

СВЯЗЬ ЛСГ РУССКИХ И БЕЛОРУССКИХ ГЛАГОЛОВ ВОСПРИЯТИЯ С ДРУГИМИ ГРУППАМИ. ПРОБЛЕМЫ МАШИННОГО ПЕРЕВОДА

Одним из главных методических принципов преподавания русского языка в иноязычной аудитории является «целесообразность подачи и закрепления лексики в системе, а не изолированными единицами» [3, с. 5]. Процесс освоения колоссальных лексических пластов, так или иначе, предполагает выделение в них групп, рядов, классов слов, в зависимости от грамматических, семантических, стилистических и др. признаков. Данный принцип понятен и разумен, если речь идёт о студенте. Когда же лексический фонд нового языка «пытается» *усвоить* и *освоить* не человек, а компьютерная программа, принцип этот становится и вовсе ведущим. Так, *традиционные* машинные переводчики, использующие технологию, основанную на применении алгоритмов (Rule-based machine translation), обрабатывают входную лексику в несколько этапов. На первом этапе устанавливаются морфологические характеристики переводимых единиц (род, число, лицо...), затем находит разрешение вопрос грамматической многозначности (определяются значения слов, которые могут относиться к разным частям речи). Следующий шаг, представляющий для нас особый интерес — систематизация: программа объединяет отдельные слова в группы (именные, глагольные...). На заключительных этапах производится синтаксический анализ и синтез полученных данных. *Статистические* же машинные переводчики, рассматриваемые в данной работе, базируются не на языковых правилах, которые программе даже не известны, но на технологии статистического вычисления вероятности текстовых совпадений (Statistical-based machine translation). Такие переводчики группируют, систематизируют обрабатываемую лексику, однако иным образом.

Структура статистического машинного переводчика состоит из трёх частей: «модели перевода», «модели языка» и «декодера» [6]. «Модель перевода» представляет собой таблицу, «в которой для всех известных системе слов и фраз на одном языке перечислены все возможные их переводы на другой язык и указана вероятность этих переводов» [6]. Создаётся «модель перевода» путём подбора параллельных документов: оригинальных и переводных текстов (художественных, научных, публицистических), сайтов, функционирующих одновременно на двух языках и т.д. В параллельных текстах система отбирает сначала пары соотносимых предложений, затем — пары соотносимых слов или словосочетаний. В сущности, программа, фиксируя различные варианты перевода одной и той же языковой единицы, *сама того не зная*, составляет ряды синонимов, фрагменты лексико-семантических групп, подбирает эквивалентные данному слову перифрастические выражения. «Изучив» сотни миллионов различных параллельных документов, статистический переводчик приступает к созда-

нию «модели языка». Программа составляет частотный список *всех* слов и словосочетаний, употреблённых в текстах. Принцип работы «декодера», отвечающего за конечный перевод, прост: для каждого предложения исходного документа он подбирает все варианты перевода, сочетая между собой синтагмы из «модели перевода», и сортирует их по частотности употребления в порядке убывания. Затем все получившиеся варианты декодер соотносит с «моделью языка». Так, например, пословный перевод английской фразы *to be or not to be?* на русский как *бывает или не бывает?*, выполненный «моделью перевода», декодер исправит на *быть или не быть?*, выбрав наиболее частотное употребление уже комплекса из данных четырёх форм из «модели языка» [6].

Наибольший интерес для белорусского языкознания сегодня представляют именно статистические машинные переводчики, а точнее, один конкретный статистический переводчик, разработанный поисковой системой Google¹. Данный продукт по совокупности характеристик не имеет аналогов: 1) Это единственный *статистический* машинный переводчик подобных масштабов, поддерживающий белорусский язык; 2) Это единственный на сегодняшний день *машинный* переводчик (речь идёт как о традиционных, так и о статистических переводчиках, поддерживающих белорусский язык) *высокого качества*; 3) Наконец, это единственный переводчик такого рода, функционирующий в сети интернет в свободном доступе. Фактически, для человека (особенно для представителя дальнего зарубежья), только начинающего изучать белорусский язык, Google-переводчик — безальтернативный продукт. Важно также отметить, что корпорация Google, являясь владельцем крупнейшей в мире и популярнейшей поисковой системы, располагает огромными ресурсами для постоянного улучшения своих разработок. Учитывая, что качество статистического машинного перевода, практически полностью зависит от объёма параллельных текстов, с активным развитием белорусскоязычного сегмента сети интернет, Google-переводчик будет постоянно совершенствоваться.

