

Опыт лингвистического исследования коммуникации в Twitter: причины и особенности использования англицизмов в русскоязычных микроблогах

Е.В. Карнуп

Российский государственный педагогический университет им. А.И. Герцена
ekarnup@gmail.com

Аннотация

В статье описывается опыт исследования коммуникации в Twitter с использованием методов корпусной лингвистики. Предпринята попытка количественного и качественного анализа заимствованных лексических единиц в языке общения в русскоязычном сегменте сервиса микроблогов Twitter.

Факт зарождения сети Интернет в англоязычной стране и неиссякаемый поток технических нововведений в этой области, который и по сей день имеет свои истоки в США, поначалу приводил исследователей языков национальных сегментов сети Интернет к мысли об угрозе целостности и чистоте мировых языков. Существовали опасения, что под мощным влиянием английского языка, подкрепленным стремительным распространением сети Интернет по всему миру и статусом английского как языка международного общения, национальные языки превратятся в своеобразные лингвистические гибриды, навсегда утратив свою полноту и самобытность [5].

На раннем этапе развития сети Интернет такие опасения были небеспочвенны, поскольку английской была не только терминология, обозначающая все без исключения реалии сети (например, Internet, web-site, browser), но и вся информация, представленная в Интернете, была доступна лишь англоговорящим людям. Кодировка ASCII, применявшаяся на заре интернетизации, разрабатывалась как американский стандарт и была рассчитана только на латинские буквы, используемые в английском языке. Вследствие такого ограничения обмен информацией на каком-либо другом языке, кроме английского, был невозможен. Только по мере распространения сети Интернет по всему миру стали разрабатываться кодировки для представления символов других национальных языков.

Несмотря на то, что и по сей день проблема неодинаковой степени приспособленности сети Интернет для разных национальных культур решена не полностью, статус русского языка в сети Интернет

вызывает меньше беспокойства, чем раньше, благодаря активной поддержке языкового разнообразия в глобальных информационных сетях на государственном и международном уровнях [3]. О вытеснении русского языка английским речи уже не идет, но, тем не менее, Интернет, как свободное пространство общения без национальных границ, накладывает свой отпечаток на характер коммуникации. Так, тенденция глобализации, опирающейся, среди прочего и на развитие сети Интернет, сопровождается интенсивным контактированием языков, что приводит к их взаимному влиянию и активному проникновению заимствований в языковую систему русского языка.

С учетом описанных выше тенденций, говоря о заимствованиях, исследователи зачастую подразумевают англицизмы, так интенсивно используемые в современном русском языке. Однако пласт англицизмов состоит не только из упомянутой выше компьютерной терминологии.

С целью выяснить, как русский язык приспосабливается к условиям коммуникации в сети Интернет, в частности, как он справляется с влиянием англоязычной традиции компьютерно-опосредованного общения, проведено исследование особенностей языка межличностного общения в Интернете. В рамках исследования предпринята попытка количественного и качественного анализа заимствованных лексических единиц в языке общения в русскоязычном сегменте сервиса микроблогов Twitter, позволяющего представить степень распространенности, освоенности заимствованных элементов и особенности их функционирования на современном этапе развития русскоязычной компьютерно-опосредованной коммуникации.

Сервис микроблогов Twitter был выбран для исследования неслучайно. Созданный сравнительно недавно — в 2006 году, Twitter занимает сейчас одну из первых строчек рейтинга популярности среди сервисов общения в Интернете (<http://www.ranking.com/>) и является самой популярной в мире площадкой для ведения микроблогов. И хотя интерфейс Twitter стал поддерживать русский язык лишь 26 апреля 2011 года, на сегодняшний день наблюдается быстрый рост аудитории русскоязычного сегмента сервиса. Соответственно,

можно предположить, что на сегодняшний день продолжается активное освоение нового формата общения русским языком и тщательный анализ позволит обнаружить механизмы влияния англоязычной традиции общения на язык русскоязычных микроблогов.

Twitter выполняет множество функций, он одновременно используется как чат, форум, блог, социальная сеть, рекламная площадка, средство массовой информации и т. д. Компании, государственные органы, средства массовой информации, публичные персоны используют Twitter в качестве оперативного средства информирования. Короткая запись позволяет передать суть события, после чего пользователь, заинтересовавшийся данной темой, может найти подробности в сети [2].

