***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

Новые компьютерные технологии для исследований языка и сознания

© кандидат филологических наук Г. Ю. Никипорец-Такигава, 2006

Рассуждения о преимуществах использования инструментов точного компьютерного анализа, которыми располагает «Интегрум», для объектов гуманитарного знания, требуют предварительного рассмотрения нескольких концептуальных вопросов:

1. Каковы современные возможности доказательства гуманитарной идеи при помощи точных компьютерных методов анализа?
2. Чем использование «Интегрума» отличается от использования корпусов русского языка и таких Интернет-ресурсов как Yandex, Rambler или Google?
3. Приближают ли нас точные методы анализа к истине в большей степени, чем «неточные»?

Представляется, что вопрос о том, точные или неточные методы анализа мы избираем, имеет отношение не к цели достижения истины, а к способу приближения к истине. Так, например, сбор картотеки может быть осуществлен двумя способами. Можно:

- определить количество данных, на основании которых можно обосновать гипотезу (например, 100 или 1000);
- стремиться охватить как можно больше данных.

Второй способ выбирают исследователи, которым для решения их класса задач требуется максимально возможное количество данных. Или исследователи, которых не удовлетворяет абстрактность методологии гуманитарного исследования, выражающаяся, в частности, в допущении, что 1000 примеров достаточный материал для формулирования выводов. Таким исследователям эмпирическую базу непременно хочется увеличить.

Безусловно, компьютер предоставил такую возможность. Картотеку можно собирать в корпусах, таких, как Национальный корпус русского языка, корпус университета Тюбингена и др. Для поиска примеров и определения частотности слова широко используется и Yandex наряду с другими Интернет-ресурсами. Однако как только гуманитарии овладели компьютерными технологиями, радость от общения с новым медиумом сменилась осознанием недостатков. Стало очевидно еще и другое: с той

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

же скоростью, с какой развиваются Интернет и компьютерные технологии, наши методы использования компьютера для сбора данных устаревают и требуют замены, соответственно, мы оказываемся на том этапе компьютеризации научной мысли, когда поиск в Яндэксе – вчерашний день.

Что касается корпусов русского языка, то они морфологически и синтаксически размечены, однако тематически и количественно (65 миллионов слов в РНК, 25 миллионов в Тюбингене) ограничены, что делает их непригодными для решения ряда задач. Yandex, Rambler, Google количество материала не ограничивают, предоставляя массив текстов для обработки, многократно превышающий тот, который филолог анализировал двадцать лет назад. Но при использовании Интернета неизбежны следующие проблемы:

Первая проблема: проблема качества и достоверности информации. Сбор картотеки «руками» из печатных источников не ставил проблему неаутентичности текста. Поиск в Интернете не гарантирует соответствия электронной копии оригиналу, так как размещение документов в Интернете происходит почти бессистемно и бесконтрольно. Быстро растущий объем Интернета и количества информации в нем приводит к тому, что его поисковое пространство превращается в огромную «кучу» «плевел» и для того, чтобы отделить зерна, требуется много времени и усилий.

Вторая проблема заключается в выборе источников. Огромный массив материалов заставляет решать проблему выбора релевантных для данного анализа источников. Так остро проблема выбора источников ранее не стояла из-за их несравнимо меньшего числа. Поиск в Интернете – это поиск во всем пространстве Интернета, при котором ограничение области поиска на этапе запроса невозможно и выбор нужных документов превращается в длительную ручную процедуру отбора тех самых «зерен» из выдачи.

Третья проблема: отсутствие в Интернете полнотекстовых архивов средств массовой информации, что затрудняет сравнительный и исторический анализ.

Четвертой проблемой является язык запросов. Исследователя, использующего Yandex, Rambler или Google для научных целей, обычно сначала подкупает известная легкость в общении. Гуманитариям плохо известно, как оптимизировать запрос, и отзывчивостьYandex, Rambler или Google при ответе на самый «человеческий» запрос нас радует. Если речь идет о бытовых проблемах, то Интернет великолепно удовлетворяет любопытство. Вопрос: «Какая сегодня погода?» – может быть

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

задан и собеседнику, и Интернет-поисковой машине – она ответит точнее.

