

УДК 81-139

Кирина Маргарита Александровна
младший научный сотрудник Лаборатории языковой конвергенции
Национальный исследовательский университет «Высшая школа экономики»
г. Санкт-Петербург, Россия

Margarita Kirina
Junior Research Fellow of the Linguistic Convergence Laboratory
National Research University «Higher School of Economics»
St. Petersburg, Russia
mkirina@hse.ru

Лукьянчикова Алиса Сергеевна
стажер-исследователь Лаборатории языковой конвергенции
Национальный исследовательский университет «Высшая школа экономики»
г. Санкт-Петербург, Россия

Alisa Lukyanchikova
Research Assistant of the Linguistic Convergence Laboratory
National Research University «Higher School of Economics»
St. Petersburg, Russia
lukyanchikova@edu.hse.ru

**«ЗВЕЗДЫ РЕКОМЕНДУЮТ ВЕСАМ ПИТЬ СЛИВОВОЕ ВИНО»:
ИССЛЕДОВАНИЕ АСТРОЛОГИЧЕСКОГО ДИСКУРСА НА ОСНОВЕ
РАСПРЕДЕЛЕНИЙ ЧАСТОТНОЙ ЛЕКСИКИ И СЕНТИМЕНТ-АНАЛИЗА**

В статье рассматриваются характерные особенности гороскопических текстов как части астрологического дискурса. Материалом исследования выступает представительная выборка ежедневных предсказаний на русском языке, опубликованных в открытых группах социальной сети «ВКонтакте», суммарным объемом 1185425 словоупотреблений. С использованием методов корпусной и компьютерной лингвистики анализируются содержательные лексические единицы – как общие, так и отличительные для каждого знака зодиака (в сопоставлении с устоявшимися архетипами). Кроме того, проводится сравнение верхней зоны частотных словарей, полученных на основе собранного корпуса гороскопов и подкорпуса «гороскопы», входящего в Национальный корпус русского языка (НКРЯ). Немаловажным этапом настоящей работы стало определение эмоциональной валентности текстов в соответствии с тринарной классификацией (положительная, отрицательная, нейтральная). Анализ тональности, проведенный с использованием словаря оценочной лексики RuSentiLex, позволил установить, что гороскопы имеют преимущественно положительную или нейтральную эмоциональную окраску, а в среднем наиболее позитивным оказался поджанр «любовный гороскоп». В результате исследования с опорой на количественные данные удалось выявить специфические черты современных астрологических предсказаний на лексическом уровне и в контексте их эмоциональной «заряженности», а также отметить некоторые ограничения изучения этого вида текстов.

Ключевые слова: астрологический прогноз; корпус гороскопов; зодиакальные архетипы; корпусная лингвистика; частотный словарь; анализ тональности.

“STARS RECOMMEND LIBRAS TO DRINK PLUM WINE”:
A STUDY OF ASTROLOGICAL DISCOURSE BASED
ON WORD FREQUENCY DISTRIBUTIONS AND SENTIMENT ANALYSIS

This article explores the distinctive features of horoscope texts as part of astrological discourse. The research material consists of a representative sample of daily horoscopes in Russian, published in open groups on the social network VKontakte, with a total volume of 1,185,425 word occurrences. Using methods from corpus and computational linguistics, the study analyses substantive lexical units—both general and unique to each zodiac sign (in comparison with established archetypes). Additionally, a comparison is conducted between the most frequent words derived from the collected horoscope corpus and those from the “horoscopes” subcorpus of the Russian National Corpus (RNC). An important stage of the work involved determining the emotional valence of the texts according to a trinary classification (positive, negative, neutral). Sentiment analysis based on the RuSentiLex evaluative lexicon revealed that horoscopes predominantly have a positive or neutral emotional tone, with the “love horoscope” subgenre being on average the most positive. As a result, relying on the quantitative data, it was possible to identify lexical and emotional characteristics of modern astrological forecasts, as well as to note some limitations when studying this type of text.