События начала 2011 г., когда в Северной Африке и на Ближнем Востоке произошел ряд восстаний, называемых Twitter-революциями, или революциями социальных медиа, подтвердили важную роль сетевых сервисов общения, в частности, Twitter, в современных процессах коммуникации в обществе [1].

Twitter является не только актуальным и обширным, но также исключительно удобным полем для различных исследований, в том числе и лингвистических. Любый человек и в любое время может присоединиться к общению в этих формах коммуникации; высказывания участников виртуальной беседы последовательно появляются на мониторе в виде текста, который может быть сохранен в памяти компьютера, а впоследствии при желании распечатан или отредактирован. По этой причине не наблюдается потерь информации из-за трудностей при чтении, шумов или, например, фонетических дефектов, которые обычно осложняют лингвистические исследования в других сферах.

Кроме того, для исследователя ценно то, что все записи в Твиттере оперативно индексируются поисковыми системами, что значительно облегчает поиск и обработку исследуемого материала.

Кроме всех перечисленных сервисов не нужно забывать и о возможностях самого сайта twitter.com, предлагающего услуги расширенного поиска по ключевым словам с указанием языка, по людям, местам и даже по типу сообщения (вопросительное, утвердительное, отрицательное).

Таким образом, современный исследователь вооружен множеством инструментов для анализа текстов блогосферы и, в частности, Твиттера, являющихся ценным лингвистическим материалом, отражающим современные тенденции развития языка.

В рамках проведенного исследования сформирован корпус текстов сообщений системы микроблогов Twitter, в который вошли 376791 сообщений, выделенных методом сплошной выборки с использованием ресурса Datasift, представляющего собой платформу сбора данных из социальных сетей. Всего корпус содержит 1876286 словоупотреблений.

Для обеспечения репрезентативности корпуса фильтр для сбора сообщений был настроен таким образом, чтобы в коллекцию были отобраны только сообщения, автор которых указал в своем профиле русский язык как предпочитаемый язык общения. Кроме того, в коллекцию сообщений не вошли массовые рассылки и сообщения рекламного характера, а также ретвиты (повторяющиеся сообщения), таким образом, в корпусе представлены тексты, порожденные русскоязычными пользователями в процессе межличностной коммуникации в системе микроблогов Twitter.

Все сообщения в корпусе разбиты на токены, каждому из которых приписан условный тип. Для исследования влияния английского языка на русский интересно рассмотреть частотные характеристики трех типов токенов: тип RW (Russian Word), обозначающий последовательность кириллических символов (т.е. слово, написанное кириллицей, например, просто), тип LW (Latin Word), обозначающий последовательность латинских символов (т.е. слово, написанное латиницей, например, people) и тип RLW (Russian & Latin Word), в котором последовательность символов состоит одновременно из кириллицы и латиницы (например, fashion-фотограф) (Таблица 1).

Таблица 1. Наиболее частотные токены типа RW, LW и RLW

Тип токена	Словоформа	Частота	Место в частотном списке
RW	не	40467	1
	я	35507	2
	и	34898	3
LW	I	1810	1
	you	1774	2
	the	1507	3
RLW	xD	55	1
	XD	15	2
	xDD	11	3

Данные табл. 1 свидетельствуют о том, что русскоязычные пользователи довольно редко прибегают к латинскому алфавиту, предпочитая буквы родного языка — частота кириллических словоформ (RW) во много раз превышает частоту словоформ, написанных латинскими символами (LW). При этом частота словоформ, объединяющих в себе кириллические и латинские буквы (RLW), очень низкая, что можно объяснить, во-первых, действием закона экономии усилий (пользователи не хотят тратить силы и время на переключение раскладки), а во-вторых, спадом популярности такого способа словообразования, при котором заимствованный элемент — корневая морфема или аббревиатура — сохраняется в латинской графической форме (HD-каналы, gif-анимация и т.д.)

Одним из мотивов комбинирования в одном слове символов двух раскладок является стремление пользователя сохранить графический облик упоминаемого бренда: ДжаZ (Animal ДжаZ невероятно выросли...), МакСим (хочу на концерт МакСим!!!!!!). В приведенных примерах пользователи соблюдают точность написания названий групп, тем самым не только показывая свою компетентность в этой области, но и проявляя уважение к упоминаемым исполнителям.

Преобладание кириллицы в исследуемом материале является доказательством доминирующего положения русского языка в межличностном общении русскоязычных пользователей. Этот вывод подкреплен и статистическими данными по языку сообщений: корпус содержит 282268 русскоязычных сообщений, 44613 англоязычных сообщений и 49910 смешанных сообщений.

