

АНАЛИЗ ЛЕКСИКО-СЕМАНТИЧЕСКИХ ОСОБЕННОСТЕЙ РЕГИОНАЛЬНОЙ ПРЕССЫ (НА ПРИМЕРЕ ГАЗЕТ ГРОДНЕНСКОГО РЕГИОНА БЕЛАРУСИ)¹

Шайкевич А. Я. (lingstat@yandex.ru),

Савчук С. О. (savsvetlana@mail.ru)

Институт русского языка им. В. В. Виноградова
РАН, Москва, Россия

В статье приводятся результаты применения дистрибутивно-статистического анализа к корпусу газет Гродненского региона Беларуси. Было выделено три подкорпуса — районные газеты, городская газета «Вечерний Гродно» и комментарии читателей в «Вечернем Гродно». В каждом подкорпусе с помощью дистрибутивно-статистического метода были выделены списки маркеров, которые на основе лингвистического анализа удалось сгруппировать в кластеры, отражающие как тематические, так и стилистические особенности подкорпусов.

Для районных газет ведущими оказались маркеры, связанные работой местной власти, сельским хозяйством, охраной здоровья, охраной порядка и др. Большая группа маркеров определяет стиль текстов районных газет как официальный и книжный. В «Вечернем Гродно» наряду с темами, отражающими повседневную жизнь города, неожиданно на первый план по количеству маркеров выдвинулись темы, связанные с достопримечательностями города и его историей. В стилистическом отношении газете свойственна разговорность и диалогичность. Маркеры комментариев наследуют маркеры и кластеры маркеров из подкорпуса ВГ и демонстрируют логическое завершение основных стилистических тенденций газеты. Предложенный метод может быть использован для сопоставительного анализа других корпусов текстов.

Ключевые слова: дистрибутивно-статистический анализ, корпус региональных газет

¹ Работа выполнена при поддержке РГНФ, проект № 13-24-01004 (м)

DISTRIBUTIONAL-STATISTICAL ANALYSIS OF REGIONAL PRESS (NEWSPAPERS OF GRODNO REGION OF BELARUS)

Shaikovich A. Y. (lingstat@yandex.ru),

Savchuk S. O. (savsvetlana@mail.ru)

Vinogradov Russian Language Institute
of the Russian Academy of Sciences, Moscow, Russia

The paper is an application of distributional-statistical analysis (DSA) to the sub-corpora of Grodno region newspapers corpus. The sub-corpora under study are district newspapers, "The Evening Grodno" and commentaries to the latter. With the help of DSA hundreds of keywords have been elicited for each sub-corpus. The linguistic interpretation of those three lists showed that the keywords grouped into clusters reflect both thematic and stylistic features of the sub-corpora.

The district newspapers are specific in the choice of domains (mostly of local interest) and stylistic flavor (mostly official and bookish, to some extent resembling Soviet use). "The Evening Grodno" is more colloquial stylistically; its domains are naturally connected with the day-to-day city life and some topics which were unexpected, such as a large cluster of words denoting places of interest for tourists and inhabitants of the city. The keywords of the commentaries brings the stylistic trend of "The Evening Grodno" to its logical end. The method may be used for comparative analysis of other corpora, which might bring about new results depending on the composition of the corpus.

Key words: distributional-statistical analysis, regional press, newspaper corpus, Grodno region

Введение

В статье приводятся результаты исследования лексико-семантических и стилистических особенностей корпуса газет Гродненского региона Беларуси. Исследование проводится в рамках международного проекта, выполняемого коллективами ИРЯ им. В. В. Виноградова РАН и Гродненского государственного университета им. Янки Купалы. Цель проекта состоит в том, чтобы на материале газетных текстов выявить лексико-семантические и культурно-специфические особенности русской речи на территории Гродненского региона. В качестве экспериментальной базы используется создаваемый корпус региональных газет, который составит основу нового модуля в Национальном корпусе русского языка. В настоящее время в белорусскую часть корпуса входят 5 русскоязычных газет за 2012 год: городская газета «Вечерний Гродно»

и 4 районные газеты — «Берестовицкая газета», «Островецкая правда», «Ивьевский край», «Свислочская газета», общий объем корпуса составляет около 900 тыс. словоупотреблений. Российская часть корпуса формируется из областных и районных газет, а также региональных выпусков «Комсомольской правды».

Одновременно с составлением корпуса ведется поиск наиболее эффективных методов и приемов анализа материала. Как известно, в корпусной лингвистике принято разделение на corpus-based (CBA) и corpus-driven (CDA) подходы к изучению языковых данных. Мы предполагаем использовать оба подхода. Перспективы применения первого на основе инструментария, предоставляемого НКРЯ, изложены в [Кустова, Савчук 2013]. Ко второму подходу относится дистрибутивно-статистический анализ, опыт использования которого на материале гродненского корпуса излагается в настоящей статье.

