.д.) и содержащих их пословиц были выделены концепты 1-го уровня, передающие составность человека в ПКМ и вычислены их частоты как суммы частот, передающих их токенов (Табл. 3).

Далее на основании отношения "часть-целое" были выделены концепты 2-го и последующих уровней (всего для обоих языков 5, хотя условие равенства при проведении исследования изначально не задавалось). Их частоты вычислялись как суммы частот, входящих в них концептов более низкого уровня (Табл. 4).

При этом оказывается, что распределения частот концептов, передающие составность человека, резко неравночисленные и в ранговой настолько резко убывающие, так, что их удобно представлять в полулогарифмических координатах (рис. 3).

Характер этих распределений оказывается однотипным для распределений частот концептов 1÷5 уровней русских и китайских пословиц (рис. 4). При этом, чем выше уровень концептов, тем их меньше и тем круче оказывается падение их частот.

В связи с таким резким падением частот концептов была изучена динамика накопления их частот. При этом также обнаружилась совершенно однотипная картина (рис. 5; концепты четвёртого уровня выбраны в связи с тем, что их число невелико, а картина динамики совершенно такая же, как и для концептов других уровней русских и китайских пословиц).

Таблица 3. Концепты русских и китайских пословиц (первый уровень)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| R | Русский | Кол-во | Доля% |  | R | Китайский | Кол-во | Доля% |
| 1 | [ДУША (неспец.)] | 5152 | 19,82 | 1 | [РОТ] | 1377 | 13,96 |
| 2 | [ТЕЛО] | 4058 | 15,61 | 2 | [СЕРДЦЕ] | 1318 | 13,36 |
| 3 | [ГОЛОВНАЯ ДУША] | 3235 | 12,44 | 3 | [ГЛАЗ] | 825 | 8,36 |
| 4 | [ДУХ] | 2356 | 9,06 | 4 | [СЕРДЕЧНАЯ ДУША] | 732 | 7,42 |
| 5 | [РОТ] | 2028 | 7,80 | 5 | [ЖИВОТ] | 570 | 5,78 |
| 6 | [СЕРДЕЧНАЯ ДУША] | 1797 | 6,91 | 6 | [КИСТЬ] | 558 | 5,66 |
| 7 | [ГЛАЗ] | 1247 | 4,80 | 7 | [ТЕЛО] | 503 | 5,10 |
| 8 | [ГОЛОВА (неспец.)] | 851 | 3,27 | 8 | [ГУБА] | 410 | 4,16 |
| 9 | [ТЕЛО]+[ДУША]+[ДУХ] | 715 | 2,75 | 9 | [СТОПА] | 406 | 4,12 |
| 10 | [<КИСТЬ>] | 601 | 2,31 | 10 | [ГОЛОВА (неспец.)] | 405 | 4,11 |
| 11 | [ЛИЦО (неспец.)] | 315 | 1,21 | 11 | [ЛИЦО (неспец.)] | 314 | 3,18 |
| 12 | [УХО] | 306 | 1,18 | 12 | [ДУХ] | 295 | 2,99 |
| 13 | [ДУША]+[ДУХ] | 300 | 1,15 | 13 | [ГОЛОВНАЯ ДУША] | 286 | 2,90 |
| 14 | [<СТОПА>] | 264 | 1,02 | 14 | [УХО] | 255 | 2,59 |
| 15 | [НОГА] | 216 | 0,83 | 15 | [НОГА] | 196 | 1,99 |
| 16 | [РУКА] | 193 | 0,74 | 16 | [МЕРТВЕЦ] | 190 | 1,93 |
| 17 | [ЖЕЛУДОК] | 190 | 0,73 | 17 | [РУКА] | 173 | 1,75 |
| 18 | [ЖИВОТ] | 164 | 0,63 | 18 | [ПОЯСНИЦА] | 104 | 1,05 |
| 19 | [ЗУБ] | 156 | 0,60 | 19 | [ЯЗЫК] | 97 | 0,98 |
| 20 | [ГОРЛО] | 141 | 0,54 | 20 | [КОЖА] | 87 | 0,88 |
| 21 | [СПИНА] | 141 | 0,54 | 21 | [ПЕЧЁНОЧНАЯ ДУША] | 84 | 0,85 |
| 22 | [НОС] | 138 | 0,53 | 22 | [ЗУБ] | 76 | 0,77 |
| 23 | [СЕРДЦЕ] | 127 | 0,49 | 23 | [ГОРЛО] | 65 | 0,66 |
| 24 | [ПОКОЙНИК] | 126 | 0,48 | 24 | [СПИНА] | 61 | 0,62 |
| 25 | [ПАЛЕЦ] | 99 | 0,38 | 25 | [ЖЁЛЧНЫЙ ПУЗЫРЬ] | 59 | 0,60 |
| 26 | [ПЛЕЧО] | 93 | 0,36 | 26 | [КОСТЬ] | 55 | 0,56 |
| 27 | [ГУБА] | 90 | 0,35 | 27 | [НОС] | 54 | 0,55 |
| 28 | [БОК] | 86 | 0,33 | 28 | [КУЛАК] | 52 | 0,53 |
| 29 | [КОЖА] | 86 | 0,33 | 29 | [СЛЕЗА] | 47 | 0,48 |
| 30 | [СЛЕЗА] | 86 | 0,33 | 30 | [БРОВЬ] | 44 | 0,45 |
| 31 | [ЯЗЫК] | 84 | 0,32 | 31 | [ПЛЕЧО] | 38 | 0,39 |
| 32 | [ГРУДЬ] | 71 | 0,27 | 32 | [ГРУДЬ] | 32 | 0,32 |
| 33 | [КОСТЬ] | 70 | 0,27 | 33 | [ПЕЧЕНЬ] | 32 | 0,32 |
| 34 | [КУЛАК] | 65 | 0,25 | 34 | [КРОВЬ] | 30 | 0,30 |
| 35 | [НОГОТЬ] | 45 | 0,17 | 35 | [ДУША (неспец.)] | 12 | 0,12 |
| 36 | [ЛОБ] | 43 | 0,17 | 36 | [ЩЕКА] | 12 | 0,12 |
| 37 | [ГОЛЫЙ] | 42 | 0,16 | 37 | [ЛЁГКОЕ] | 9 | 0,09 |
| 38 | [КРОВЬ] | 33 | 0,13 |  |  |  |  |
| 39 | [ЩЕКА] | 26 | 0,10 |  |  |  |  |
| 40 | [БРОВЬ] | 24 | 0,09 |  |  |  |  |
| 41 | [ЛОКОТЬ] | 24 | 0,09 |  |  |  |  |
| 42 | [ПЕЧЁНОЧНАЯ ДУША] | 20 | 0,08 |  |  |  |  |
| 43 | [ПЯТКА] | 20 | 0,08 |  |  |  |  |
| 44 | [КУКИШ] | 18 | 0,07 |  |  |  |  |
| 45 | [ТИТЬКА] | 15 | 0,06 |  |  |  |  |
| 46 | [КОЛЕНО] | 14 | 0,05 |  |  |  |  |
| 47 | [ПУП] | 11 | 0,04 |  |  |  |  |
| 48 | [ПЕЧЕНЬ] | 8 | 0,03 |  |  |  |  |
| 49 | [<МАТКА>] | 7 | 0,03 |  |  |  |  |