Тем не менее, исследователя, который хочет на основании примеров, найденных в Интернете при помощи Интернет-поисковых машин, строить систему доказательств, очень скоро перестает удовлетворять процент шума и невозможность снятия некоторых принципиальных затруднений, например, омонимии. Так, поставив задачу поиска стереотипов, которые наиболее активно разрабатываются в русской языковой картине мира в связи с японским характером, японцами, Японией во всем пространстве Рунета, исследователь сталкивается с тем, что Интернет-поисковые машины не могут исключить из выдачи контексты, в которых **японец** оказывается машиной (автомобилем, грузовиком, вездеходом, джипом), бытовой техникой, электроникой и так далее.

По сравнению с Интернет-поисковыми машинами «Интегрум» предложил принципиально новое поисковое пространство и новые средства поиска.

«Интегрум» старался решить **первую проблему**: проблему качества информации, так как искал не в Интернете, а в собственных базах данных, составленных из наиболее надежных источников[1].

«Интегрум» решил **вторую проблему** выбора релевантных для анализа источников. Поиск и исследования при помощи «Интегрума» можно осуществлять в ограниченной по желанию исследователя на этапе запроса области: во всех СМИ или в группах СМИ («Центральные газеты» (ЦГ), «Центральные журналы» (ЦЖ), «Региональная пресса» (РСМИ), «Теле-Радио эфир» (Т-Р), «Интернет-издания», «Компьютерная пресса») или в одном выбранном источнике (электронной версии одного СМИ), а также в «Библиотеке всемирной литературы» и т. д.[2].

Как и Интернет-ресурсы, «Интегрум» дал гуманитариям возможность исследовать русскоязычные массивы информации, состоящие из сотен миллионов документов[3], однако, в отличие от Интернета, содержал **полные архивы тысяч СМИ**, большинство из которых начиналось с середины 90-х годов XX века. Это позволило собирать данные на

[1] Информация поступает в «Интегрум» от издателей, которые стараются соблюдать определенные соглашения по форматам, достоверности информации, актуальности и тематической привязке. Источники, главным образом, издания прессы, описываются редакторами.

[2] См. об этом Приложения к книге «Интегрум: точные методы и гуманитарные науки». Сб. статей / Под ред. Г. Никипорец-Такигава. (В печати.)

[3] На начало 2006 года в «Интегруме» имелось 6000 баз данных, состоящих из 350 миллионов документов.

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

большом качественном материале и анализировать период продолжительностью десять-пятнадцать лет.

И, наконец, «Интегрум» **имел точную поисковую систему.** «Артефакт» (информационно-поисковая система, которая обрабатывает базы «Интегрума») располагает разветвленным языком запросов со сложным синтаксисом (операторами «или» для изучения группы, «не» для снятия омонимии и других ограничений запроса, «*» и «!п» для того, чтобы система искала все слова, начинающиеся с определенной комбинации букв или полную парадигму слова, «/п» и «:» с последующей цифрой для ограничения поиска указанной после буквы «п» группой предложений или указанным после двоеточия расстоянием до следующего слова), осуществляет сортировку результатов по разным критериям, поиск с опечатками, поиск дат и людей и т. п.

Например, запрос, который может обработать «Интегрум» для решения задачи выяснения стереотипов восприятия японцев, выглядит следующим образом:

(настоящ* или типичн* или среднестатистич* или средн* или обычн* или обыкновенн*):0 (япон* или самурай или «житель японских островов» или «житель Японии» или «житель страны восходящего солнца») не машина не автомобиль не внедорожник не вездеход не джип не видеомагнитофон не магнитофон не видео не телевизор не камера не видеокамера не фотоаппарат[4].

В 2003 году на этом сравнение «Интегрума» с Интернетом и корпусами русского языка можно было закончить. Революционное стремление «Интегрума» сохранить любое русское слово позволяло анализировать широкий и разнообразный материал и было несомненным преимуществом по сравнению с Интернетом и потенциалом корпусной лингвистики. Возможности точного поиска делали эффективнее сбор материала. Однако для профессионального анализа собранных данных не хватало инструментов, которые позволили бы:

– изучать динамику функционирования единицы языка на выбранных временном отрезке и материале;

– изучать частотность единицы языка при помощи вычисления и сравнения:

- абсолютного количества документов, в которых встречается единица языка, в заданной временной точке;

[4] Чем длиннее, подробнее будет запрос, чем больше потенциально возможных контекстов придумает исследователь, тем точнее будут соответствовать результаты на выходе ожиданиям. Важно, что «Интегрум» не имеет ограничений на длину запроса, обрабатывает запрос любой сложности.

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

- процентного отношения документов, в которых встречается единица языка, к общему количеству документов, которые содержатся в «Интегруме» в каждой заданной временной точке;
- процентного отношения документов, в которых встречается единица языка в определенном контексте, к количеству документов, в которых она встречается;

– фиксировать первое употребление единицы языка;

– создать электронный частотный словарь СМИ с функцией обновления.