1. Метод и процедура анализа

Существуют разные меры определения неслучайности концентрации той или иной лексической единицы в конкретном подкорпусе текстов, входящем в более широкий корпус. В настоящем проекте принимается путь, намеченный в публикации [A. Shaikevich, 2001, 229–255].

Как данное принимается следующая ситуация: существует какой-то корпус текстов (и соответствующий частотный словарь), в котором выделяется подкорпус (со своим частотным словарем). Зная долю подкорпуса в общем корпусе, мы можем подсчитать математическое ожидание (m) частоты какой-то лексической единицы в подкорпусе в предположении, что вероятность появления единицы не меняется по сравнению с общим корпусом (нулевая гипотеза). Затем реальная частота единицы в подкорпусе (f) сравнивается с математическим ожиданием, и в случае значительного расхождения двух величин делается вывод о неслучайности такого расхождения. Мера неслучайности (S) определяется по формуле:

$$S = \frac{f-m-1}{\sqrt{m}}$$

Для отрицательных значений S :

$$S = \frac{f-m+1}{\sqrt{m}}$$

При $S=2$ следует обратить внимание на данное слово, при $S=3$ возникнет подозрение в неслучайности отклонения, при $S=4$ подозрение превращается в уверенность.

Предположим, что корпус гродненских газет включен как подкорпус в общий корпус русскоязычных белорусских и российских газет объемом около 4 млн словоупотреблений. Доля гродненского подкорпуса составит 0.22, тогда степень специфичности следующих шести слов составит (F — частота в общем корпусе):

Слово	F	m	f	S
<i>без</i>	3802	836	842	0
<i>больница</i>	1067	234	236	0
<i>белорусский</i>	824	182	764	42
<i>ветеранский</i>	102	22	74	10
<i>бизнес</i>	1158	255	81	-11
<i>власть</i>	2098	462	191	-12

Вывод очевиден: частота слов *без* и *больница* ничем не отличается от общей нормы, слова *белорусский* и *ветеранский* крайне специфичны для гродненского подкорпуса, а частота слов *бизнес* и *власть* значимо меньше, чем в общем корпусе.

Будем называть лексическими маркерами те лексические единицы, реальная частота которых значимо превышает математическое ожидание (S превышает некоторый порог).²

В корпусе гродненских газет мы выделили три подкорпуса: районные газеты (РГ) (80% всего объема), «Вечерний Гродно» (18%) и комментарии читателей ВГ (2%). К ним применена та же процедура выявления маркеров. В подкорпусе РГ обнаружено 217 маркеров ($S \geq 2$), в «Вечернем Гродно» — 1080 маркеров ($S \geq 3$), в комментариях — 360 ($S \geq 2$). Расхождения в числе маркеров в значительной мере связаны с самим устройством нашей формулы. Она легко находит маркеры в подкорпусе, составляющем небольшую долю общего корпуса (скажем, менее четверти). Когда подкорпус составляет половину общего корпуса и больше, маркеры выделяются с трудом.

Чтобы обойти эту трудность и все-таки найти характерные черты подкорпуса РГ, прибегнем к двум приемам. Первый прием возможен при данной структуре изучаемого корпуса, в котором больший и меньший подкорпусы покрывают почти весь объем корпуса. Тогда отрицательные маркеры меньшего корпуса можно использовать в качестве кандидатов в положительные маркеры большего корпуса. Рассмотрим как пример слово *ввод*. Оно встретилось в РГ 59 раз и за их пределами не встретилось ни разу. В большем подкорпусе частота 59 дает S меньше 2, но в городской газете частота 0 делает слово отрицательным маркером и тем самым кандидатом в маркеры подкорпуса РГ. Такими же кандидатами становятся *божественный*, *ветеранский*, *вклад*, *возглавлять*, *воинский*, *воспитанник*, *вправе*, *встреча*, *выборы* и многие другие слова.

Второй способ пополнения маркеров возможен при любой структуре общего корпуса. Он сводится к группировке нескольких слов, что увеличивает реальную частоту и может вести к повышению S . Слова *ВИЧ*, *ВИЧ-инфекция* и *ВИЧ-инфицированный* порознь не обладают частотой, достаточной для $S=2$, но вместе эти три слова набирают частоту 141 в РГ (при 145 во всем корпусе), значение

² Строго говоря, их следовало бы называть положительными лексическими маркерами. Отрицательными лексическими маркерами можно называть слова, чья частота существенно ниже математического ожидания, но их в корпусе обычно не так много, и они хуже интерпретируются содержательно.

С превысит 2 и сделает группу полновесным маркером.³ Подобным образом маркерами РГ станут *выяв-ить, животновод-ство, заготов-ить, информ-ация, кредит-, назнач-ить, налог-, нарко-тик, необходим-о, поруч-ить, страхов-ой, удел-ять, уплат-а, Христ-ос*. Так же пополняется круг кандидатов в маркеры: *горд-иться, достав-ить, предостав-ить, прокур-ор, эффективн-ость* 4. Группировку кандидатов в маркеры будем ниже давать в круглых скобках, заключая список указанием частоты в изучаемом подкорпусе на фоне остальных двух подкорпусов.