Табл. 4. Число концептов разных уровней и их число, приходящееся на 80% покрытия

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень концептов | Русские пословицы | | Китайские пословицы | |
| количество | 80% покрытия | количество | 80% покрытия |
| 1 | 49 | 8 | 37 | 12 |
| 2 | 36 | 8 | 27 | 9 |
| 3 | 24 | 6 | 21 | 7 |
| 4 | 18 | 6 | 15 | 5 |
| 5 | 5 | 2 | 3 | 1 |

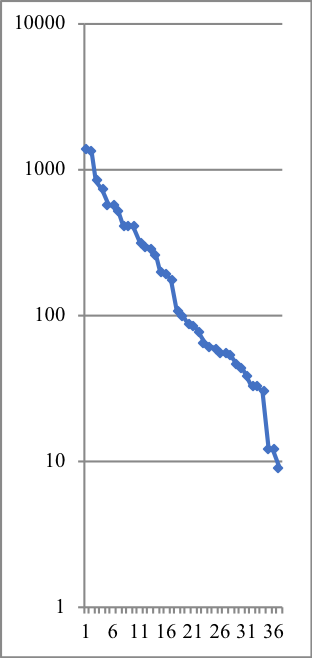
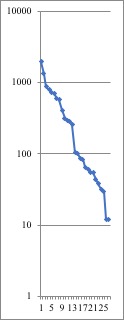
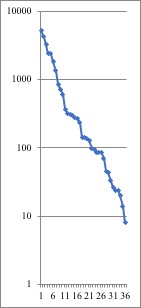
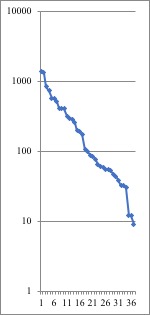
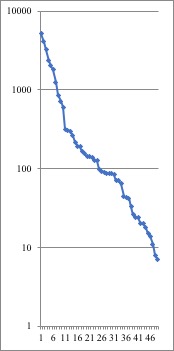
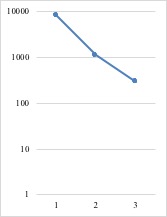
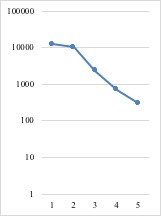
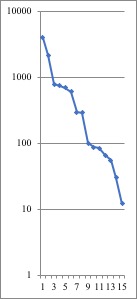
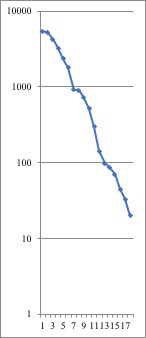
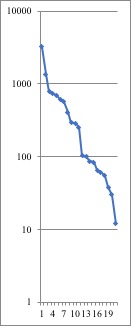
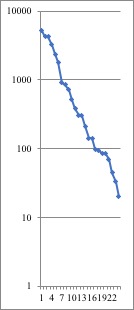


Рис. 3. Частоты концептов, передающих составность человека, 1-го уровня в китайских пословицах (в полулогарифмических координатах)



1р 1к 2р 2к



3р 3к 4р 4к 5р 5к

Рис. 4. Распределение частот концептов, передающих составность человека, 1÷5-го уровней в русских (Р) и китайских (К) пословицах (в полулогарифмических координатах; масштаб по оси абсцисс точно не выдержан из-за несоизмеримости числа концептов разных уровней)