Объектом исследования в данной работе стали возможности сервиса Google при двустороннем переводе глагольной лексики, находящейся на периферии ЛСГ русских и белорусских глаголов восприятия. Особенности состава и функционирования лексико-семантической группы заявленных глаголов уже подробно рассматривались нами в ряде публикаций². Ранее методом «челночного перевода» в сочетании с семантическим анализом нами была построена достаточно полная ЛСГ белорусских глаголов вос-

¹ <http://translate.google.ru>

² 1) Красковский, Н. И. Лексико-семантическая группа глаголов восприятия в русском и белорусском языках / Н. И. Красковский ; Сборник работ 68-ой научной конференции студентов и аспирантов БГУ. — Минск, 2011; 2) Русский язык: система и функционирование (к 90-летию БГУ и 85-летию профессора П. П. Шубы) : сб. материалов V Междунар. науч. конф., 11–12 окт. 2011 г., Минск / редкол. : И.С. Ровдо (отв. ред.) [и др.]. — Минск: Изд. центр БГУ, 2011.

приятия (в белорусской лингвистике отсутствует словарь глагольных ЛСГ) и, одновременно, значительно дополнена исходная ЛСГ русских, приведённая в словаре-справочнике «Лексико-семантические группы русских глаголов», 1988 г. (от словарной статьи с 83-мя рус. глаголами мы пришли к параллельным словарным статьям со 159-ю рус. и 147-ю бел. глаголами) [2]. Также путём анализа фразеологических единиц, содержащих архисему 'восприятие' (для их выделения было проанализировано 4000 русских и 6000 белорусских фразеологизмов), были выявлены смежные группы, наиболее часто пересекающихся с ЛСГ глаголов восприятия в русском и белорусском языках, среди которых в обоих случаях самой частотной оказалась ЛСГ глаголов *перемещения в пространстве* (*доходить до ушей, бросаться в глаза, вочы разбязгаюцца...*), что, вероятно, объясняется, наличием в семной структуре глаголов данной группы «большого количества разнообразных дифференциальных сем, помогающих представить силу, внезапность и т. п. процесса» [3, с. 7].

Равно как любая ЛСГ имеет базовые глаголы (наиболее конкретные и простые по семантике), так и лексический фонд языка имеет базовые ЛСГ (состоящие из частотных, незаменимых в употреблении слов). Именно в силу своей «незаменимости», русские и белорусские глаголы, принадлежащие к лексико-семантической группе восприятия и её периферии, регулярно становятся объектом перевода. Трудно представить художественный текст, в котором бы отсутствовали глаголы, выражающие семантику зрительного, слухового, обонятельного восприятия: *Іра спалохана страляла ў яго вачамі: маўчы!* (І. Шамякін); *Доктор бросил взгляд на бумажку и ахнул.* (Б. Пастернак). Исследователи отмечают, что и «семантические сцепления в тексте художественного произведения часто осуществляются путём взаимосвязи глагольных слов и их перифрастических замен» (*Он посмотрел / кинул взгляд / бросил взгляд / вскинул глаза...; Яна пачула / да яе вушэй дайшло / да яе слыху дайшло...*) [1, с. 11]. Как пишет О. И. Аладьина: «это один из типичных способов создания связного текста» [1, с. 11].

На начальном этапе исследования было экспериментально проверено предположение о том, что, если на уровне фразеологизмов возникают регулярные пересечения двух ЛСГ, то такие пересечения должны происходить и вне устойчивых выражений. Методом сплошной выборки из национального корпуса русского языка³ и десятков произведений классиков белорусской литературы было выявлено, как и ожидалось, значительное количество подобных случаев. Фиксировались исключительно глаголы, принадлежавшие к ЛСГ перемещения в пространстве и выражавшие архисему 'воспринимать' в синтаксической позиции предиката (позиция сказуемого как смысловый центр предложения практически исключает возможность случайного пересечения ЛСГ). Наиболее частотными русскими глаголами, относящимися к ЛСГ перемещения в пространстве, но передающими семантику восприятия стали глаголы: остановиться, поднять, проводить,

³ <http://ruscorpora.ru> .