Key words: astrological forecast; horoscope corpus; zodiac archetypes; corpus linguistics; frequency dictionary; sentiment analysis.

Астрологический прогноз – особый жанр, получивший широкое распространение в медийном пространстве постсоветской действительности, который берет свое начало еще в донаучном периоде развития общества. Понятие гороскопа, не имея четких границ и исключительного определения, сегодня во многих культурах воспринимается как предсказание на день. Хотя сфера бытования гороскопов крайне обширна, из-за их маргинального, даже псевдонаучного, статуса [1], эти тексты не столь популярны в научном сообществе. Тем не менее такие особенности гороскопического текста, как функциональная направленность, рекомендательный характер, а также специфика выражения модальности и оценки, вызывают лингвистический интерес и описываются в ряде исследований [2–3]. Языковые средства, свойственные этим прогнозам, рассматриваются в рамках как магического [4], так и астрологического дискурсов [5]. Неслучайно характерные признаки названных текстов изучаются и как черты медижанров [6]: гороскопы встречаются в журналах, газетах, распространяются по радио, публикуются на форумах в сети Интернет.

Популярность астрологических прогнозов среди широкой аудитории объясняется конструируемой «персонализированностью» текстов – каждый с нетерпением ждет, что сегодня скажут «про него». Более того, считается, что за знаками зодиака закреплены определенные архетипы [7], обобщение которых представлено, в соответствии с [2; 8], в Таблице 1. Например, Овна относят к образу воина, Тельца – к образу хранителя [Там же]. В этой связи интерес представляет то, в какой степени гороскопическое предсказание оказывается в действительности «адресным», то есть насколько в нем (намеренно или нет) отражаются ассоциируемые с тем или иным знаком зодиака стереотипы.

Т а б л и ц а 1

Характеристика знаков зодиака по архетипам

№	Знак зодиака	Архетип
1	Овен	Воин, первооткрыватель
2	Телец	Хранитель
3	Близнецы	Коммуникатор, посредник
4	Рак	Хранитель домашнего очага (мать)
5	Лев	Король, созидатель
6	Дева	Служитель
7	Весы	Миротворец
8	Скорпион	Преобразователь
9	Стрелец	Философ, учитель
10	Козерог	Стратег
11	Водолей	Новатор, гуманист
12	Рыбы	Мечтатель, мистик

Цель настоящего исследования – проследить особенности лексического состава гороскопов, в том числе проверить, проявляются ли архетипические черты в астрологических прогнозах. Под гороскопами будут пониматься ежедневные предсказания, то есть тексты, которые предписывают то или иное поведение или пророчат некоторые события в определенный день согласно знаку зодиака человека. Методологической основой этой работы станет корпусный подход. В отличие от классических методов, компьютерные инструменты позволяют, во-первых, принять во внимание больший объем данных, а во-вторых – независимо от интроспективных допущений, рассмотреть общие лексико-тематические особенности групп текстов, разделенных в нашем случае в зависимости от целевой аудитории, то есть соответствующего знака зодиака.

Материалом выступили гороскопы, размещенные в открытых группах «ВКонтакте», которые занимаются регулярной публикацией ежедневных астрологических прогнозов для каждого из 12 знаков зодиака с 2014 по 2025 гг. (включительно)¹. Пример выгрузки, за исключением некоторых не используемых в рамках этого исследования метаданных (количество лайков, комментариев, репостов, просмотров, оставленных пользователями), представлен в Таблице 2. В «сыром» виде полученный датасет насчитывал 90 642 строки (около 17 тыс. строк на каждый знак зодиака). Далее осу-

¹ Вопрос о возможности попадания в корпус текстов, сгенерированных искусственным интеллектом, остается за рамками текущего исследования.