Сейчас «Интегрум» располагает почти всеми перечисленными инструментами, являя собой качественно новое, постоянно расширяющееся информационное пространство[5] с новыми возможностями поиска и анализа. Возвращаясь к концептуальным вопросам, заданным в начале статьи, использование «Интегрума» можно назвать способом приближения к истине или методологией, главные принципы которой:

- репрезентативность – максимальное количество данных;
- точность – методы точных наук (цифры, вычисления, формулы, таблицы, графики и диаграммы);
- качество данных – поиск в базах «Интегрума» вместо поиска в Интернете, дающего впечатляющие, но абсолютно ненадежные цифры;
- доказательность и доказуемость каждого утверждения.

Такая методология при помощи «Интегрума» была применена в ряде исследований, краткий ретроспективный обзор которых продолжит статью.

1. Поскольку средства массовой иноформации являются главной частью баз «Интегрума», использование сервиса оказалось целесообразным для наших исследований языка СМИ и, прежде всего, его современного состояния, тенденций развития, нормативного аспекта, отражения в нем процессов, происходящих в русском языке и влияния СМИ на эти процессы. Лингвисты отмечали высокую пропорцию жаргонизмов, арго, просторечия, заимствований в языке современных СМИ, но такого рода заключения основывались часто на репрезентативных примерах, хотя, по существу, любое утверждение о частотности того или иного явления, а тем более, о возрастании частотности является абстракцией в отсутствие точного метода квантитативного исследо-

[5] Базы «Интегрума» ежедневно увеличиваются на два источника (под источником в группе «СМИ» понимается одно конкретное издание за все даты с момента существования его электронной версии).

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

вания этого явления на материале всего языка. Требовалось отделить общую тенденцию от частных проявлений, делать заключения о состоянии языка СМИ на всем материале СМИ, проводя анализ частотности употребления исследуемых единиц, сравнивая эту частотность в языке разных СМИ и исследуя динамику изменения этой частотности. Такой материал и средства быстрого поиска можно было найти только в электронном сервисе с большим количеством баз русскоязычных СМИ и с возможностями быстрого и эффективного поиска. В 2002–2003 годах таким сервисом был «Интегрум». Цифры частотности исследованных заимствований, жаргонизмов и арготизмов на протяжении десяти лет позволили сделать выводы об активном росте частотности их употребления[6]. С развитием технологии «Интегрума» стало возможно всестронне и тщательно доказывать гипотезы и разнообразнее иллюстрировать доказательства.

2. Изучение языка СМИ в нормативном аспекте сопровождалось исследованиями понятий «языковая норма», «культура речи», для чего было необходимо изучить сферу рецепции и трансляции: как воспринимаются попытки государственного вмешательства в языковую политику, как отражается тема «языковая политика», «языковая реформа», «политика в области языка» в СМИ, как часто русский язык упоминается в контексте «порча, разрушение, гибель, смерть и т. д.». Инструмент «Интегрума» «Относительная статистика» помог выяснить отношение количества документов, в которых сочетание «русский язык» упоминалось в контексте «порча русского языка», «разрушение русского языка», «гибель русского языка» в центральных газетах, журналах, в региональных СМИ, а также в эфире[7] к количеству документов, в которых это словосочетание употреблялось в любом контексте[8].

3. При исследовании проявлений агрессии в языке СМИ при помощи «Интегрума» выяснялась частотность группы слов «агрессивной семантики», чья семантика и яркая суггестивность в большинстве кон-

---

[6] Подробнее об этом: Языковая норма и культура речи. Эволюция понятий в современном российском дискурсе // Slavic Culture Studies. 2003; Труды и материалы II Международного конгресса исследователей русского языка «Русский язык: исторические судьбы и современность» (МГУ, 2004).

[7] **Инструмент**: относительная статистика/раздельно для групп источников/ежегодно/график. **Запрос**: (порч* или портить или погибел* или гибел* или разруш* или умира* или уничтож* или смерт* или снижен* или ухудш* или расшат*) и «русский язык» не украина не украинск* не латв*/п2. **Временной интервал**: 1998–2004. **Область поиска**: ЦГ, ЦЖ, РСМИ, Т-Р.

[8] Графики и анализ темы содержатся в: Языковая политика в России и роль российского лингвистического сообщества // Journal of the Institute of Language Research. Tokyo, 2005.