Среди англоязычных сообщений преобладают два типа:

1) цитаты, представляющие собой строки песен (Listen, listen — I will take a whisper, if That's all what you have to give @ #NP Jason Walker — Echo) и изречения известных людей ("If I don't answer on Twitter, it doesn't mean that I don't read and don't smile to your words." — Justin.);

2) обращения к англоязычным собеседникам (@tyrabanks You are great, I admire you as a model, when will the new season of "Top Model in American"?).

Если во втором типе английский язык выполняет функцию средства международной коммуникации, ставшей возможной благодаря сети Интернет, то в первом типе причины использования английского языка кроются в вовлеченности пользователя в массовую музыкальную культуру, крупным «поставщиком» которой является США. Таким образом, в первом случае выбор английского языка в качестве средства общения обусловлен факторами, не связанными с каналом общения.

Сравнительно большое число смешанных сообщений объясняется тем, что хештеги (начинающиеся с символа #, например, #followback, #EXO1stAnniversary) и упоминания (начинающиеся с символа @, например, @justinbieber, @xenia_sobchak), являющиеся неотъемлемой частью традиции общения в Твиттере, написаны, в основном, на латинице, поэтому если даже полностью русскоязычное сообщение содержит такой хештег или упоминание, оно автоматически относится к смешанным сообщениям.

Преобладание кириллицы не означает, однако, что русский язык межличностного общения в Твиттере игнорирует англоязычные приметы новой площадки коммуникации. Проанализировав частотный список кириллических словоформ, можно увидеть большое число заимствований, которые русский язык вобрал в себя и переработал, подстроившись под нужды нового формата общения. Частота транслитерированных заимствований гораздо выше англоязычных аналогов. Например, в парах лол

(частота 580) — lol (частота 66), ок (частота 430) — ok (частота 46), хд (частота 672) — XD (частота 152), фак (частота 126) — Fuck (частота 60) частота кириллической словоформы во много раз превышает частоту не транслитерированного аналога. Выделенные заимствования, переданные на основе кириллической графики, можно распределить по следующим группам (в скобках рядом с примерами указана частота словоформ в исследовательском корпусе):

— заимствования, обозначающие новые понятия, связанные с сетью Интернет: твиттер (481), твит (444), тви (298), юп (208), инстаграм (68), скайп (64), фолловер (58), юн (53), фф (46), ава (44), гугл (42);

— заимствования, обозначающие технические новинки вне сети: трек (72), айфон (69), комп (60), компьютер (49), симс (31), айпад (29), смартфон (13);

— заимствования, передающие общеупотребительные слова и фразы английского языка из молодежного сленга, рекламных слоганов, игр, фильмов и пр.: ок (430), ту (180), окей (153), бро (129), го (79), ван (77), тайм (68), лав (67), пати (52), плиз (28), тру (26), вин (25), айс (24), слоу (21), лов (20);

— заимствования из обценной лексики, служащей для эвфемизации: фак (126), бич (18);

— заимствованные междометия и акронимы, несущие функции междометия: лол (580), лал (219), лалка (34), омг (37), ноу (45), воу (92), вау (72), бла (58), йоу (51), омномом (32);

— заимствованные эмодзи: хд (672), хдд (170), хДД (44), тт (157), 33 (29), хх (38).

Таким образом, не только компьютерная и техническая лексика составляет основу пласта заимствований. Сильное влияние на русскоязычного пользователя сети Интернет оказывает американский вариант массовой культуры — общеупотребительные фразы из англоязычных фильмов, музыкальных и рекламных роликов окутаны ореолом престижности, позволяют блеснуть перед собеседником знанием современного стиля общения. По этой же причине заимствуется и множество междометий, «подслушанных» на видео — не с целью добавить высказыванию информативности, а с целью придать модную интонацию.

Кроме того, на русскоязычную почву привносятся и элементы англоязычной традиции общения в сервисах коммуникации в сети. Так, акроним LOL, еще несколько лет назад писавшийся исключительно латиницей и заглавными буквами, уже преобладает в кириллической форме, и включен в процесс словообразования путем аффиксации и чередования корневой гласной: лол—лал—лалка. А эмоджон XD, призванный изображать лицо с зажмуренными от смеха глазами и широко растянутым в улыбке ртом, превратился в кириллический хд (наиболее частотный из всех заимствований), уже мало чем внешне напоминающий улыбку, но, тем не менее, успешно ее обозначающий. Русский язык, выясняя возможности компьютерно-опосредованной комму-

никации, идет вслед за «старшим товарищем» английским, перенимая и перерабатывая наиболее удобные средства общения.