Оглядываясь постоянно на тексты, мы можем сводить маркеры в некоторые кластеры, открывающие для нас тематическую и стилистическую специфику изучаемого подкорпуса⁵.

2. Анализ результатов: маркеры в районных газетах

Естественно предположить, что в районных газетах будут представлены маркеры, семантически связанные именно с данным **административным образованием**: *районный* $f=1553$ $S=8$, *район* $f=2260$ $S=6$, *сельский* $S=4$, *агрогородок* $S=3$ *совет* $S=3$ *административный, населенный, отдел, пункт, сельсовет, территория (межрайонный, муниципалитет, полномочия, поселковый, село* 226:9). К этому же кластеру следует отнести жителей района с их жильем и контактами с администрацией: *работник* $f=1030$ $S=6$, *заявление* $f=431$ $S=3$,

³ Для краткости одно из слов такой группы будем делать ее представителем, выделяя в нем основу. Так за символом *выплат-а* будут скрываться четыре слова — *выплата, выплатить, выплачивать, выплачиваться*.

⁴ Группировку кандидатов в маркеры будем ниже давать в круглых скобках, заключая список указанием частоты в изучаемом подкорпусе на фоне остальных двух подкорпусов.

⁵ Такой анализ необходимо проводить при постоянном обращении к текстам, чтобы избежать ошибок при классификации многозначных лексем, а также контролировать влияние привычных ассоциативных связей. Так, например, первоначально предполагалось, что слово *сердце* имеет отношение к кластеру «медицина», а *минировать* — к «военной истории». Однако в ходе проверки выяснилось, что *сердце* (в районных газетах $f=229$ $S=2$) употребляется почти исключительно в метафорическом значении (*чистые сердца, доброе сердце, зов сердца, найдется место в сердце, прикипела душой и сердцем* и под.), и поэтому не годится на роль «медицинского» маркера. А слово *минировать* (в ВГ $f=8$, $S=4$) вообще, как выяснилось, к войне не имеет отношения, поскольку все 8 вхождений встретились в составе терминологического наименования *минирующая моль*.

⁶ Высокий показатель маркера *сельский* связан, скорее всего, с тем, что в подкорпус РГ попали газеты четырех районов с преимущественно сельским населением.

⁷ Слова *агрогородок* (в Белоруссии это официальное название одного из типов сельских населенных пунктов) и *агроусадьба* (объект экологического туризма) имеют выраженную региональную специфику, которая проявляется в различии частотных характеристик. Так, в белорусских газетах в составе регионального корпуса слово *агрогородок* встретилось 202 раза (на 0,9 млн словоупотреблений), а в российских — лишь 2 раза (на 9 млн); в корпусе современных российских СМИ (173 млн) зафиксировано всего 19 вхождений этого слова.

подача =3, жилищ-е, жилой, житель-, помещение, проживать, уплат-а (односельчанин, паспорт, пустующий, сельчане 299:15)

Но среди маркеров РГ представлены все звенья вышестоящей **власти** (кроме самого слова власть) со всеми их функциями: организация f=1321 S=5, республика f=1193 S=5, государственный S=3, законодатель-ство S=3, исполнительный S=3, ответствен- S=3, учреждение S=3, документ, закон, исполнение, комитет, контроль-, меры, назнач-ить, пленум, поруч-ить, проверка, республиканский, управление, установленный (декрет, должност-ь, заседание, надзор, нормативный, президиум, распорядиться, распоряжение, рассмотреть, регламентировать, регулировать, указ 896:47). Персонификация власти отражена в маркерах руководитель f=458 S=5 и председатель f=983 S=4 (возглавлять, глава, управляющий 244:17). Индивиды обозначены здесь маркерами граждане f=1035 S=6, лицо f=738 S=3, гражданин.

Обобщенные маркеры «**НИЗОВ**» — население f=603 S=3, профсоюз f=228 S=3, ОО [общественное объединение], «Белая Русь» (первичка 34, первичный 124:3). От «верхов» исходит информация — информ-ация, обращение, проинформировать; информация сопровождается другими средствами убеждения, часто идеологизированными — акция f=319 S=3, мероприятие f=555 S=3 (агитбригада, акцентировать, идеолог-, комсомольский, месячник, нравствен-ый: октябренок, октябрятский, оповещение, опубликование, пионерия, подчеркнуть, разъян-ить, убеждать 507:19). Особое значение имеет воспитание патриотизма — ветеран- f=364 S=4 (афган-ский, беззаветный, воин-, горд-иться, мемориальный, отечество, партизанский, патриотический, подвиг, преданность, пьедестал, увековечение, фронтовой 540:24).

