



Казанский федеральный  
УНИВЕРСИТЕТ

# О корпусных исследованиях учебной литературы

на материале учебных пособий  
по гуманитарным дисциплинам

**Кулигин А.С.**

к.ф.н. ст. преподаватель

Подготовительный факультет для иностранных учащихся КФУ



## Корпусное исследование учебной литературы КФУ

### Ресурсы в открытом доступе:

- электронный архив КФУ
- сайты институтов / факультетов
- открытые ресурсы научной библиотеки им. Лобачевского

## Состав корпуса

Тематика	Документов	Словоупотреблений
Журналистика	9	159 212
Юриспруденция	15	308 298
Филология	14	300 331
Политология и МО	11	312 109
Психология и педагогика	11	303 062
<b>Всего</b>	<b>60</b>	<b>1 383 012</b>



# Программное обеспечение

## Морфологический анализатор TreeTagger

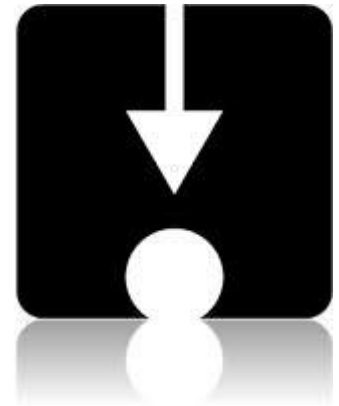
Schmid H. Probabilistic part-of-speech tagging using decision trees // International Conference on New Methods in Language Processing. – 1994. – pp.44–49, Manchester, UK.

## Корпус-менеджер LancsBox X

Brezina, V. & Platt, W. (2023)

*#LancsBox X* [software],

Lancaster University, <http://lancsbox.lancs.ac.uk>.



## Вывод Tree Tagger

*На смену остросоциальным произведениям приходит психологический роман.*

```
<S>
  На          на          Sp-a
  Смену       смена       Ncfsan
  Остросоциальным  остросоциальный  Afpmpdf
  Произведениям  произведение     Ncnpdn
  приходит       приходит        Vmip3s-a-e
  психологический  психологический  Afpmsnf
  роман          роман          Ncmsnn
  .              .              -
</S>
```

*формат tsv*

# Программное обеспечение

#LancsBox X 2.1.0

Q [lemma="явление"]

KFU Humanities whole corpus 1M

Q [lemma="явление"] Hits: 927 (670,28) Texts: 58/60

File	Left	Node	Right
jour_01.xml	таком популярном в XXI веке	явлении	как storytelling . 5. « Разница между новостью и
psy_15.xml	искал новые комбинации , преобразовывая вещи ,	явления ,	процессы действительности , искал неизвестные связи
psy_15.xml	целостной картины мира , диалектических взаимосвязей	явлений	природы . Межпредметные связи с точки
polit_02.xml	делает акцент на практических видоизменениях	явлений .	Таким образом , политический процесс раскрывает
ling_22.xml	омонимов , паронимы и их виды ,	явление	параномазии . Ключевые слова : асистемные отношения ,
ling_23.xml	мире , лежащем между миром внешних	явлений	и внутренним миром человека » . В
psy_03.xml	информацию согласием и принятием . Возможно ,	явление	сопротивления возникает как сигнал о
ling_04.xml	драматургией последнего порубежья ? 3. Когда возникает	явление ,	получившее название « новая немецкая драма » ? 4.
jour_04.xml	типологическое и генетическое в возникновении	явлений .	Акцентируется внимание на такие понятия ,
ling_06.xml	и внутреннего потенциала самого воспринимаемого	явления	с точки зрения его стимулирующего
ling_06.xml	которым принадлежат воспринимающее и воспринимаемое	явления .	Историко-литературное и теоретическое значение дифференциаци
psy_21.xml	одномерное , систематическое и целенаправленное восприятие	явлений	с целью их изучения . 4. процесс ,
psy_01.xml	как о субъективном качестве восприятий	явлений	жизни [ 7 ] . О.С . Редкозубова , А.А . Сычев приводят
ling_06.xml	В чем специфика критического восприятия	явлений	инонациональной литературы ? 7. Какую роль играют
psy_24.xml	для демонстрационного и лабораторного воспроизведения	явлений ,	их качеств и количеств , исследования (
ling_23.xml	когда наша задача не восстановление	явлений	прошлого , а вообще - раскрытие законов
polit_01.xml	науках . Он требует изучения всех	явлений	жизни в последовательном временном развитии ,
ling_05.xml	всего сущего - всех людей , всех	явлений ,	всех деяний , всех вещей ; - равенство
psy_15.xml	понятие и его характеристики Всякое	явление ,	любой процесс представляет собой единство .

Searched KWIC for "[lemma="явление"]".