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

текстов могла оказывать сильный психологический эффект, провоцируя состояние тревоги, панику, страх, депрессию или агрессию у реципиента. Была исследована частотность группы лексики, которую можно объединить по этому признаку, а затем для прояснения динамики и причин актуализации этой группы лексики была выделена подгруппа из десяти лексем, для каждой из которых был построен отдельный график. На основании графиков стало возможным сделать вывод о динамике частотности присутствия этих лексем в центральных печатных СМИ на протяжении временного отрезка 1993–2005. Графики, построенные для отдельных лексем, позволили увидеть колебания их частотности и сделать предположения о факторах, которые влияют на актуализацию той или иной лексемы[9].

Еще одним проявлением агрессии в СМИ можно считать частотность употребления в них военно-политической лексики. Однако чтобы утверждать, что такое проявление агрессии существует, надо было доказать высокую частотность военно-политической лексики. Частотный словарь «Интегрума»[10] позволил убедиться, что **война**[11] входит в пятьсот самых употребительных в СМИ слов русского языка и с 1995 года появляется, как показывал график, более чем в каждом десятом тексте. Для достижения наибольшей точности результатов из дальнейшего анализа слово **война** было исключено. Однако и без слова **война** группа военно-политической лексики обнаруживала исключительно высокую частотность[12]. Тонкий запрос позволил избежать нерелевантных для анализа контекстов и построить точное статистическое выражение реальной частотности военно-политической лексики. График на рисунке 1

[9] Подробнее об этом: Агрессия в языке СМИ: опыт статистического анализа // Bulletin of the Japan Association for the Study of Russian Language and Literature. 2005.

[10] Словарь частотности языка СМИ (как только закончится работа над индексацией корпуса художественной литературы, словарь будет назван «Частотный словарь русского языка») сделан на материале 8 миллиардов слов. Для сравнения: словарь Засориной делался на материале 8 советских газет за одну определенную дату и некоторых литературных источников. Словарь исследует около миллиона слов, но неизбежно включает непропорционально большое даже для ситуации 60-х годов, если оценивать частотность русского языка в целом, количество идиологем. Это следствие выбора материала: советских газет и других идиологических источников. Корпус машинного словаря Шарова около 40 миллиона слов.

[11] Индекс частотности слова **война** – 6,64, *хорошо* 6,28, *народ* 6,26, *думать* 6,20, *женщина* 6,0.

[12] Здесь использовался инструмент «Интегрума» «Сравнительная статистика по относительной шкале», который позволяет выяснять процентное отношение количества публикаций, в которых употреблялись лексемы (откладывается по шкале y), к общему количеству публикаций, которыми располагал «Интегрум» в каждой заданной временной точке (обозначается на шкале x). Временной интервал вводится в зависимости от задачи.

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

свидетельствует, что военно-политическая лексика даже без слова **война** стала в два раза частотнее за два года с 1993 до 1995, а с 2002 года по сегодняшний день употребляется в каждом четвертом тексте СМИ, став частотнее по сравнению с 1993 годом в три раза. Такую частотность справедливо считать проявлением речевой агрессии в СМИ[13]:

Рис. 1

Следовало выяснить, интенциональна ли агрессия в российских СМИ или это требование контекста и времени? Для этого попытаться определить факторы, которые могли обусловить возрастание частотности военно-политической лексики, что было сделано и представлено в виде отчета. Наиболее интересную динамику обнаружили слова с корневой морфемой **агресс**, которые занимали первую строчку в отчете (см. рис.2).

[13] **Инструмент**: сравнительная статистика/по относительной шкале/график/ежегодно. **Запрос:** агресс* или атак* или батал* или битва или бой!п или бомба* или «военная угроза» или «вооруженное нападение» или зачистка или захват* или конфронтац* или кровопролит* или ликвидация!п или нашествие!п или перестрелка!п или разгром* или сражение!п или террор* или уничтож* или штурм*.**Область поиска:** Центральные газеты, центральные журналы (ЦСМИ). **Временной интервал**: 1993–2005.