ществлялась очистка данных в несколько этапов. Так, были удалены тексты, не являющиеся гороскопами вида «ежедневное предсказание»: рекламные и поздравительные посты, гороскопы на неделю, год и иные временные промежутки, а также гороскопы, составленные не в соответствии с зодиакальной системой (например, кельтские гороскопы). Затем автоматически были «развернуты» посты, которые в одном тексте включали несколько видов, – общий, бизнес- и любовный гороскопы.

Т а б л и ц а 2

Пример выгрузки (фрагмент датасета)

ID	Заголовок	Текст	Год публикации	Знак зодиака
1436	Любовный гороскоп на 16 марта 2025 года	Спокойствие, только спокойствие – так проще достичь желаемого!	2025	Стрелец
4911	Гороскоп на 30 мая 2024 года	Нынче у вас должна обостриться интуиция. Ей и следуйте.	2024	Лев
31540	Гороскоп на 25 апреля 2018	Наиболее трудные дела пока отложите.	2018	Рак
33931	Бизнес-гороскоп на 01 октября 2017	Будьте осторожны в финансовых делах. Возможны происки неожиданных конкурентов.	2017	Телец
46142	Общий гороскоп на завтра, 9 сентября, 2016	Вы прекрасный исполнитель и способны многого добиться. Сегодня Вам предстоит работать с общественностью.	2016	Козерог

После удаления «мусорных» постов и гороскопов-дублей (без учета знака зодиака), токенизации и лемматизации, выполненных с помощью MyStem [9], а также сегментации на предложения с использованием библиотеки NLTK [10], объем корпуса составил 28 149 строк или 1 185 425 словоупотреблений (15578 лексических типов). Средний размер группы гороскопов, приходящихся на каждый знак зодиака, равен 2346 текстам со средним суммарным объемом 99 427,42 токена. Текст гороскопа как единица анализа имеет следующее лингвостатистическое описание (в средних значениях): длина в словах – 42,39 (σ – 13,97), длина в предложениях – 3,42 (σ – 1,18).

Перейдем к результатам. Сначала рассмотрим лексические особенности гороскопических текстов. Верхняя зона частотного словаря, составленного по данным всего корпуса, представлена в Таблице 3, а характерные содержательные лексемы на Рисунке 1. На примере наиболее релевантных сочетаний слов прослеживаются прагматико-дискурсивные черты подобного текста: предсказание событий на завтра (*день, завтра*), ориентация на положитель-

ный исход дня (*благоприятный, день*), а также межличностные отношения и рекомендации по их организации, в частности через открытость к новым встречам и знакомствам (*человек, новый, знакомство, отношение*).

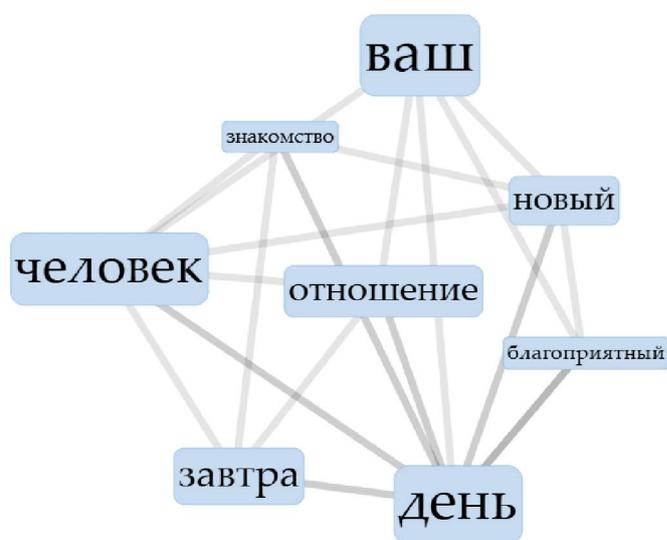


Рис. 1. Сетевая визуализация со встречаемости наиболее частотных слов (исключая стоп-слова)