Не только слова, но и **материальные блага** идут сверху вниз: социальный f=619 S=4, обеспеч-ение f=651 S=4, пенси-я S=3, выплат-а, заработ-ок, плата, пособие, предостав-ить (льгот-а, надбавка, нетрудоспособн-ость, оклад, перерасчет, престарелый 378:22). Благожелательство «верхов» выражено маркерами благополучие и удел-ять (благосостояние, поощр-ить, процветание, удовлетворение, чутк-ий 171:0).

Характерная черта советского быта, сохранившаяся в РГ, — **приуроченность** государственной благотворительности к **каким-то датам**: праздн-ик f=765 S=3, наград-а f=321 S=3, поздрав-ление f=332 S=3, благодар-ность, вруч-ить, грамота, подарок, почетный (выраз-ить, годовщина, заслуг-а, искренн-ий, преддверие, предпраздничный, предстоящий, премия, приветств-овать, признательность, приуроченный, торжественный, удостоить, чествова-ние, юбиляр 1119:43).

Обратная связь снизу обеспечивается **готовностью масс к действию** (маркер активн-ый f=362 S=3) и участием в выборах: выдвинуть f=273 S=3, депутат f=298 S=3, избиратель, кандидат (выборы, голосование, делегат, отчет-ный, предвыборный, проголосовать, созыв 408:3). Главное же для «низов» — **труд на благо общества**: работа f=2798 S=7, труд f=688 S=5, трудовой S=4, коллектив (бескорыстный, вдохновлять, дисциплина, добросовестный, передовик, потрудиться, самоотверженный, слаженн-о, созидательный, сплоченный, старательный, терпение, труженик, умел-ый 429:5).

Общий итог выражен маркером **порядок** (*стабильн-ый, устойчивый* 66:2).

В районах Гродненской области, представленных в корпусе РГ, основной сектор — **сельское хозяйство**, поэтому вполне ожидаем широкий круг соответствующих маркеров: *зерно- f=420 S=4, сельскохозяйственный S=3, агро-промышленный, гектар, животновод-, заготов-ить, комбайн, совхоз, участок*. Из длинного перечня кандидатов в маркеры назовем самые частые — *корм, молоко, зимний, посев, корова, скот, озимый, механизатор* (всего 80 слов, 2683:76).

Но не только сельское хозяйство, вся **экономика** образует мощный кластер маркеров в РГ. Просматривая этот список, так и слышишь оптимистические экономические отчеты на съездах КПСС: *хозяйство f=851 S=5, выполн-ить f=621 S=4, осуществ-ление f=409 S=4, размер S=3, текущий S=3, величина, достижение, задача, итоги, материальный, млн, ОАО, обслуживание, объединение, период, показатель, приобретение, продукция, произвести, производственный, процент, результат, рост, сроки, темп, товар, услуги, цель*. Список кандидатов включает 30 слов (например, *валовой, обязательство, отрасль, совершенство-вание, ускорение*; 1264:31). Рядом находим **финансовый** кластер, кажущийся более современным: *бюджет- f=258 S=3, денежный, имущество, кредит, налог, руб., страхов-ой, сумма (задолженность, затраты, израсходовать, наниматель, оборот, пеня, плательщик, подходящий, расходы, ревизия, рента, ссуда* 480:25).

Максимум эмоциональности сосредоточен в сверхкластере маркеров РГ, сводимом к трем словам — **семья, дети, школа**. *Семья f=975 S=4, брак, рождение; дети f=1374 S=3, возраст S=3, воспита-ние, ребенок, родитель; профессиональный f=468 S=4, школа f=923 S=3, образование f=625 S=3, учащийся S=3, филиал S=3, базовый, знание, культура, лицей, молодежь, педагог-, подросток, профессия, ребята, учитель*. Длинный список кандидатов в маркеры содержит очень разные слова: от официальных (*алименты, безнадзорный, времяпрепровождение*), высоких (*напутствие, наставник, юный*) до умиленно-ласкательных (*деткишки, детвора, ребяташки*).

Здесь же примыкает и **физкультурный** кластер: *СШ [спортивная школа] f=384 S=3, команда, победитель, турнир, борьба (велопеход, гиревой, гребля, ДЮСШ, купаться, мяч, перетягивание, пионербол, подопечный, соперник, спартакиада, стритбол и т. п., всего 26 слов 684:10)*. Особый кластер в РГ образует единственный маркер **ХРИСТ-ос** с очень длинным списком ассоциированных слов (*акафист, божественный, вечерня, воскреснуть, духовн-ый, молебен, Пасха, прихожане и т. п. всего 35 слов 867:10*).

Два кластера угрожают жизни гродненских районов — **болезни и преступники**.

