

Башкирский сегмент Интернета: семиотика и лингвистика  
(Тезисы выступления)

Ввиду малого объёма и числа пользователей башкирским интернетом до сих пор не интересовались.

Под Башнетом мы будем понимать совокупность доступных в сети Интернет текстовых документов, написанных на башкирском языке.

По нашим данным на конец января 2012 года объём башкирского сегмента составляет 66.199 страниц<sup>1</sup>. В проведённом нами замере участвовало 30 доменных имён. Это не выборка, а более-менее полный перечень сайтов, которые вообще содержат тексты на башкирском языке. По самой грубой оценке, современный объём Рунета не может быть менее 5 миллиардов документов, то есть разница с объёмом Башнета составляет 5 порядков. Показательно, что в Рунете по данным компании «Яндекс» на 2009 год опубликовано около 2,3 триллиона слов<sup>2</sup>. По нашим данным в Башнете содержится 27.252.251 слово. Таким образом, разница в объёмах Рунета и Башнета в словоупотреблениях будет составлять те же 5 порядков, что и при сравнении страниц.

Логически сайты Башнета следует разделить на две группы. В первую войдут страницы республиканских СМИ, которые печатаются на бумаге или выходят в радиоэфир и одновременно с этим вывешивают свою продукцию в Интернете. Мы будем называть такие сайты «медийными». Во вторую группу будут входить все остальные сайты, то есть те, наполнение которых изначально рассчитано на Интернет. Медийная группа сайтов – это примерно треть всего перечня (9 из 30 сайтов). Однако её доля в страницах почти достигает половины (41.298 страниц из 87.462), а в словоупотреблениях даже превышает половину (15.155.408 из 27.252.251 слов). Именно в этой группе находится самый объёмный сайт Башнета – страница газеты «Йэшлек». На нём опубликовано 8.470.444 слова, что составляет 55.89% всего объёма словоупотреблений на медийных сайтах и 31% от всего корпуса Башнета. Тексты газеты «Йэшлек» представлены на 14.229 страницах, что меньше, чем количество страниц в башкирской Википедии (14.441), однако последняя более чем в два раза уступает сайту газеты в количестве слов (3.981.524).

Вторая группа сайтов демонстрирует некоторое тематическое и жанровое разнообразие, какое только возможно в ограниченном двумя десятками пунктов кругу. Здесь имеются литературные странички, блоги и форумы, сайт спортивных болельщиков и общественного движения. Наиболее значительный ресурс в этой категории – свободная мультязычная интернет-энциклопедия «Википедия». В башкирской Википедии на данный момент около 15.000 статей и 5.000 участников, из которых примерно 50 (1%) активно участвуют в создании и улучшении статей.

Тематический и жанровый обзор свидетельствует, что для башкирского сегмента Сети фактически отсутствуют все основные сервисы, обычно занимающие большую часть времени пользователя. Проще говоря, современному человеку ходить в Башнет незачем, башкирские сайты не предоставляют для него ничего полезного. Как минимум на треть он является всего лишь приложением к СМИ. В гораздо меньшей степени, чем это развито в Рунете и всемирной паутине в целом, Башнет является ресурсом общения. Башкирскую Википедию следует, скорее, счесть имиджевым проектом, призванным не столько поставлять пользователю информацию, сколько заявлять о существовании башкирского языка и башкирской культуры в глобальном культурном пространстве. Ту же представительскую функцию выполняют и башкироязычные разделы на разного рода официальных сайтах (мэрии Уфы, Башкирского университета, литературного конкурса

<sup>1</sup> Весь объём сайтов, которые публикуют тексты на башкирском языке, составляет 87.462 страницы, из них 21.263 не содержат башкирских текстов.

<sup>2</sup> Контент Рунета [Электронный ресурс] // URL: [http://company.yandex.ru/researches/reports/ya\\_content\\_09.xml](http://company.yandex.ru/researches/reports/ya_content_09.xml).

переводов тюркоязычной поэзии «Ак Торна» и пр.). Это означает, что, по большому счёту, размещённые в Интернете башкирские тексты не предполагают существования читателя.

Самым очевидным параметром детального описания Башнета как целого, разумеется, становится частотность лексики.

По этому параметру язык Интернета заведомо отличается от языка печатных изданий (это ещё одна причина разделять сайты на медийные и остальные). Самым частотным словом из лексики сетевой коммуникации в башкирском интернете выступает «форум», который находится в середине третьего десятка. Следующий за ним специфический интернет-термин – «Википедия» – обнаруживается в середине четвёртого десятка.

Очевидно, что такое положение вещей обусловлено неразвитостью интернет-коммуникации на башкирском языке и, соответственно, отсутствием сетевого метаязыкового уровня. Башнет – это менее всего способ коммуникации на башкирском языке. Скорее, это специальная площадка для представления разного рода официальных документов.