Стоит отметить, что самые распространенные слова, то есть такие, которые встречаются в каждом документе корпуса, составляют 18,83 % от общего числа лексических типов (с учетом повторений – 92,27 % от объема всего корпуса). При этом уникальные для каждого из подкорпусов (то есть знаков зодиака) единицы не оказываются интерпретируемыми и в своем большинстве не обнаруживают статистически значимой разницы в частотных распределениях (эксперимент по извлечению ключевых слов осуществлялся при использовании Log-Likelihood и пороговом значении $p\text{-value} < 0,05$, в качестве референтного для каждого целевого корпуса брались оставшиеся 11). Однако некоторые наблюдения путем фронтального анализа лексико-тематических групп и характерных архетипических мотивов представляется возможным. Обратимся для этого к Рисунку 2, на котором для каждой «зодиакальной» выборки визуализированы по 50 наиболее частотных содержательных и не входящих во множество общих для всего корпуса лексем.

Например, в облаке слов, характеризующем Овна заметен фокус на карьеру (*шеф, торговля, совещание*) и советы «заботиться» (возможно, из-за стереотипов недостатка созидательности у представителей этого знака зодиака). Значительное место занимает лексика, связанная со знаниями, их получением и применением (*образование, изучение, философский*). В отношении Тельца заметна «предостерегающая» лексика (*угроза, агрессивный*), а также связанная с работой – примечательно, что появляется фигура подчиненного (*работник, офис, совещание*), а также материальный аспект (*вложение*). Вновь появляется указание на сферу взаимоотношений (*привязанность, заботиться*). Тема работы наблюдается и в частотных распределениях гороско-

пов, адресованных Козерогу (*перегрузка, масштаб, нервничать*), что согласуется с образом «трудяги», через который часто описывается этот знак. Для Льва тоже отмечается лексика, связанная с работой (*шеф, офис, специалист*), однако еще и с указанием на такие, действительно, приписываемые ему черты, как успешность в карьере и лидерские качества (*победитель, руководить*).