Кластер «**охрана здоровья**» содержит маркеры *ВИЧ, здоровый, здоровье, нарко-тик, профилактик-а* и множество кандидатов в маркеры (*алкоголизм, бюл-летень, вирус, инфекция, клещ-, курение, очаг, сыпь и т. п., всего 26 слов 971:64*).

Неожиданным для нас оказался громадный кластер «**охраны порядка**». Казалось бы, эта тема присуща всем СМИ, однако именно районные газеты выдвигают ее на первый план: *безопасность f=459 S=4, несовершеннолетний S=3, органы S=3, внутренних, возмест-ить, кодекс, милиц-ия, оперативн-ый, правонарушение, преступ-ный, РОВД, уголовный* (более 50 кандидатов в маркеры, например *антиобщественный, арест, ДТП, кража, МВД, наказание, обвиняемый,*

опьянение, пресеч-ение, причин-ить, прокур-ор, раскрываемость, самогон, след-ственный, ст. и т. п. 2119:67).

Перечисленные выше тематические кластеры маркеров не исчерпывают специфики подкорпуса РГ. Остается большая совокупность **стилистических маркеров**, тяготеющих одновременно к официальности и книжности: *внимание S=4, качество S=4, внес-ти, вопрос S=3, проводится, путем, являться S=3, данный, действовать, иметь, иметься, иной, категория, необходим-о, оказание, определ-ить, отношение, принять, провед-ение, в соответствии, соответствующий, степень, в сфере, участие, учет, в ходе.* Остается еще 36 кандидатов в маркеры (например, *аспект, вследствие, вышеназванный, значим-ый, ибо, исходя, каков, констатировать, контекст, менталитет, насыщенный, поприще, совокупный, спектр, столь, сторицей 1309:65*). Лишь одно слово выбивается из этого перечня — слово *районка*, 14 раз встретившееся в подкорпусе РГ.

3. Анализ результатов: маркеры в газете «Вечерний Гродно»

Анализ маркеров газеты «Вечерний Гродно» проводился аналогичным образом. Напомним, что тексты ВГ составляют не более 20% всех текстов совокупного корпуса гродненских газет, поэтому в результате применения формулы было получено более 1080 положительных маркеров с величиной отклонения $S \geq 3$. Следующим шагом было объединение маркеров в кластеры на основе семантического анализа.

Как и в подкорпусе районных газет, ожидаемым стало выделение группы маркеров, связанных городом Гродно и его жителями. Однако в отличие от районных газет город в ВГ представлен не как муниципальное образование с административным аппаратом, а как живая среда обитания, со своей инфраструктурой и организацией жизни горожан — их быта, труда и досуга.

Значительным по объему оказался кластер «**городская застройка и жилье**»: *ул-ица S=16, микрорайон S=6, проспект 6, переулок, планировка, застройка, квартал, спальный и др., здание S=8, дом S=4 (строение S=4, постройка, фасад S=8, двухэтажный S=5, трехэтажный, крыша S=5, комната S=4, хрущевка, подвал S=5, лифт S=4, окно и др.)*. В городе выделяются **городские зоны** разного назначения: *центр S=4 и окраина, окрестности S=4, зоопарк S=5, парк S=9, ограждение, лесопарк S=5, рынок S=7, магазинчик S=4, салон, гостиница S=6*. Самым многочисленным оказался субкластер «**транспорт**»: *автобус- S=11, троллейбус S=4, билет S=10, автовокзал, маршрут S=4, рейс S=4, остановка S=4, пересадка S=5, стоянка, такси, вокзал S=8, железнодорожный S=5, поезд S=5, вагон, авто, машина S=10, грузовик, мотоцикл- S=7, велосипед-, парковка S=6, ГАИ, поворот, транспорт, разметка, развязка S=6, светофор, трасса S=6, шоссе*.

«Вечерний Гродно» отличается от районных газет по двум параметрам. Первое отличие в целевой аудитории: газета ориентируется на жителей крупного города, промышленного и культурного областного центра, а читателем районных газет является житель райцентра или сельского поселения. Вторых, в отличие от остальных, ВГ — это газета, предназначенная для чтения

на досуге и потому в ней меньше пропагандистских материалов, и больше познавательных, просветительских и занимательных. Помимо городских новостей читателя привлекают темы искусства, досуга, в том числе путешествий, истории края и др. Все эти темы богато представлены в корпусе ВГ, что проявляется в составе маркеров.

Сильной стороной газеты ВГ является краеведческая тематика. Публикации под рубрикой «Красная книга Гродно» — очерки об истории и архитектуре Гродно — можно считать «брендом» газеты. Эти публикации — основной источник маркеров кластера **«история»**: *истор-ия* — S=8, *стар-ый* S=16, *сохраниться* S=6, *век* S=11, *столетие* S=4. Маркеры этого кластера характеризуют Гродно как город с богатым историческим наследием. Крупный субкластер связан с описанием архитектурных **исторических достопримечательностей** города: *архитектур-ный* S=4, *достопримечательность* S=4, *памятник, замок* S=15, *средневековый* S=5, *дворец* S=10, *крепость, баш-ня* S=9, *особняк, домик* S=7, *мельница* S=4, *колокольня* S=4, *казарма* S=8, *конюшня, мостовая, мост* S=11, *кованый* S=4, *кладбище* S=4 и др.