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

Рис.2

	31.12 1993	31.12. 1994	31.12 1995	31.12 1996	31.12 1997	31.12 1998	31.12 1999
Агрессия	1,03	1,23	1,95	1,96	1,99	2,17	3,10

	31.12 2000	31.12 2001	31.12 2002	31.12 2003	31.12 2004	31.12 2005
Агрессия	2,55	2,66	2,91	3,09	2,84	2,71

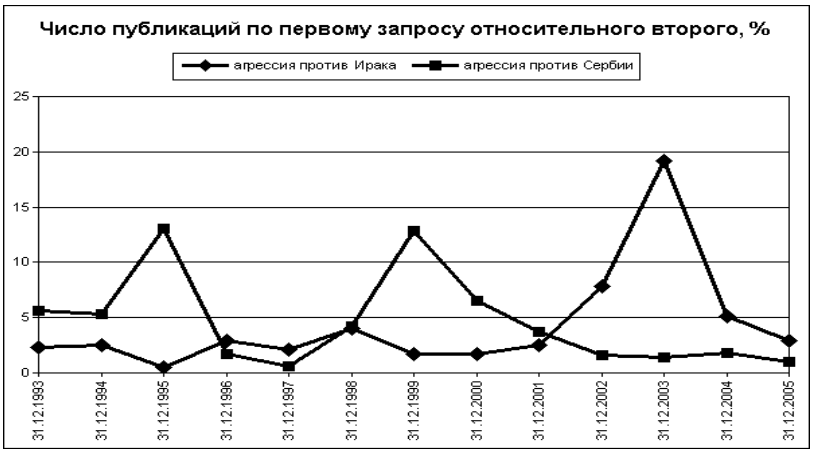
Цифры свидетельствовали о пиках частотности лексем **агрессия**, **агрессивность**, **агрессивный**, **агрессор**, **агрессивно** в 1999 и в 2003 годах, что могло быть связано с агрессией против Сербии и Ирака. Еще один инструмент «Интегрума» - «Относительная статистика», позволил уточнить гипотезу[14]. График «Относительной статистики»[15] на рисунке 3 демонстрирует, что слова с корневой морфемой **агресс** становились частотными трижды – в 1995 и в 1999 годах в связи с актуализацией контекста «**агрессия** против Сербии», в 2003 году «**агрессия** против Ирака».

[14] При использовании этого инструмента система выясняет процентное отношение количества документов, в которых лексема употреблялась в заданном контексте (в данном случае: «агрессия против Сербии» или «агрессия против Ирака»), к количеству всех документов, в которых она употреблялась в любом контексте.

[15] **Инструмент**: относительная статистика/график/ежегодно. **Запросы**: первая пара запросов: агресс* (серб* или белград*)/п2 относительно агрессия!п; вторая пара запросов: агресс* (США или америк*) (ирак* или багдад*)/п2 относительно агрессия!п. **Область поиска**: ЦСМИ. **Временной интервал**: 1993–2005.

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

Рис.3



Сравнение показателей графика на рисунке 3 и отчета на рисунке 2 обнаружило существенное различие между поведением кривых относительной частотности и цифр отчета. Кривые показывали, что реальная война немедленно актуализирует слова с корнем **агресс**. В отчете зависимость от военных действий должна была бы приводить к такому же резкому, как на графиках, взлету и падению цифровых значений в начале и в конце войны. Даже представив всемирную новейшую историю как сплошную череду войн и агрессий, следовало ожидать некоторой скачкообразности и отсутствия роста. Таких ожиданий цифры отчета не оправдывали, соответственно, зависимость актуализации лексем от военных действий не была абсолютной.

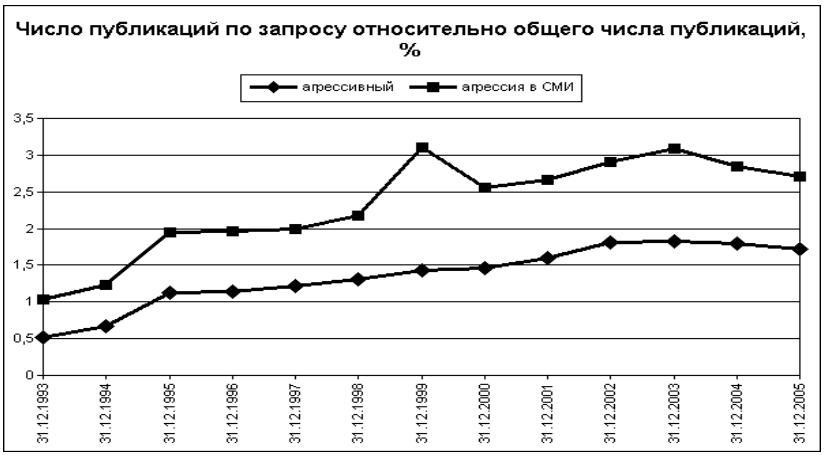
Для большей доказательности были выбраны еще две агрессии, которые стали заметным событием и достаточно активно обсуждались в СМИ: агрессия Ирака в Кувейте, о которой заговорили в связи с агрессией США в Ираке, и «агрессия чеченских боевиков в Дагестане». Графики подтверждали тенденцию, обозначенную графиком на рисунке 3: начало войны немедленно «вбрасывало» слова с корневой морфемой **агресс** в СМИ, прекращение или снижение активности боевых действий (или потеря интереса к ним со стороны СМИ) почти полностью снимали необходимость в их употреблении в контексте этих конкретных агрессий и войн.