Соединительный союз *һәм* 'и' и послелог *менән* 'с' вполне ожидаемо занимают верхние строчки в частотнике Башнета. Столь же высокие позиции у этих слов и в других частотных словарях башкирского языка. Но вот третье слово в частотном словаре Башнета в сравнении с другими частотниками довольно неожиданно, и, как мы считаем, отражает крен в сторону официального языка. Это послелог *буйынса*. Употребительность этого послелога может быть связана с такими случаями, как этот: *баш мөхәррирзең дөйөм мәсьәләләр буйынса урынбасары* 'заместитель редактора по общим вопросам'. Четвёртое место в частотнике занимает слово *башкорт* 'башкирский' (66.945 вхождений). Неожиданно высокие позиции в частотном списке занимают формы *ағасым* 'моё дерево' (6 место и 59.792 вхождений), *кошом* 'моя птица' (8 место, 59.739 вхождений), *ырыуым* 'мой род' (10 место, 47.789 вхождений), *ораным* 'мой клич' (16 место, 43.791 вхождений). Такое необычное положение вещей объясняется форматом башкирского форума (<http://www.bashforum.net/>), регистрация на котором предполагает указание пользователем признаков единства своего племени<sup>3</sup>. Глагол *үзгәртәргә* 'изменить' (9 место и 49.017 вхождений в частотном списке Башнета) тоже не характерное часто употребляемое слово в башкирском языке за пределами Интернета. Его высокие позиции в перечне лексики объясняются дизайнерским решением «Википедии», в которой слово «изменить» сопровождает каждый раздел статьи (которых на одной странице может быть много). А так как «Википедия» составляет 31,28 % всего объёма немедийного Башнета, частотности слова *үзгәртәргә* удивляться не следует. То же относится и к слову *мәкалә* 'статья' (12 место 45.984 вхождений), отсутствующему в ряду самой частотной лексики в других частотных словарях. При этом можно было бы ожидать высокие позиции этого слова в Н, однако, в реальности мы сталкиваемся с иным положением вещей.

Специального очерка заслуживает слово *йүнәлештәр* 'направления' (14 место, 44.049 вхождений). Его мы тоже вряд ли смогли бы ожидать на верхних позициях частотного списка, составленного на основе сбалансированного корпуса текстов данного языка. Однако чрезмерно канцеляризированный язык Башнета демонстрирует аномальную популярность этой словоформы. Она появляется в характерных контекстах, в которых речь идёт о развитии новых направлений, стремлении к новым направлениям исследований, о работе по следующим направлениям, об основных направлениях работы<sup>4</sup>. Самая частая коллокация с участием этой лексемы в корпусе Башнета: *төп йүнәлеш* 'основное направление'. Обилие контекстов свидетельствует, что *йүнәлеш* относится к ряду особенно «модных» слов с предельно абстрактным значением, об обилии которых в современной русской языковой

<sup>3</sup> См. статью «Структура родоплеменной организации башкир» в Краткой энциклопедии «Башкортостан» (1996).

<sup>4</sup> В русском словаре слово *направление* имеет такие показатели: общая частота леммы: 179.2, показатель R: 100, коэффициент вариации D: 92, количество документов, в которых встречается это слово: 6394. Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка на материалах Национального корпуса русского языка. – М.: Издательский центр «Азбуковник», 2009. – С. 180.

ситуации писал М. А. Кронгауз<sup>5</sup> (ср. также *нанотехнологии, инновации, модернизация* и т. д.).

Интересным показателем жанрового состава попавших в базу документов являются инвективные и обценные слова. Их наличие свидетельствовало бы о наличии ситуации живого общения. Здесь мы наблюдаем довольно ущербную парадигму, в которой актуализированы лишь слова со значением 'meretrix' (*уйнаш, уйнашсы*: 21 и 105 употреблений соответственно), а остальные лексемы, связанные с именованием телесного низа и под., в Башнете не встречаются.

Среди слов, имеющих высокую частотность и в Рунете, и в Башнете, можно упомянуть *год / йыл* (3 место в списке существительных Рунета / 50 место для начальной формы в Башнете и сравнительно высокие позиции у омонимичных форм типа *йылга* в общем списке, см. выше<sup>6</sup>), *новость / яңылыктар* (4 место в Рунете / 176 место в Башнете, 12.530 вхождений), *форум / форум* (7 место в Рунете / 25 место в Башнете), *поиск / эзләү* (8 место в Рунете / 26 место в Башнете, 33.217 вхождений), *день / көн* (9 место в Рунете / 339 место, 6.137 вхождений), *пользователь / кулланушы* (14 место в Рунете / 562 место в Башнете, 3.333 вхождений), *время / вакыт* (19 место в Рунете / 199 место в Башнете, 10.790 вхождений), *человек / кеше* (20 место в Рунете / 93 место в Башнете, 16.745 вхождений). В этом списке видны ключевые слова, отражающие, как можно представить, универсальные для сайтостроительства понятия, независимо от языка оккупирующие верхнюю часть списка. Отметим особую, как мы видим, важность для Интернета слов, относящихся к семантическому полю «время» (*йыл, көн, вакыт*) и ассоциативно с ним связанных (*яңылыктар*). Можем заключить, что с аксиологической точки зрения Сеть ориентирована на категорию времени более, чем на какую-то другую из философского репертуара.

---

<sup>5</sup> Кронгауз М. А. Русский язык на грани нервного срыва. – М.: Языки славянских культур, 2007.

<sup>6</sup> Цифры для сравнения весьма приблизительные, так как рейтинги для Рунета и Башнета составлены по разным принципам: в Рунете считаются леммы, а в Башнете словоформы; в Рунете каждая часть речи приводится в составе собственного рейтинга, а в Башнете учитывается место слово в общем (смешанном с точки зрения частеречной принадлежности) рейтинге. Тем не менее, есть слова, для которых разница в употребимости достаточно показательна, несмотря на принципы подсчёта.