Еще более детально представлена **история в событиях и лицах**. Примечательно, что большинство маркеров в этой группе связано с военными действиями, победами и поражениями: *княжество* S=4, *князь* S=10, *Витовт, король* S=14, *корона* S=6, *Карл* S=9, *Сапега, Речь Посполита* S=5, *Наполеон* S=12, *австро-венгерский, повстанцы, штрафник, шпион, царский, войска* S=5, *улан, гусар, полк* S=9, *орудие* S=6, *пушка* S=6, *война, бой* S=5, *сражение* S=6 (*атаковать, штурмовать, переправа, отступ-ать, разбит-ы*), *узник, гетто* и др. Среди других исторических реалий *магистрат* S=5, *купец* S=6, *Ожешко* S=11.

Кластер **«искусство и культура»**, многосторонне отражает культурную жизнь города. В нем выделяются субкластеры **«театр»**: *театр-, актер, артист, режиссер, комедия, драм-а, драмтеатр, кукла, кукольный, роль, спектакль, сцена*; **«кино»** S=11, *кинотеатр* 6, *серия, фильм* 12, *экран, кадр* 4, *камера*; **«музыка, пение, танец»**: *музыкант, певица, танец, танцор, филармония, хореограф-, студия*; **«литература»**: *писатель-* 6; **«изобразительное искусство»**: *скульпт-ура* S=6, *фотограф-ия* 7, *экспозиция, художник* 9, *вернисаж, галерея* 6, *музей* 7; кроме того *поклонник* 5, *публика, зритель* 8, *персонаж* 6, *сюжет* и др.

Продолжением этого кластера будут маркеры, объединенные более широким понятием **«досуг»**: *развлечение* S=4, *розыгрыш* S=5, *зрелище, аквапарк, бассейн, вечеринка, бар, кафе* S=8, *игра-* S=4, *клубы* (по интересам), *турист-* S=10, *путешеств-ие* S=4, *экскурсовод, яхта* S=5 и др.

Путешествия, а также приграничное положение Гродненского региона обуславливают значительную долю материалов на международные темы. Это и впечатления о поездках, и реклама туров, и отчеты о международных конкурсах и деловых встречах, и аналитические публикации о жизни соседей. Поэтому кластер **«зарубежье»** достаточно представителен: *Россия* S=4 (*российский* 4, *россияне*), *Поль-ша* S=7 (*польский* 11, *Варшава* 4), *Литва* S=7 (*литовский* 6, *виленский* 7, *Вильнюс, Вильно*), *Европа* S=11 (*европейский, Евросоюз* 4), *Германия* S=6, *немецкий* 5, *Франц-ия* S=7, *Америк-а* S=9, *Итал-ия* S=8, *шведский, армянский, Лондон* S=4, *Турция* S=4, *виза* S=6.

Как представлена повседневность — работа, учеба, быт, семья, дети, забота о здоровье и охрана общественного порядка — в корпусе ВГ? Понятно, что эти темы не уникальны и занимают значительное место в других газетах. В чем специфика их разработки на страницах ВГ?

В отличие от районных газет практически отсутствует тема **«взаимодействия с властью»**. В состав соответствующего кластера собраны маркеры, отражающие наиболее доступный рядовому обывателю уровень власти и формы общения с ней: *горисполком S=11, облизполком, мэр, анкета, бланк, заявка, запрет, очередь* и др.

Тема общественно-полезного труда также мало представлена в составе маркеров. Не велик кластер, связанный с промышленным или сельскохозяйственным **«производством»**: *завод- S=9, комбинат, фабрика 5, пивзавод 5, мастерская 4, агроусадьба 4*. Часто эти маркеры встречаются в текстах с исторической, экологической тематикой.

Зато маркеры, связанные со сферой **«предпринимательства»**, представлены в большом количестве: *бизнес, бизнесмен S=4, бренд, деньги 7, доллар 11, евро 13, ЗАО 5, фирма 5, фирменный 4, инвест-ции 8, торги, маркетинг, менеджер, распрод-ажка 6, рекламный 17, купон 9, сертификация, скидка 5, стоимость 4, стоить 6, супермаркет, част-ный 4*.

Удивительно, что в отличие от районных газет, **криминальный кластер** ВГ не заполнен: единственный претендент на роль маркера — *насилие* встречается в тексте о домашнем насилии, разъясняющем и комментирующем статьи закона. Правда, как будет показано ниже, криминальная тема получает некоторое развитие в комментариях читателей ВГ.