Поскольку процесс освоения некоторых заимствований еще не завершен, их написание может варьироваться в зависимости от предпочтений пользователя (фолловер — фоллоуер, ретвит — ритвит и пр.). Многие из распространенных заимствований из-за их недолгого существования еще не зафиксированы в авторитетных словарях, и могут встретиться лишь в любительских онлайн-словарях, не имеющих отношения к кодификации современной нормы [4]. Поэтому варьирование написания некоторых заимствований нельзя оценить с точки зрения правильности или ошибочности. Однако отсутствие устоявшихся орфографических норм не препятствует активным словообразовательным процессам, в которые вовлечены заимствования.

Кроме аффиксации (твиттерские (41)), наиболее активными механизмами словообразования в процессе освоения англоязычных заимствований в языке пользователей Твиттера можно назвать аббревиацию (юп (208), юн (53), фф (46)) и усечение (тви (298), комп (60), ава (44), акк (41), ник (32), фик (27), инста (11)), популярность которых также объясняется действием закона экономии усилий.

Таким образом, в русскоязычном сегменте Twitter преобладает коммуникация на русском языке. Английский язык используется в основном только для международного общения и для цитирования англоязычных прецедентных текстов.

В русскоязычной межличностной коммуникации происходят процессы активного освоения заимствований из английского языка. По мере распространения заимствований они довольно быстро проходят стадию графического освоения, формального освоения слова, включаются в словообразовательные процессы.

Английский язык проникает не только через профессиональные жаргоны компьютерщиков и терминологию, связанную с техническими новинками, заполняя семантические пробелы в лексике русского языка, но и через англоязычную массовую коммуникацию, массовую теле-, кино- и музыкальную продукцию, рекламу и пр. В последнем случае употребление заимствований представляет собой своеобразное следование моде, здесь ключевым фактором в выборе заимствованных лексических средств являются стремление авторов сообщений подражать харизматичным персонажам западной культуры, фактор престижности англицизмов, желание быть включенным в современную массовую культуру.

Кроме того, английский язык выступает образцом в рамках традиций компьютерно-опосредованной коммуникации, у которого русский язык перенимает различные элементы, присущие только дан-

ному виду общения — эмотиконы, акронимы, аббревиатуры, междометия.

Несмотря на обилие заимствований, можно отметить, что русский язык в условиях компьютерно-опосредованной коммуникации достойно справляется с воздействием английского, качественно перерабатывая иноязычный материал и обеспечивая эффективность коммуникации и ее соответствие нуждам современного общества.

Литература

- [1] Валюгин А.С. Проблемы диалога культур в современном коммуникативном пространстве // Вестник Нижегородского государственного лингвистического университета им. Н.А. Добролюбова. Нижний Новгород, 2011. №15. С. 207—214. URL: <http://elibrary.ru/item.asp?id=16961684> (дата обращения: 12.07.2013).
- [2] Глазкова С.А. Особенности коммуникации в поддельных микроблогах сети Интернет // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. Тамбов: Грамота, 2011. № 3 (9). Ч. I. С. 27—30. URL: <http://elibrary.ru/item.asp?id=17010796> (дата обращения: 12.07.2013).
- [3] Рекомендация ЮНЕСКО «О развитии и использовании многоязычия и всеобщем доступе к киберпространству» // Языковое разнообразие в киберпространстве: российский и зарубежный опыт. Сборник аналитических материалов. М.: МЦБС, 2008. С. 130—136;
- [4] Словарь молодежного сленга [Электронный ресурс]. URL: <http://teenslang.su> (дата обращения: 12.07.2013).
- [5] Crystal D. English as a Global Language. 2nd Edition. Cambridge, 2003.

A piece of twitter-based linguistic research of communication: causes and characteristics of the use of anglicisms in russian-language microblogging

E. V. Karnup

The author of the article describes the experience of the study of communication in Twitter using corpus linguistics methods. The article contains quantitative and qualitative analysis of borrowed lexical items in the language of communication on Russian Twitter. The causes of English influence on the Russian language used in microblogging are analyzed.