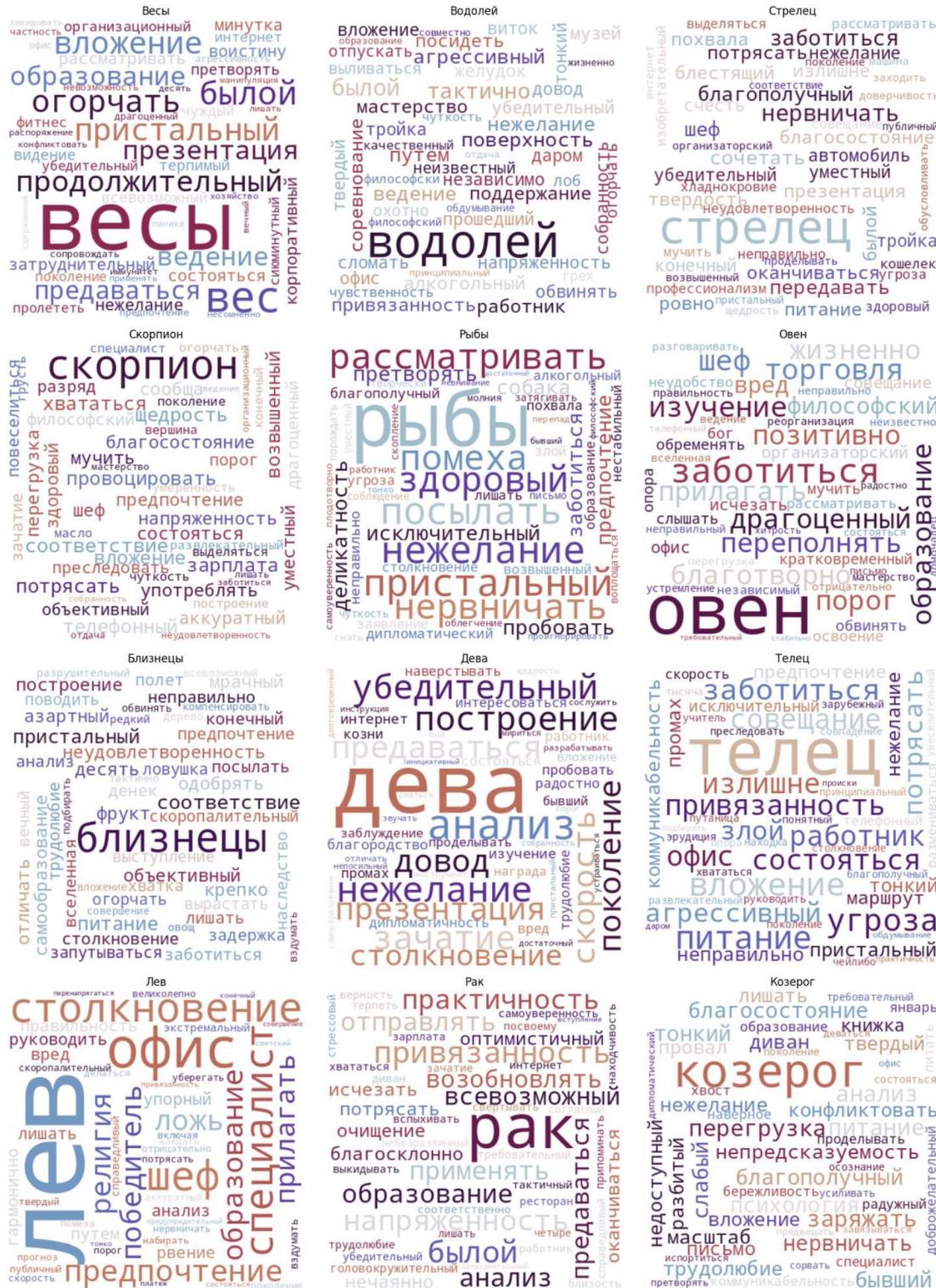


Рис. 2. Облака частотных слов для каждого подкорпуса (исключая стоп-слова и общие для всего корпуса)

Любопытно, что при всей доминантной для архетипа созидательности (*возобновлять, благосклонно*), в подкорпусе Рака мало представлена рабочая сфера. Однако отмечаются слова, описывающие отношение к жизни (*практичность, трудолюбие, знание*) и к людям (*привязанность, верность*), а также экспрессивная, говорящая о чувственном восприятии действительности лексика (*всевозможный, потрясать, головокружительный, вспыхивать, напряженность*). Для Скорпиона выделяются лексико-тематические группы со значением богатства/успеха (*щедрость, благосостояние, состояние, успех, вложение, зарплата*), а также «напряжение – отдых» (*повеселиться, напряженность*). Однако это в большей мере отражает тематическую направленность гороскопов, предсказания в которых строятся с опорой на сферу рабочих или личных отношений, чем указывает на особенности характера этого знака зодиака. Дева же, напротив, описывается через слова, связанные с логикой, аналитическим мышлением (*убедительный, анализ, довод, построение*), что согласуется со стереотипом о ее педантичности.

Несмотря на отмечаемые в некоторых случаях различия, проведенный анализ указывает на необходимость обращения к гороскопу как к единому дискурсивному полю, отличающемуся слабо выраженной дистинктивностью (по крайней мере, если в качестве дифференцирующего параметра рассматривать знак зодиака). В связи с этим мы провели еще один эксперимент, а именно сравнили частотные распределения лемм в пользовательском корпусе, описанном в этом исследовании (и условно названном «гороскопы в соцсетях»), и в подкорпусе гороскопов Национального корпуса русского языка (НКРЯ) [11]. Подкорпус «гороскопы» НКРЯ насчитывает 130 текстов суммарным объемом 105877 словоупотреблений. Как позволяют судить метаданные, гороскопы, включенные в НКРЯ, были составлены в период с 1991 по 2010 гг. (включительно) и представляют собой принципиально другой вид медиа. Так, их источниками выступают журналы, газеты, а также, что немаловажно, встречаются авторские астрологические прогнозы.