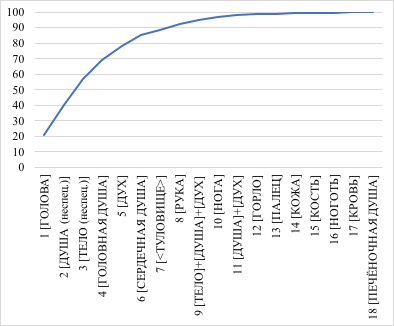


Рис. 5. Накопленные частоты концептов четвёртого уровня русских пословиц

Примечательно и то, что соотношение Парето 80:20, выступающее в данном случае как 80% покрытия частот употребления всех концептов, достигается для концептов 1÷4 уровней русских и китайских пословиц (для 5 уровня число концептов так мало, что данное отношение теряет смысл) за счёт 5–9 концептов (при одном исключении для китайского – 12, что может быть связано с особенностями строя китайского языка) и слабо зависит от общего числа концептов (15÷49; табл. 4). Это в точности соответствует магическому числу Миллера 7±2 компонентов, которыми оптимально манипулирует оперативная память (Миллер, 2010).

Другим направлением квантитативной концептологии может быть изучение не распределений концептов, а распределение частот токенов, с помощью которых выражаются концепты.

Один из вариантов таких исследований был намечен Б. И. Кудриным в связи с его представлением о пойнтер-точке (Кудрин, 2007, с. 25–33). Исследование Е. Б. Кудриной показывает, что при изучении частоты упоминаний героев в романе М. А. Булгакова "Мастер и Маргарита" к пойнтер-точке ***R***=34 примыкают Левий Матвей, Гелла, Н. И. Босой, Варенуха, Римский, Стёпа Лиходеев и Га-Ноцри, которые и отражают булгаковскую специфику повествования и его отличие от "Фауста" Гёте (Кудрин, 2007, с. 31). В связи с этим было предпринят поиск пойнтер-точки для распределения токенов, описывающих составность человека в русских и китайских пословицах. Для этого был использован способ расчёта пойнтер точки по методике С. Л. Пущина (Пущин, 2014, с. 21–28).

В результате было обнаружено, что пойнтер-точка для распределения русских лексем соответствует частоте 186 (ближайшие лексемы Воля – 190 и Сердце – 175), а для китайских – 92 (ближайшие лексемы 听Слышать – 96, 腿Нога от пояса до стопы – 87). Среди русских лексем вблизи пойнтер-точки оказываются лексемы *добро, видеться, добрый, воля, сердце, чёрт, грех*, характеризующие [СЕРДЕЧНУЮ ДУШУ], [ГОЛОВНУЮ ДУШУ], [ДУХ], а через *видеть* и [ТЕЛО], что вполне соответствует образу русского человека как живущего душой и духом. Среди лексем китайских пословиц вблизи пойнтер-точки находятся *饱 сыт, 死 умереть, 饥 голод, 讲 (книжн.) говорить, 食еда (книжн.), 腰 поясница*, которые передают потребность в еде как важнейшую потребность [ТЕЛА], отмечая и такую важную для китайской культуры часть [ТЕЛА] как [ПОЯСНИЦА] (которая присутствует на 3-х из 5-и уровней иерархии концептов) и включая только две лексемы, обозначающие действия [ГОЛОВНОЙ ДУШИ] *–* *слышать* и *говорить*. Картина оказывается весьма контрастной и соответствующей клишированным образам представителей двух народов, что позволяет рассматривать нахождение пойнтер-точки как перспективный способ нахождения токенов, претендующих на представление ядра концепта (в данном случае концепта [ЧЕЛОВЕК] в русской и китайской ПКМ).

Другим способом количественного анализа токенов, выражающих концепт, является энтропийно-анэнтропийный анализ (метод ***RHA*** Т. Г. Петрова, 2008). Анализ заключается в том, что после получения ранговой формулы ***R*** (которой в данном случае является частотный словарь токенов, с помощью которых выражается тот или иной концепт или их группа), осуществляется вычисление информационной энтропии Шеннона (***Н****=*–Σ*p*i*\**ln*p*i где *р*i *–* нормированная к 1 частота *i*-ой лексемы *–* Петров, Фарафонова, 2005, с. 48), характеризующей равномерность распределения токенов, и анэнтропии (***A***=–[(Σln*p*i)/*n*] – ln(*n*), где *n* – число токенов, представляющих концепт *–* Петров, Фарафонова, 2005, с. 61), введённой Т.Г.Петровым для характеристики неравномерности вкладов компонентов (в данном случае – токенов) в их распределение, на основании сопоставления которых у изучаемых объектов делаются содержательные выводы. При этом ***H*** и ***A*** могут рассчитываться для полных (что позволяет полнее охарактеризовать их индивидуальность) или усечённых составов объектов (с целью их сопоставления вне зависимости от природы объектов и способов их изучения).

Для концептов, представляющих составность человека, были вычислены ***H*** и ***A*** для токенов, представляющих концепты [ТЕЛО], [ДУША], [ДУХ] и их совокупности в русских и китайских пословицах (табл. 5).