# Программное обеспечение

#LancsBox X 2.1.0

Q [lemma="явление"]

KU Humanities whole corpus 1M

Summary table

Q [lemma="явление"] Hits: 927 (670,28) Texts: 58/60

Node

word

Value	Hits	Texts	gramm	lemma	lemma_pos	lemma_rki	pos	rki
явлений	289	45	1	1	1	1	1	1
явления	265	49	3	1	1	1	1	1
явление	181	46	2	1	1	1	1	1
явлениями	43	19	1	1	1	1	1	1
явлением	35	21	1	1	1	1	1	1
явлениях	29	15	1	1	1	1	1	1
явлении	27	17	1	1	1	1	1	1
явлению	23	14	1	1	1	1	1	1
явлениям	22	15	1	1	1	1	1	1
Явление	8	6	1	1	1	1	1	1
Явления	3	2	2	1	1	1	1	1
Явлению	1	1	1	1	1	1	1	1
Явлением	1	1	1	1	1	1	1	1

Close

Opened KWIC summary table.



# Программное обеспечение

#LancsBox X 2.1.0

Q [lemma="явление"]

KFU Humanities whole corpus 1M lemma

Terms: 48 189

Term	Frequency	Rel. frequency	ARF (average ...	Range	Range %	CV (coefficien...	Juilland's D	DP (deviation ...
за	3 106	2 245,82	1 594,78	60	100,00	0,58	0,92	0,22
тот	3 025	2 187,26	1 775,68	60	100,00	0,31	0,96	0,10
лицо	2 967	2 145,32	863,00	59	98,33	1,96	0,74	0,61
слово	2 949	2 132,30	935,38	58	96,67	2,30	0,70	0,58
форма	2 929	2 117,84	1 305,32	60	100,00	0,94	0,88	0,29
государство	2 852	2 062,17	911,88	50	83,33	1,47	0,81	0,53
язык	2 760	1 995,64	517,85	50	83,33	2,59	0,66	0,73
себя	2 737	1 979,01	1 502,30	60	100,00	0,48	0,94	0,18
государственный	2 569	1 857,54	708,98	53	88,33	1,95	0,75	0,57
новый	2 523	1 824,28	1 173,04	60	100,00	0,77	0,90	0,27
вопрос	2 515	1 818,49	1 257,39	59	98,33	0,51	0,93	0,20
так	2 482	1 794,63	1 441,79	59	98,33	0,43	0,94	0,15
власть	2 481	1 793,91	703,78	51	85,00	2,20	0,71	0,59
время	2 463	1 780,90	1 333,10	60	100,00	0,51	0,93	0,18
основной	2 452	1 772,94	1 264,32	60	100,00	0,53	0,93	0,19
она	2 429	1 756,31	1 248,91	59	98,33	0,54	0,93	0,21
также	2 347	1 697,02	1 359,90	60	100,00	0,51	0,93	0,21
иметь	2 314	1 673,16	1 276,45	60	100,00	0,50	0,94	0,15
только	2 278	1 647,13	1 291,44	60	100,00	0,41	0,95	0,15

Changed unit to lemma.





# Квантитативные данные

## Доля частей речи в корпусе

Часть речи	Доля (%)	Часть речи	Доля (%)
Существительные	42,2	Наречия	2,5
Прилагательные	15,5	<b>Причастия</b>	<b>2,2</b>
Предлоги	10,9	Числительные	2
Глаголы	8,2	Частицы	1,6
Местоимения	7,5	Деепричастия	0,3
Союзы	6,8	Междометия	0,02

## Частота причастий в тематических подкорпусах

Тематика	Частота	
	Абсолютная	Относительная (на 100 000)
Юриспруденция	9 297	3 015,58
Психология, педагогика	6 340	2 091,98
Филология	6 022	2 005,12
Журналистика	3 189	2 002,99
Политология и МО	6 135	1 965,65

Распределение причастий между тематическими подкорпусами носит равномерный характер.

## Частота причастий в тематических подкорпусах

Тематика	Частота	
	Абсолютная	Относительная (на 100 000)
Юриспруденция	9 297	3 015,58
Психология, педагогика	6 340	2 091,98
Филология	6 022	2 005,12
Журналистика	3 189	2 002,99
Политология и МО	6 135	1 965,65

Юридический подкорпус является исключением.

## Сочетания с причастиями *предусмотренный* и *установленный* в юридическом подкорпусе

<i>предусмотренный</i> <i>установленный</i>	<i>законом</i>	<i>порядок</i>
	<i>статьёй</i>	<i>случай</i>
	<i>кодексом</i>	<i>срок</i>
	<i>конвенцией</i>	<i>предел</i>
	<i>конституцией</i>	<i>ответственность</i>

- *предусмотренный* (576 вхождений)
- *установленный* (367 вхождений)

## Причастия, превосходящие по частоте производящие глаголы

- *выраженный*
- *достигший*
- *закреплённый*
- *направленный*
- *обусловленный*
- *объединённый*
- *опубликованный*
- *ориентированный*
- *основанный*
- *посвящённый*
- *поставленный*
- *построенный*
- *предназначенный*
- *представленный*
- *предусмотренный*
- *признанный*
- *принятый*
- *производящий*
- *разработанный*
- *расположенный*
- *распространённый*
- *регулирующий*
- *связанный*
- *сложившийся*
- *совершённый*
- *совершивший*
- *созданный*
- *указанный*
- *установленный*



# Причастные обороты

## Примеры

1. Акт агрессии может быть **выражен в различных формах**.
2. Во французской Западной Африке был принят ряд **мер, направленных на улучшение** положения женщины.
3. Юмор – явление, **обусловленное** душевными особенностями человека.
4. Это первый в политантропологии **труд, специально посвящённый** данной дисциплине.
5. Системно **представленные результаты** должны быть корректно интерпретированы.