В результате сравнения верхней зоны частотных словарей (см. Таблицу 3) стала заметна ориентированность гороскопов, публикуемых в социальных сетях, на прогнозирование событий завтрашнего дня, в то время как гороскопы в журналах и газетах, что несколько ожидаемо, охватывали конкретный *день* или *неделю* целиком. Объяснение находится в особенностях сферы функционирования этих текстов, а также в исключении из нашего корпуса не ежедневных предсказаний. Кроме того, кажется, что предсказания из соцсетей стремятся быть более личными (см. ранг местоимения *вы*), отсутствие чего в подкорпусе гороскопов из НКРЯ, вероятно, является метой ограничений, накладываемых на периодические издания, их внешней модерлируемостью.

В обоих корпусах заметна тема рабочих отношений. Однако в подкорпусе НКРЯ лемма *работа* имеет ранг 24, а в корпусе «гороскопов в соцсетях» – 50. Похожая тенденция на «опущение» по рангу заметна и для

некоторых других лексем: *здоровье* – 31 vs. 234, *любовь* – 34 vs. 87, *деньги* – 37 vs. 289. По всей видимости, для ежедневных онлайн-предсказаний, хоть и важна, но в меньшей мере свойственна привязка к какому-либо конкретному аспекту жизни, в результате чего они представляются более общими и несколько абстрактными. В то время как для публикуемых в официальных СМИ гороскопах характерна выраженная категоризация «по сфере действия».

Т а б л и ц а 3

Сравнение частотных словарей:
гороскопы в соцсетях vs. подкорпус НКРЯ (верхняя зона)

Гороскопы в соцсетях				НКРЯ (подкорпус "гороскопы")			
Ранг	Лемма	Частота		Ранг	Лемма	Частота	
		Абс.	IPM			Абс.	IPM
1	и	40706	34338,74	1	и	3954	37345,22
2	вы	38437	32424,66	2	в	3297	31139,91
3	в	32133	27106,73	3	не	1831	17293,65
4	не	29601	24970,79	4	с	1675	15820,24
5	быть	22860	19284,22	5	на	1615	15253,55
6	на	17127	14447,98	6	вы	1612	15225,21
7	что	15461	13042,58	7	быть	1338	12637,31
8	с	14595	12312,04	8	что	1024	9671,60
9	то	14137	11925,68	9	а	862	8141,52
10	ваш	13956	11772,99	10	по	768	7253,70
11	день	12758	10762,38	11	к	675	6375,32
12	мочь	11737	9901,09	12	для	655	6186,42
13	свой	11717	9884,219	13	это	617	5827,52
14	сегодня	10877	9175,612	14	то	613	5789,74
15	человек	10506	8862,644	15	свой	581	5487,50
16	для	10174	8582,58	16	мочь	569	5374,16
17	он	8644	7291,90	17	но	556	5251,38
18	или	8568	7227,79	18	ваш	555	5241,93
19	к	8370	7060,76	19	он	518	4892,47
20	это	8214	6929,16	20	этот	511	4826,36

21	если	7995	6744,42	21	день	510	4816,91
22	этот	7722	6514,12	22	если	507	4788,58
23	завтра	6574	5545,69	23	неделя	467	4410,78
24	они	6209	5237,78	24	работа	459	4335,22
25	который	6032	5088,47	25	или	451	4259,66

Следующим шагом этой работы стало выявление особенностей эмоционального воздействия гороскопических текстов, определяемых на лексическом уровне. Для этого с использованием русскоязычного словаря оценочной лексики RuSentiLex [12] и программного обеспечения Orange [13] был проведен анализ тональности (или сентимент-анализ). Задача анализа тональности сводится к определению текста, выражающего мнение, и его характеристики с точки зрения направленности и степени выраженности эмоционального отношения (положительного, отрицательного или нейтрального) [14]. Расчет оценки тональности в настоящем исследовании производился в соответствии с [15] по следующей формуле:

$$(sum(pos) - sum(neg)) / document_length * 100,$$

где pos – количество положительных слов (в типах), neg – количество отрицательных слов (в типах), а document_length – длина документа (количество слов в тексте) [16].