Таблица 5. Число токенов, представляющих концепт

|  |  |
| --- | --- |
| Концепт | Число токенов, представляющих концепт |
| Китайские пословицы | |
| [ТЕЛО] + [ДУША] +[ДУХ] | 246 |
| [ТЕЛО] | 182 |
| [ДУША] | 60 |
| [ДУХ] | 20 |
| Русские пословицы | |
| [ТЕЛО] + [ДУША] +[ДУХ] | 490 |
| [ТЕЛО] | 336 |
| [ДУША] | 119 |
| [ДУХ] | 43 |

При этом расчёты были произведены по полным и усечённым (в последнем случае по 20 самым частым токенов *–* минимальное число токенов, выражающих концепт, в данном случае концепт[ДУХ]китайских пословиц) словарям токенов.

В итоге получены в целом ожидаемые результаты (рис. 6), которые, тем не менее, ставят некоторые частные вопросы.

Изменение величины энтропии *H* совершенно однотипно и понятно для полных и усечённых словарей русских и китайских пословиц. Для каждого из четырёх сопоставимых комплектов данных *H* является минимальной для словарей представляющих концепт [ДУХ], которые являются самыми короткими во всех случаях, следующими по величине являются энтропии для токенов концепта [ДУША], более детально представленного в русских и китайских пословицах, а ещё больше для концепта [ТЕЛО], наиболее подробного представленного в пословицах. Естественно, *H* для токенов, передающих совокупность концептов [ТЕЛО]+[ДУША]+[ДУХ] является максимальным.

Сравнение этих четырёх комплектов друг с другом даёт более сложную картину: для полных и усечённых составов все величины *H* для русских пословиц сдвинуты в сторону больших величин на два шага, так что *H* концепта [ТЕЛО] китайских пословиц сопоставима с таковой концепта [ДУХ] русских, что свидетельствует о большей дифференциации представления о человеке в русских пословицах. При этом основной вклад в *H* совокупности концептов [ТЕЛО]+[ДУША]+[ДУХ] вносят токенов, передающие концепт [ТЕЛО], причём этот вклад заметно больше для китайских пословиц, в то время как разницы значений *H* для словарей [ДУША] *–* [ТЕЛО] и [ДУХ] *–* [ДУША] примерно одинаковы. При этом, как это следует из самой логики расчётов, начинаясь с одной и той же минимальной величины, для полных словарей размах варьирования *H* оказывается примерно в два раза больше, чем для усечённых.

|  |  |
| --- | --- |
|  |  |
| Рис. 6. Энтропийно-анэнтропийные характеристики наборов токенов, реализующих концепты. Точки 4 и 8 совпадают в связи с тем, что число токенов, реализующих концепт [ДУХ] в китайских пословицах, принято за объём усечённого словаря | |

Таким образом, использование *H* для оценки разнообразия словарей качественно ничего нового не даёт, однако позволяет давать порядковым оценкам (больше*–*меньше) количественную оценку.

Ситуация с анэнтропией оказывается иной *–* характер её изменений для полных и усечённых словарей оказывается совершенно разным.

Для усечённых словарей анэнтропия ***A*** уменьшается с увеличением словаря, т.е. с ростом энтропии *H*, что вполне понятно, т.к. ***A*** является характеристикой своеобразия словаря, которое связано прежде всего с редкими токенами, доля которых при фиксированном количестве токенов, принимаемых в расчёт, уменьшается по мере увеличения полного словаря. Неожиданным оказалось то, что все точки, за исключением одной (12 – концепт [ДУХ] русских пословиц) лежат на одной прямой, причины чего требуют прояснения.

Для полных словарей значения величин анэнтропии ***A*** в несколько раз больше, чем для усечённых, а характер изменения совершенно другой *–* есть тенденция её роста с увеличением объёма словаря и соответственно ростом энтропии *H*. Эта тенденция вполне понятна, т.к. с увеличением словаря может увеличиваться и число токенов с низкой частотой. Поэтому объяснимо увеличение ***A*** при переходе от словаря токенов, выражающих концепт [ДУХ] к выражающим концепт [ДУША], а далее к выражающим концепт [ТЕЛО]. Эта закономерность выражена для концептов русских пословиц (точки 16-15-14), но нарушается для китайских за счёт понижения ***A*** при переходе к токенам, выражающим концепт [ДУША], что связано, по-видимому, с тем, что для его выражения используется большое число высокочастотных специфических токенов. Однако, при переходе от токенов, выражающих концепт [ТЕЛО], к токенам совокупности концептов [ТЕЛО]+[ДУША]+[ДУХ] происходит некоторое падение ***A*** из-за того, что одни и те же токены могут использоваться в зависимости от контекста для передачи разных концептов. Поэтому в результате объединения словарей может поменяться статистический статус токенов. Такой эффект вполне объясним, хотя не был ожидаем.

Проведённое исследование даёт основание говорить о том, что обнаружен новый статистический объект со своеобразным набором характеристик. Его особенности таковы.

- Имеется некоторый текст (в данном случае *–* собрания русских и китайских пословиц), в котором выделен набор однотипных концептов (напр., [ТЕЛО], [ДУША], [ДУХ]; однако, их должно быть достаточно много *–* около десятка и более).