## Примеры

6. *Принятый закон* подписывает и обнародует глава субъекта Федерации.
7. Аллюзия как *наиболее распространённый* вид цитаты.
8. Ещё одна *проблема, связанная с образом Воланда*, – это проблема земной власти.
9. *Натурализм* – литературное направление, *сложившееся в последней трети XIX века*.
10. Как показала история, *государства, созданные после Версальского договора*, оказались маломощным, неэффективным буфером.

# Инфинитивные обороты

## Слова, управляющие инфинитивными оборотами

Предикатив	Журналистика		Юриспруденция		Филология		Политология		Психология		Док. из 60
	отн.	абс.	отн.	абс.	отн.	абс.	отн.	абс.	отн.	абс.	
<i>можно</i>	189,68	302	68,76	212	72,92	219	97,08	303	130,34	395	60
<i>необходимо</i>	69,72	111	24,65	76	33,30	100	28,20	88	55,76	169	57
<i>вправе</i>	0,63	1	51,25	158	0,33	1	0,96	3	3,30	10	19

Прилагат.	Журналистика		Юриспруденция		Филология		Политология		Психология		Док. из 60
	отн.	абс.	отн.	абс.	отн.	абс.	отн.	абс.	отн.	абс.	
<i>должен / должный</i>	197,85	315	126,83	391	44,95	135	105,09	328	172,57	523	59
<i>способен / способный</i>	16,96	27	6,76	27	18,31	55	35,24	110	34,98	106	51
<i>обязан</i>	3,77	6	20,43	63	0,67	2	7,37	23	3,96	12	30



## Слова, управляющие инфинитивными оборотами

Глагол	Журналистика		Юриспруденция		Филология		Политология		Психология		Док. из 60
	отн.	абс.	отн.	абс.	отн.	абс.	отн.	абс.	отн.	абс.	
<i>мочь</i>	294,58	469	418,43	1290	210,43	632	289,00	902	370,55	1123	60
<i>следует</i>	39,57	63	62,60	193	44,95	135	50,62	158	47,19	143	58
<i>стать</i>	49,62	79	13,95	43	23,64	71	61,20	191	27,06	82	48
<i>позволять</i>	47,11	75	17,52	54	17,65	53	28,52	89	62,69	190	56
<i>пытаться</i>	16,33	26	2,59	8	19,64	59	25,95	81	22,44	68	40
<i>начинать</i>	19,47	31	4,87	15	17,65	53	12,50	39	33,00	100	45
<i>стремиться</i>	23,87	38	2,92	9	26,64	80	19,86	62	14,85	45	41
<i>хотеть</i>	24,50	39	3,24	10	13,98	42	9,29	29	33,66	102	34
<i>помогать</i>	23,87	38	0,97	3	5,33	16	6,41	20	35,64	108	36
<i>начать</i>	23,87	38	2,27	7	17,65	53	19,22	60	7,59	23	35

## Примеры конструкций со *следует*:

- При изучении рекомендованной литературы **следует делать** конспект.
- Уже во время встречи **следует думать**, как ваше интервью будет выглядеть на бумаге или в эфире.
- В начале консультирования **следует избегать** вопросов, на которые возможен однозначный ответ [...].
- Каждое из анализируемых произведений **следует воспринимать** в контексте творчества писателя
- Интересы **не следует смешивать** с целями [...].
- Какие требования **следует соблюдать** при выставлении оценок?

## Примеры конструкций с **позволяет**:

- *Функциональный подход **позволяет** выявить своеобразие журналистики [...].*
- *Состав преступления **позволяет** отграничить одно преступное деяние от другого.*
- *Путешествие героев **позволяет** Ильфу и Петрову **создать** широкую сатирическую панораму нового мира.*
- *Таблица 4 **позволяет** увидеть, что население Японии, ЕС и РФ, будет иметь наибольшую долю лиц старше 65 лет.*
- *Выделение субсистем **позволяет** четко **обозначить** структуру семьи.*

## Конструкции в стилевом сопоставлении

Значение	РКИ	НСР
Необходимость	[кому] надо [инф.] [кому] нужно [инф.]	[кому] необходимо [инф.] [кому] следует [инф.] [кому] сто́ит [инф.]
Обязанность	[кто] должен [инф.]	[кто] обязан [инф.]
Возможность	[кто] может [инф.]	[кто] вправе [инф.] [кто] способен [инф.] [кто] имеет возможность [инф.]
Интенция	[кто] пробует [инф.] [кто] старается [инф.]	[кто] пытается [инф.] [кто] стремится [инф.]



**СПАСИБО ЗА ВНИМАНИЕ!**