В результате были получены следующие распределения по количеству текстов: положительные – 12093 (42,96 %), нейтральные – 10788 (38,32 %), негативные – 5268 (18,71 %), что согласуется с заключениями, представленными в [17]. Так, и в нашем случае, кажется, что лексико-эмоциональное наполнение текстов имеет целью вызвать у читателя положительные эмоции: спокойствие, надежду, стабильность. Это гарантируется чтением предсказания, что функционально выступает действием эквивалентным возможности контролировать свое будущее. При этом, как видно исходя из Рисунка 3, знак зодиака не влияет на эмоциональную направленность текста, что вновь свидетельствует о генерализованном характере астрологических предсказаний.

Приведем несколько примеров положительных (1–2), негативных (3) и нейтральных предсказаний (4–5):

(1) *Сегодня вы одержите пусть небольшую, но победу. Этот день принесет вам уважение окружающих.*

(2) *Сегодня вы просто обречены на успех в амурных делах. Не играйте в карты!*

(3) *Сегодня вы рискуете допустить ошибку или в чем-то просчитаться. Вам могут предоставить недостоверную информацию, так что будьте бдительны.*

(4) *Сегодня лучше действовать сообща. Одному столько не выжить.*

(5) *На работе не будет происходить ничего особенного. Будете выполнять рутинные обязанности и мелкие поручения.*

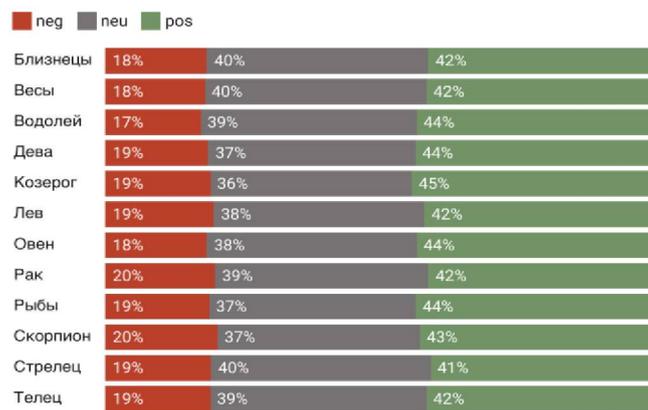


Рис. 3. Распределение гороскопов, сгруппированных в зависимости от подкорпуса, по тональности (в процентах)

Интересным показалось посмотреть, какое выражение принимает эмоциональная компонента в зависимости от цели текста. Рассматриваемые данные можно охарактеризовать по «поджанру» в соответствии с назначением при публикации заголовком (см. Таблицу 2): гороскопы общего характера превалируют (58,01 %), затем следуют любовные гороскопы (32,01 %) и бизнес-гороскопы (9,98 %). Как заметно по Рисунку 4, положительная лексическая тональность главным образом характерна для прогнозов в сфере любви. Обратившись к долям постов, сгруппированных по гороскопическим поджанрам и классам тональности, удалось подтвердить это наблюдение: так, 46,77 % любовных гороскопов и 42,22 % общего характера вне зависимости от знака зодиака имеют положительную «заряженность». Наиболее негативными, составляя 21,58 % постов своей подгруппы, в свою очередь являются бизнес-гороскопы (вероятно, так проявляется свойственный гороскопам этого поджанра мотив *предостережения о неприятностях на работе*), а наименее – вновь любовные (доля отрицательных постов внутри этой категории – 16,54 %).