Повседневная жизнь не может обойтись без темы **«кулинария»**, чрезвычайно популярной во всех СМИ: рецепты приготовления блюд, кухонная утварь и техника, оценка качества продуктов и учреждений общепита, дегустация блюд, кулинарные конкурсы и мн. др. Этот кластер заполнен и в ВГ: *кухня, мясо, рыба S=5, кофе 4, картошка, спирт, сыр, торт, хлебный, лук, чеснок* и др. Сюда же можно включить и учреждения общепита: *ресторан, кафе, кофейня*.

Актуальной для горожан является тема утилизации бытовых **отходов** и связанная с ней экологическая тема: *отходы S=7, мусор S=4, контейнер, раздельный, полигон, макулатура*.

Наконец, всегда востребованы такие темы, как забота о здоровье и здоровом образе жизни, воспитание и образование детей. Кластер **«здоровье и медицина»** велик и разнообразен по составу, что свидетельствует о разных аспектах разработки этой темы в публикациях газеты: *врач — пациент, профилактика — лечение, лекарства и методы лечения, медицинские учреждения и управление медициной: больница S=7, клиника 5, поликлиника 11, врач 5, пациент 5, диагноз, инфаркт, курорт- 6, санаторий, фитнес 5, гигиена, лекарство 6, лечебный, лечиться, медучреждение, Минздрав 5, узи 4, препарат 8, антибиотики 5, целитель* и др.

Что касается образовательного кластера, то в нем преобладают маркеры, связанные с университетским образованием: *ГрГУ S=8, доцент, кафедра, лаборатория 6, ректор, студенческий 4, университет 6, ученый 5*.

Здесь мы затронули лишь тематические кластеры, однако серьезного внимания заслуживают группы маркеров, в которых проявляются стилистические различия между ВГ и районными газетами. В частности, огромная группа оценочной лексики (*вредный, громкий, грустный, заброшенный, знакомый, знаменитый, идеальный, известный, могучий, мрачный, небольшой, невероятный, необычный, неприятный* и мн. др. — более 100 единиц) выделяется в качестве маркеров в ВГ именно потому, что в РГ оценка практически отсутствует. Это свидетельствует об относительной эмоциональной свободе и раскованности повествования, эксплицитном выражении авторской оценки. О более широком спектре и разнообразии модальности текста, помимо разъяснения, убеждения и долженствования, присущих районным газетам, свидетельствует большая группа глаголов речи (*рассказывать, говорить, объяснять, пояснять, обещать, советовать*, и др.). Группы слов, обозначающих «положение в пространстве» и «восприятие», можно рассматривать как стилистический маркер, связанный с присутствием в корпусе ВГ значительного количества текстовых фрагментов в регистре описания — пейзажей, строений, интерьеров (*внутри, возле, рядом, около, вокруг, вдоль, размещать-ся, располагаться, расположиться, стоять, виден, смотреть-ся, увидеть* и др.).

Что касается комментариев к ВГ, то они, конечно, наследуют маркеры и кластеры маркеров из подкорпуса ВГ, часто увеличивая долю глаголов. Наряду с *автобус, перекресток, пересадка, светофор* в кластере «**транспорт**» появляются *сходить, возить, ездить, поехать, пробка*. Разрастается кластер «**больница**» *врач, пациент, медицина, леч-ить, персонал, рожать, маммолог, морг, больной, лежать*. Появляется и новый кластер «**бедности**»: *зарплата, нищета, зарабатывать, бесплатный, дешев-ый, платить, продать*.

Очевидно должен появиться кластер «**переписки с газетой**»: *статья S=22, я 21, вы 13, написать 12, автор, комментарий 9, пожалуйста, уважа-емый 5, журналист, мой 4*. За ним следуют кластеры «**одобрения**» (*молодец S=18, спасибо 16, нормальный 12, правильно, супер, хорош-о 8, здорово 6, умница 5, великолепный, наконец-то 3*) и несогласия (*пиар S=19, ненависть, плох-о 10, бред, бардак, быдло, жаль, злоба, позор, тупой, урод 7, гадость 5, грустно, плевать 4, глупый, зря 3*). Обоснование оценки мы видим в маркерах *думать S=13, знать 13, подумать 6, понимать 5, видно, забыть, логично, умный, ум, ясно 3*.

В отсутствие цензуры в этом подкорпусе становятся маркерами и такие единицы, которые в двух других подкорпусах встречаются редко (*гей S=14, аборт 12, взятка, травести 8, гомофобия, отсидеть 7, критиковать, ориентация 4, забастовка 3*) или не встречаются совсем (*власть имущие, гетеро 7*).