- В таком случае распределение частот этих концептов в ранговой форме будет описываться резко убывающим распределением, которое обладает следующими свойствами:

- - Миллеровское число 7±2 концептов из этого набора покрывает 80% употреблений концептов из этого набора (выполняется закон Парето 80:20).

-- Как правило, частота самых редких концептов измеряется несколькими единицами – десятком в том случае, если эти концепты связаны с тематикой текста. Возможны ситуации более редкого или однократного употребления концепта (в тексте, тематика которого слабо связана с семантикой концепта), но они никогда не образуют длинного хвоста *hapax legomena*, как в случае *h*-распределения токенов (на которое похоже описываемое распределении при условии очень резкого падения *h*-распределения). Это даёт основание квалифицировать данные распределения как распределения с толстыми хвостами (Фуфаев, 1996; Anderson, 2006), которые обрезаны.

- Распределение частот совокупности токенов, выражающих этот набор концептов в данном тексте, аппроксимируется *h*-распределением, вблизи пойнтер-точки ***R*** которого концентрируются токены, наиболее полно представляющие данный набор концептов.

- Для усечённых словарей токенов существует обратное соотношение энтропии ***H*** и анэнтропии ***A***, отклонения от линейности которого характеризует текст.

- Для полных словарей токенов характерно прямое отношение энтропии и анэнтропии, причём анэнтропия полнее отражает особенности текста.

После того, как характеристики такого объекта были выявлены, оказалось, что он даже на момент обнаружения представлен не только русскими и китайскими собраниями пословиц. Аналогичные свойства обнаруживаются и у других наборов концептов.

Такие концепты были выявлены в работах учеников С. В. Чебанова, выполненных в связи с другими задачами. Поэтому их материал не всегда даёт возможность проследить все указанные свойства и дать их количественную характеристику.

Наиболее полный в обсуждаемом контексте оказался материал исследования Д. М. Семёновой распределения частот описаний жестов в первом томе романа Л. Н. Толстого "Война и мир" (Семёнова, 2012).

При этом оказалось, что распределение токенов описания жестов (любое упоминания или описание жеста безотносительно к тому, что это за жест, какой частью тела он осуществляется и т.д., т.е. без всякой категоризации) по героям апроксимируется *H*-распределением (рис. 7 и 8; ср. Рис. 1 и 2).



Рис. 7. Распределение токенов описаний жестов героев 1-го тома романа

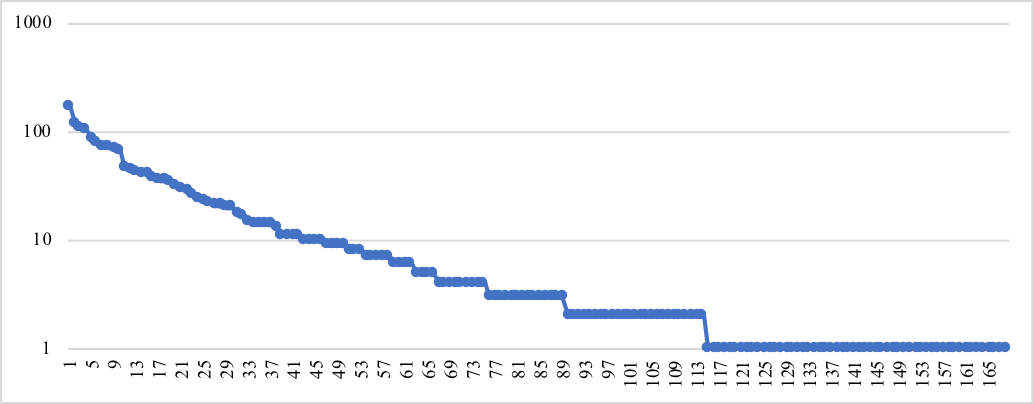


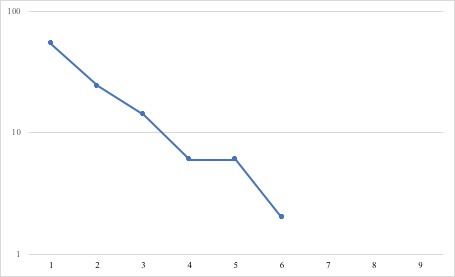
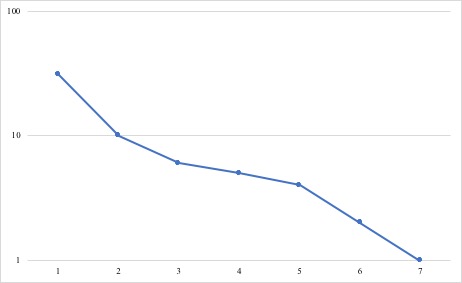
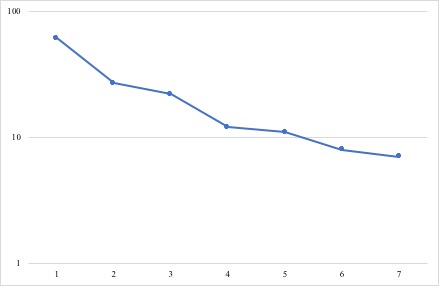
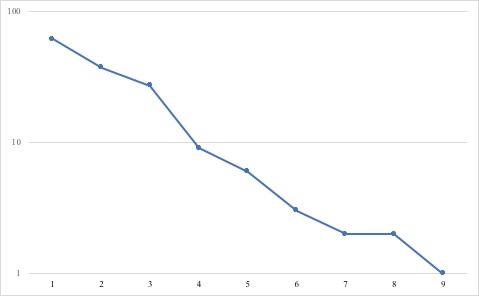
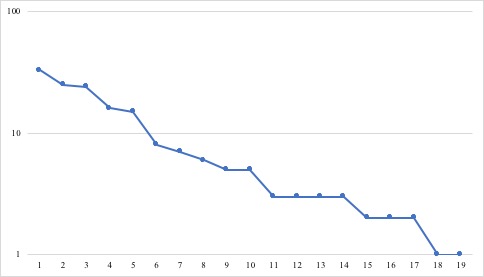
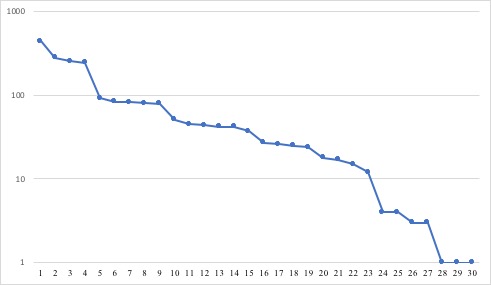
Рис. 8. Распределение токенов описаний жестов героев 1-го тома романа (в полулогарифмических координатах)