Рис. 4. Распределение гороскопов, сгруппированных в зависимости от тематической категории и подкорпуса, по тональности (в процентах)

В результате проведенного исследования были выделены лексические и дискурсивные особенности астрологических прогнозов. Частотный анализ позволяет проследить лексику, тематически соответствующую архетипическим образам знаков зодиака, однако из-за высокой генерализованности предсказаний (наблюдается значительное количество лексических повторений) и слабых различий между тематическими группами этот анализ скорее говорит о том, что гороскопы имеет смысл рассматривать как некоторый речевой жанр с маловыраженной внутренней различимостью (по крайней мере, по параметру адресата, определяемого в соответствии со знаком зодиака). Представляется, что гороскоп в современном медиаполе функционирует как «аффирмирующее послание», на что указывает также и преобладающая положительная тональность этих текстов, но не как персонализированная рекомендация.

Благодарности. Статья подготовлена в результате проведения исследования в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ).

ЛИТЕРАТУРА

1. Wuthnow R. Astrology and marginality // *Journal for the Scientific Study of Religion*. 1976. P. 157–168.
2. Азначеева Е. Н. Астрологический дискурс: семиотический и когнитивный аспекты // *Вестник Челябинского государственного университета*. 2011. № 33. С. 19–21.
3. Вепрева И. Т. Базовый лексикон астрологического прогноза: о словах с "размытым" содержанием // *Экология языка и коммуникативная практика*. 2017. № 1. С. 57–63.
4. Бабаева Е. В. Лингвокультурологические характеристики русской и немецкой аксиологических картин мира: дис. ... д-ра филол. наук: Волгоград, 2004. 438 с.
5. Савицкайте Е. Р. Дискурсивные характеристики прогностических текстов (на материале немецких гороскопов): автореф. дис. ... канд. филол. наук: Воронеж, 2006. 24 с.
6. Лалетина А. О. "Гороскоп" как гендерно-маркированный медиажанр // *Жанры и типы текста в научном и медийном дискурсе*. 2007. С. 288–294.
7. Мубаракшина А. М. Дискурсивные особенности гороскопического текста // *Филология и культура*. 2021. № 4 (66). С. 101–105.
8. Фирак-Митц Ф. Астрологические коды. Москва: АСТ: Астрель, 2007.
9. MyStem 3.1 = морфологический парсер [Электронный ресурс] // Yandex. URL: <https://yandex.ru/dev/mystem> (дата обращения: 15.06.2025).
10. Bird S., Klein E., Loper E. *Natural Language Processing with Python*. O'Reilly Media Inc., 2009. URL: <https://www.nltk.org/book>.
11. Национальный корпус русского языка [Электронный ресурс]. URL: <https://ruscorpora.ru/stats> (дата обращения 01.07.2025).

12. Лукашевич Н. В., Левчик А. В. Создание лексикона оценочных слов русского языка RuСентилекс // Труды конференции OSTIS-2016. С. 377–382.
13. Orange3 Text Mining = software [Electronic resource] // Laboratory of Bioinformatics, Faculty of Computer Science, University of Ljubljana. URL: <https://orangedatamining.com> (date of access: 01.07.2025).
14. Lukashovich N. V. Automatic sentiment analysis of texts: problems and methods // Intelligent systems. Theory and applications. 2022. Vol. 26. №. 1. P. 50–61.
15. Hu, Minqing and Bing Liu (2004). Mining opinion features in customer reviews // Proceedings of AAAI Conference on Artificial Intelligence. Vol. 4. P. 755–760.
16. Колмогорова А. В., Колмогорова П. А., Куликова Е. Р. О прошлом, но в разное время: компьютерный анализ текстов учебников по истории СССР / России для шести поколений студентов // Вестник Томского государственного университета. Филология. 2024. № 89. С. 73–103. DOI:10.17223/19986645/89/4.
17. Ling Y. S. C. The use of modal auxiliary verbs in horoscope: a corpus-based study // Issues in Language Studies. 2016. Vol. 5. № 2.