Переход к живой речи, отмеченный в ВГ, здесь проявляется в максимальной степени как в маркерах диалога и местоименности (*бы S=19, там 15, зачем 10, будет, вот, нечего 9, если, кто туда 8, сколько 7, какой, откуда 6 и т. п.*), в общем тяготении к разговорности (*надо S=14, а, ага, ну, так 12, то, хотеть(ся) 10, очень 9, нельзя, такой, уж, прост-о 8, вообще, касательно 7, лично 6*), так и в специфических разговорных словах (*баксы S=10, мужик 10, телефончик 7, бюджетник 6, куча, товарищ 4*). Заметим, что минимальная доля данного подкорпуса в общем корпусе позволила стать маркерами словам, лишь двукратно

появившимся в тексте: *абсурд, додуматься, идиот, комменты, контрацепция, норовить, солярочка, сосать, феминистка, хамство, ханжить, хреново.*

4. Заключение и выводы

Применение метода дистрибутивно-статистического анализа на корпусе газет Гродненщины показало положительные результаты. Статистическая обработка корпусов с выделением маркеров и последующий лексико-семантический анализ соотношения маркеров выявляет существенные расхождения между содержательным, тематическим наполнением текстов и их стилистическим оформлением. Эти расхождения коррелируют с теми отличительными признаками районных и центральных газет, которые выделяются на других основаниях и на другом материале (см., например, Лысакова 1989, Федотова 2001, Кислая 2008). Кроме того, они позволяют выделять специфические особенности отдельных изданий, например, «краеведческую доминанту» в газете «Вечерний Гродно». Следовательно, метод может быть использован для проведения дальнейших исследований на корпусах региональных газет. Проверку этой гипотезы, а также выводов относительно специфики различных изданий, сделанных в ходе исследования, необходимо провести, используя другое соотношение корпусов. В дальнейшем мы планируем провести аналогичные исследования для того, чтобы: 1) попытаться выявить особенности районных российских газет на фоне региональных и городских, 2) российских и белорусских вечерних газет на фоне ежедневных; 3) сравнить российские и белорусские районные газеты между собой, 4) сравнить корпус российских региональных газет в целом и гродненский корпус.

Литература

1. *Кустова Г. И., Савчук С. О.* Изучение лексико-семантической и социокультурной специфики русской речи на территории Республики Беларусь (на материале текстов СМИ) // Труды Международной конференции «Корпусная лингвистика — 2013». Санкт-Петербург, 2013. С. 344–353
2. *Кислая Л. Н.* Редакционная политика районной прессы: на примере газет Новосибирской области. АКД. Екатеринбург, 2008
3. *Лысакова И. П.* Тип газеты и стиль публикации: Опыт социолингвистического исследования. Л.: Изд-во ЛГУ, 1989
4. *Федотова Л. Н.* Анализ содержания — социологический метод изучения средств массовой коммуникации. М.: Научный мир, 2001.
5. *Шайкевич А. Я., Андрющенко В. М., Ребецкая Н. А.* Дистрибутивно-статистический анализ языка русской прозы 1850–1870-х гг. Т. 1. М.: ЯСК, 2013
6. *Shaikevich A.* Contrastive and comparable corpora: quantitative aspects, *International Journal of Corpus Linguistics*, vol. 6(2), 2001, p. 229–255.

References

1. *Kustova G. I., Savchuk S. O.* (2013), The study of lexical-semantic and socio-cultural specifics of the Russian language on the territory of Belarus (on the material of mass media texts) [Izuchenie leksiko-semanticheskoy i sociokul'turnoj specifiki russkoj rechi na territorii Respubliki Belarus' (na materiale tekstov SMI)], Proceedings of the International Conference “Corpus linguistics—2013” [Trudy Mezhdunarodnoj konferencii “Korpusnaja lingvistika—2013”], St. Petersburg, pp. 344–353
2. *Kislaja L. N.* (2008), The editorial policy of the district press: the newspaper of Novosibirsk region [Redakcionnaja politika rajonnoj pressy: na primere gazet Novosibirskoj oblasti]. Abstract of PhD dissertation, Ekaterinburg
3. *Lysakova I. P.* (1989), Type of newspaper and style of publication: sociolinguistical research [Tip gazety i stil' publikacii: Opyt sociolingvisticheskogo issledovanija]. Izd. Leningradskogo universiteta, Leningrad.
4. *Shaikevich A.* (2001), Contrastive and comparable corpora: quantitative aspects, *International Journal of Corpus Linguistics*, vol.6(2), 2001, p.229–255.
5. *Shaikevich A. A., Andrjushhenko V. M., Rebetskaja N. A.* (2013), Distributional-statistical analysis of the language of Russian prose of 1850–1870-ies [Distributivno-statisticheskij analiz jazyka russkoj prozy 1850–1870-h gg.], *Jazyki russkoj kul'tury*, Moscow
6. *Fedotova, L. N.* (2001), Content analysis as the sociological method of studying mass communication [Analiz sodержanija—sociologicheskij metod izuchenija sredstv massovoj kommunikacii]. *Nauchnyj mir*, Moscow.