Если же осуществляется категоризация, т.е. переход к концептам (т.е. рассмотрение жестов, связанных с определенной частью тела, действиями, характеристиками действий, способом описания), то получаются резко убывающие распределения, такого же типа, как выявлены на материале пословиц (ср. рис. 9), с аналогичной динамикой накопления частот (ср. рис. 5).

При этом оказалось возможным для концепта [ВЗГЛЯД] рассмотреть распределение концептов его вариантов и рассчитать для него пойнтер-точку. Оказалось, что ей соответствует концепт [смотреть (пусто)] с частотой 9 в окружении [опустить глаза (пусто)] и [взгляд (пусто)], что очень хорошо отражает не только значение концепта [ВЗГЛЯД], но и всё содержание текста, который выражает взгляд Л.Н.Толстого на светское общество.

Аналогичные распределения были обнаружены при изучении частоты концептов на другом материале.

Довольно много данных было получено по социальным институтам, которые представляют собой типичные концепты. Частоты упоминания таких институтов как концептов имеется по "Народным русским сказкам" А. Н. Афанасьева (Чернышова, 2008), тому 1 "О лицах" кодекса Наполеона (Ляпунова, 2010), "Соборянам" Н. С. Лескова (Кириллова, 2009), "Истории одного города" и "Сказкам" М. Е. Салтыкова-Щедрина (Смирнова, 2008), произведениям «Жизнь и необычайные приключения солдата Ивана Чонкина" и "Москва 2042" В. Н. Войновича (Смирнова, 2008), письменным работам младших школьников 1990-х и 2000-х гг. (Курочкина, 2008), что обобщено в работе С. В. Чебанова (2012). Во всех этих случаях концепт описывается довольно большим числом токенов, что качественно соответствует распределениям обсуждаемого типа.



1 2 3 4 5 6

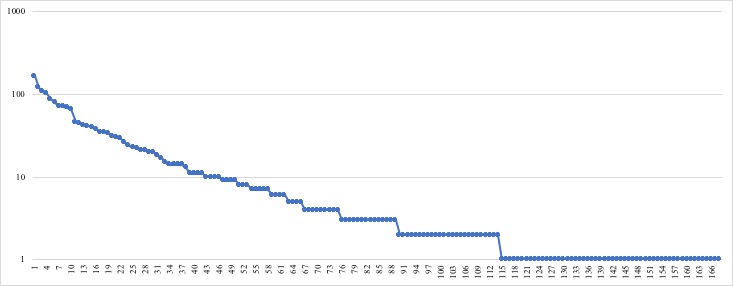


Рис. 9. Сопоставление распределения токенов, описывающих жесты героев, с типичным *H*-распределением жестов по героям в 1-ом томе романа (1 – распределение элементов описания невербальной коммуникации, 2 – распределение элементов описания жестов князя Андрея Болконского, 3 – распределение ключевых элементов по числу описывающих характеристик, 4 – распределение ключевых элементов по группам, 5 – распределение ключевых элементов, охарактеризованных хотя бы одним признаком, 6 – соотношение двух элементов характеристик жестового знака)

Ещё один блок данных по статистике концептов – данные по статистике чисел, в частности, статистике чисел в пословицах (Бабарико, Чебанов, 2014, 2015; Babariko, Jinfeng, Chebanov, 2016). В этом случае в качестве концептов выступают некоторые натуральные числа (1, 2, 3 и т.д.) – числа первого десятка, некоторые числа первой сотни (12, 13, 33, 40, 60 и т.д.), тысяча, миллион и некоторые другие. Так, в русской языковой картине мира концептами являются числа 1812, 1917, 1991 и т.п., иррациональное число √2, трансцендентные числа π и *е* и т.д. После публикации выдающейся работы генетика и мифопоэта В. В. Кушева (Кушев, 1994) число 730 превратилось в концепт.