Таким образом, можно было заключить, что если реальные военные действия и отражаются на колебаниях частотности лексем, то это отнюдь не единственный фактор, который определяет общую возрастаю-

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

щую динамику их частотности. Интересные данные обнаружил сравнительный анализ кривых частотности для отдельных лексем в парадигме слова **агрессия.** Оказалось, что для прилагательного **агрессивный** зависимости от войн и агрессий практически не существует. Его частотность возрастала непрерывно и плавно, в отличие от всей группы слов с корнем **агресс**, и увеличилась по сравнению с 1993 годом в три раза, что показывает график на рисунке 3[16]. На этом графике объединены кривая частотности для группы лексем с корневой морфемой **агресс** (верхняя кривая) и кривая частотности прилагательного **агрессивный** (нижняя кривая). Если верхняя кривая отражает зависимость лексем с корнем **агресс** от агрессии против Сербии и Ирака в виде скачка частотности, то нижняя кривая не обнаруживает такой зависимости.

Рис.4



Следовательно, слово **агрессивный** активизировали не войны и агрессии, на динамику частотности влияли какие-то иные факторы, помимо военных действий.

Следующим шагом было выяснение возможных лингвистических причин. Для этого требовался контекстный анализ максимально воз-

[16] **Инструмент**: сравнительная статистика/по относительной шкале/график/ежегодно. **Запрос:** Система предоставляет возможность объединять исследования статистики для двух запросов. Здесь верхний график отражает динамику присутствия всех слов с корневой морфемой **агресс** (запрос: агресс*), нижний график отражает частотность прилагательного **агрессивный** (запрос: агрессивный). **Область поиска:** ЦСМИ. **Временной интервал**: 1993–2005.

***Язык, сознание, коммуникация**: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. — М.: МАКС Пресс, 2006. — Вып. 32. — 108 с. ISBN 5-317-01586-3*

можного количества примеров. В базах «Интегрума» при помощи информационно-поисковой системы с функцией точного поиска было найдено и проанализировано 43.800 примеров употребления слова **агрессивный** за пятнадцать лет. «Интегрум» удобен для контекстного анализа, так как может открыть весь документ, следовательно, любой длины контекст, в котором содержится слово. Анализ позволил представить историю семантических преобразований лексемы[17] и отнести это слово к группе «вторичных заимствований», наряду со словами **шок** / **шокировать** и **амбиции**. При анализе вторичных заимствований выяснялась этимология их новых значений, история вхождения в русский язык. Было выяснено, когда «**шок**» это стало «**по-нашему**» и насколько по-нашему, частотность и история появления в русском языке кальки «**приятно шокировать**»[18].

Чтобы продемонстрировать все возможности использования «Интегрума» в научных исследованиях не достаточно формата статьи, для этого выбран формат книги «Интегрум: точные методы и гуманитарные науки», в которой собраны работы, сделанные с использованием «Интегрума»[19]. Здесь же ограничимся кратким обзором некоторых исследований, проведенных с помощью «Интегрума», и знакомством с инструментами и методологией их использования. В заключение важно отметить, что «Интегрум» – постоянно развивающаяся компьютерная технология и постоянно растущее информационное пространство. Включение корпуса разговорной речи, создание электронной коллекции русской литературы позволит делать исследования в области литературоведения и эволюции русского литературного языка, сравнения языка разных писателей, СМИ и литературного языка, создание частотного словаря русского языка и так далее.

---

[17] Описанную в работах: Агрессивная комедия. История слова агрессивный в трех частях // Slavic Culture Studies. 2005; Вторичные заимствования в русском языке XXI века // Интегрум: точные методы и гуманитарные науки. Сб. статей / Под ред. Г. Никипорец-Такигава. (В печати.)

[18] Там же.

[19] Интегрум: точные методы и гуманитарные науки. Сб. статей / Под ред. Г. Никипорец-Такигава. (В печати.)