При этом одно и то же натуральное число может передаваться разными токенами. Так, в Собрании пословиц Даля числа передаются как числительными, так и описательными числовыми концептами (ОЧК). Для первого десятка их соотношение представлено в таблице 6 (Бабарико, Чебанов, 2014).

Таблица 6. Числа, выраженные разными токенами

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Числа** | Выражения чисел | | | | | | | | |
| числительными | | | ОЧК | | | числительными + ОЧК | | |
| форм | употреблений | | форм | употреблений | | форм | употреблений | |
| Абс. | % | Абс. | % | Абс. | % |
| **1** | 18 | 1024 | 34,9 | 20 | 213 | 10,8 | 38 | 1237 | 25,2 |
| **2** | 18 | 665 | 22,6 | 14 | 179 | 9,1 | 24 | 844 | 17,2 |
| **3** | 15 | 362 | 12,3 | 11 | 49 | 2,4 | 24 | 411 | 8,4 |
| **4** | 7 | 128 | 4,4 | 3 | 40 | 2,0 | 9 | 164 | 3,3 |
| **5** | 10 | 91 | 3,1 | 10 | 57 | 2,9 | 17 | 148 | 3,0 |
| **6** | 7 | 41 | 1,4 | 2 | 2 | 0,1 | 4 | 43 | 0,9 |
| **7** | 13 | 208 | 7,0 | 7 | 108 | 5,5 | 14 | 316 | 6,4 |
| **8** | 1 | 11 | 0,4 | 0 | 0 | 0 | 1 | 11 | 0,2 |
| **9** | 8 | 29 | 1,0 | 0 | 0 | 0 | 8 | 29 | 0,6 |
| **10** | 8 | 41 | 1,4 | 2 | 12 | 0,6 | 8 | 53 | 1,1 |

Так, количества передаются следующими ОЧК (Табл. 7).

Таблица 7. Количества, передаваемые с помощью ОЧК.

|  |  |  |  |
| --- | --- | --- | --- |
| числа | ОЧК | Абс | В % |
| **1** | копейка, копеечка | 56 | 3,6 |
| **3** | алтын, алтынник | 42 | 2,7 |
| **5** | пятак, пятерка | 11 | 0,7 |
| **6** | шестёрка | 1 | 0,0 |
| **7** | седмица, неделя | 108 | 6,9 |
| **10** | гривна, десятка | 12 | 0,7 |
| **12** | дюжина, две дюжины | 4 | 0,2 |
| **24** | день, сутки | 493 | 31,8 |
| **30** | месяц | 20 | 1,2 |
| **50** | полтина | 10 | 0,6 |
| **60** | минута, час | 123 | 7,9 |
| **100** | сотня, век, рубль | 434 | 28,0 |
| **365** | год | 232 | 15,0 |
| **Итого** |  | 1565 | 100,0 |

При этом частоты употребления разных концептов резко различаются, что даёт основания ожидать и здесь присутствия распределений обсуждаемого типа.

Совокупность приводимых материалов и соображений даёт основание предполагать, что выявленный довольно своеобразный тип распределений с толстыми хвостами, которые обрублены, является более или менее универсальным распределением концептов, а пойнтер-точки распределений токенов, выражающих набор однородных концептов, является зоной концентрации токенов, наиболее полно представляющих этот набор. Вероятно, эти распределения можно назвать ***усечёнными ципфоподобными распределениями*** или распределениям Линь Цзиньфэн.

В заключение авторы благодарят В.П.Захарова за обсуждение вопросов терминологии.

Список литературы

1. Аскольдов, С. А. Концепт и слово / С. А. Аскольдов // Русская словесность: От теории словесности к структуре текста: антология / под общ. ред. В. П. Нерознака. ― М.: Academia, 1997. ― С. 267–279.
2. Бабарико, М. Н. Арифмология русских пословиц и поговорок собрания В. И. Даля / М. Н. Бабарико, С. В. Чебанов // Структурная и прикладная лингвистика. ― 2014. ― Вып. 10. ― С. 70–91.
3. Бабарико, М. Н. Русская паремиологическая арифмология XIX-XXI веков / М. Н. Бабарико, С. В. Чебанов // Структурная и прикладная лингвистика. ― 2015. ― Вып. 11. ― С. 186–219.
4. Даль, В. И. Пословицы русского народа / В. И. Даль. ― М.: в Университетской типографии, 1862. ― 883 с.
5. Захаров В. П. Корпусная лингвистика: СПб.: СПбГУ. РИО. Филологический факультет, 2013. – 148 с.
6. Карамнов А. С. Количественная оценка повторяемости и сложности лексики в корпусе учебника английского языка // [Филологические науки. Вопросы теории и практики](https://elibrary.ru/contents.asp?id=33960389). 2014. [№ 6-1(36)](https://elibrary.ru/contents.asp?id=33960389&selid=21523175). С. 82-86.
7. Кириллова, М. В. Концепты социальных институтов в произведении Н. С. Лескова «Соборяне» [Электронный ресурс] / М. В. Кириллова, В. А. Чернявский // Институт социологии РАН. Методологический семинар памяти Г. С. Батыгина. ― 2009. ― Режим доступа: http://www.isras.ru/files/File/Seminar/Seminar\_Batygin/Kirillova\_Chernyavsky.pdf
8. Кудрин, Б. И. Мои семь отличий от Ципфа / Б. И. Кудрин // Общая и прикладная ценология. ― 2007. ― № 4. ― С. 25–33.
9. Курочкина, А. С. Социальные концепты в языковом творчестве детей младшего школьного возраста (диахронический анализ). Выпускная квалификационная работа ― СПб.: БГТУ, 2008.
10. Кушев В. В. 730 шагов: Аналитическая прогулка: Пьеса / В. В. Кушев. ― СПб.: Митин журнал: Северо-Запад, 1994. ― 119 с.
11. Лихачёв, Д. С. Концептосфера русского языка / Д. С. Лихачёв // Известия РАН. Серия литературы и языка. ― 1993. ― № 1.
12. Ляпунова, Ю. И. Концепты основных социальных институтов в Гражданском Кодексе Франции и их русские соответствия. Выпускная квалификационная работа ― СПб.: БГТУ, 2010.
13. Миллер, Дж. А. Магическое число семь плюс минус два. О некоторых пределах нашей способности перерабатывать информацию [Электронный ресурс] / Дж. А. Миллер. ― Режим доступа: http://www.ebbinghaus.ru/wp-content/uploads/2010/02/Miller\_564-580.pdf.
14. Петров, Т. Г. Метод *RHA* как решение проблемы систематизации аналитических данных о вещественном составе геологических объектов / Т. Г. Петров // Отечественная геология. ― 2008. ― № 4. ― С. 98–105.
15. Петров, Т. Г. Информационно-компонентный анализ. Метод *RHA*: Учебное пособие / Т. Г. Петров, О. И. Фарафонова. ― СПб.: СПбГУ, 2005. ― 168 с.
16. Пименова, М. В. Введение в концептуальные исследования: учебное пособие / М. В. Пименова, О. Н. Кондратьева. ― Кемерово: Кузбассвузизд, 2006. ― 178 с.
17. Пущин, С. Л. О трёх теоремах Б. И. Кудрина / С. Л. Пущин // Ценологические исследования. ― М.: Техника, 2014. ― Вып. 53. ― С. 11–28.
18. Семёнова, Д. М. Ценоз описаний кинесики романа Л. Н. Толстого «Война и мир» / Д. М. Семёнова, С. В. Чебанов // Ценологические исследования. ― М.: Техника, 2012. ― Вып. 46: Специфика ценологических представлений разных школ. ― С. 181–203.
19. Смирнова, М. А. Концепты социальных институтов в русской сатирической прозе (на примерах произведений М. Е. Салтыкова-Щедрина и В. Н. Войновича). Выпускная квалификационная работа ― СПб.: БГТУ, 2008.
20. Степанов, Ю. С. Константы: словарь русской культуры / Ю. С. Степанов. ― М.: Академический Проект, 2004. ― 992 c.
21. Фуфаев, В. В. Основы теории динамики структуры техноценозов / В. В. Фуфаев // Ценологические исследования. ― Абакан: Центр системных исследований, 1996. ― Вып. 1: Математическое описание ценозов и закономерности технетики. ― C. 156–193.
22. Фуфаев, В. В. Структурно-топологический анализ динамики сообщества банков России в условиях финансового кризиса / В. В. Фуфаев // Технетика и ценология: от теории к практике. ― М.: Ценологические исследования, 2009. ― Вып. 35. ― С. 139–146.
23. Чебанов, С. В. Полнотекстовые базы данных как инструмент понимания (на материале русской лингвосоциологии) / С. В. Чебанов // Понимание и рефлексия в коммуникации, культуре и образовании: материалы Международной научно-практической Интернет-конференции, посвященной 70-летию Факультета иностранных языков и международной коммуникации Тверского государственного университета. 1 октября – 15 декабря 2011 г. ― Тверь: Тверской государственный университет, 2012. ― С. 185–197.
24. Чернышова А. П. Описание концептов социальных институтов в русских народных сказках. СПб: БГТУ, 2008
25. Anderson, Ch. The Long Tail: Why the Future of Business Is Selling Less of More / Ch. Anderson. ― N.Y.: Hyperion, 2006. ― 238 p.
26. Babariko, M. idealized cognitive model (ICM) of numbers in the Chinese (C) and Russian (R) linguistic world picture(LWP) as a basis of conceptual mapping / M. Babariko, L. Jinfeng, S. Chebanov // 3rd International Congress of Numanities (ICoN 2016). Processes, maps, narratives. Program and abstracts. 23–26.05.2016. ― Kaunas: International Semiotics Institute, Kaunas University of Technology, 2016. ― P. 43–45
27. 中国谚语资料, 中国文艺研究会资料室主编, 兰州艺术学院文学系55级民间文学小组, 上中下三册,上海文艺出版社, 1961年, 1111页 Собрание китайских пословиц / Китайская научная библиотека искусств, фольклорная группа факультета литературы Ланьчжоуского института искусств. ― Шанхай: Шанхайское издательство литературы и искусств, 1961. ― Т. 1–2. ― 1111 c.)

1. Предлагаемое использование термина, обычное для английского языка, несколько отличается от принятого в русском, хотя трактовка токена как значимой единицы языка (Захаров, 2013, с. 140) или трактовка А.М.Карамнова (Карамнов, 2014, с. 83) позволяет использовать его в таком смысле. [↑](#footnote-ref-1)