плыць, хлынуть, лезть, блуждать, пробежать, долетать (*Особенно лезло в глаза распоротое брюхо дивана...* (М. Горький); *...её взгляд остановился на недавно купленной картине.* (Л. Толстой); *И все, кроме неподвижного прокуратора, проводили взглядом Марка Крысобоя...* (М. Булгаков); *В закуте было темно, и звуки долетали сюда отрывисто и глухо.* (Ф. Гладков); *Он тут же развернул ее, пробежал глазами и подивился чистоте и красоте почерка.* (Н. Гоголь)) и др. Наиболее частотными белорусскими глаголами такого рода стали: прабегчы, прайсці, плыць, правесці, паднімаць, слізгаць, лезці, несціся, блукаць (*Дубадзел <...> мутным, важным позіркам прайшоў на тварах гасцей, запыніўся на Ганне.* (І. Мележ); *Нэля засталася на ганку і праводзіла нас паглядам...* (Н. Гілевіч); *Наводчык прыўзнімае салдатаву галаву, і на мне слізгае яго застылы нерухомы позірк.* (В Быкаў); *Лезе ў вочы зьянне месяца. Месяц яркі, як пажар.* (З. Бядуля); *Позірк яе пачынаў блукаць. На вуснах з'явілася жаласная ўсмейка.* (У. Караткевіч)) и др⁴.

При машинном переводе предложений с исследуемыми глаголами на первый взгляд не возникает существенных проблем. Например, фразу *...крик шел из железных домиков* (Г. Галахова) сервис Google переводит как *...крык ішоў з жалезных хатак*, а для предложения *Ад цёплай печы плылі пахі смажаніны* (А. Федарэнка) переводчик предлагает следующий вариант: *От теплой печи плыли запахи жаркого*. Однако при более глубоком рассмотрении проблемы всё же выявляются. Анализ в частности показал, что машинный переводчик абсолютно не соотносит глаголы с эквивалентными перифрастическими выражениями. Так машина не отождествляет, например, синтагму *прабег позірк* с глаголом *паглядзеў*, она лишь производит точный пословный перевод, а следовательно, иноязычный пользователь, захотев перевести на белорусский язык, скажем, глагол *услышаць* никогда не получит на выходе *дайсці да слыху*, программа выдаст буквальное соответствие *пачуць* без каких-либо вариантов. Ясно, что человек, изучающий таким способом глагольную лексику так и останется на базовом уровне знания языка. Вторая проблема буквального перевода в том, что на выходе мы имеем дело с фактически кальками переводимых фразеологизмов, а следовательно, неизученными останутся безэквивалентные по отношению к языку пользователя словосочетания вроде *зрывае нос* (вариант *б'е ў нос*), *кінуць вока* (вариант *кінуць позірк*), *стрыгчы вачыма* (вариант *страляць вачыма*), *метасься в глаза* (вариант *кидацься в глаза*) и т.п.

Основная причина возникновения подобных проблем состоит в том, что большинство перифрастических сочетаний относится к ФСК глаголов, способных выступать в роли предиката во фразах, передающих ситуацию восприятия, и реализуют данное значение глаголы на синтагматическом уровне, в тексте, речи (словари данные значения практически не фиксиру-

⁴ Как видно из примера, соотносимые парадигмы близки: представлены, в основном, словами, являющимися омо/паралексами омо/парасемантами с высокой частотностью.

ют). Наши наработки в области составления параллельных ЛСГ и ФСК русских и белорусских глаголов восприятия могли бы существенно улучшить статистический машинный переводчик (в данном случае сервис Google): как уже было сказано, возможности статического переводчика полностью зависят от объёма и качества параллельных текстов, а мы в наших работах составляем и соотносим именно параллельные парадигмы русских и белорусских глаголов.

ЛИТЕРАТУРА

1. Аладьина, О. И. Функциональные свойства ЛСГ глаголов чувства в их текстовой перспективе (на материале художественной и научной речи): автореф. дис. ... канд. филологич. наук; Лен. ордена трудового Красного Знамени гос. пед. ин-т. имени А.И. Герцена — Л., 1988.

2. Лексико-семантические группы русских глаголов: учеб. слов.-справ. / под общ. ред. Т.В. Матвеевой. — Свердловск, 1988.

3. Нетяго, Н. В. Лексико-семантическая группа глаголов возникновения / исчезновения в современном русском языке (системная организация, связь с другими ЛСГ): автореф. дис. ... канд. филологич. наук : 10.02.01; Лен. ордена Ленина и ордена трудового красного знамени гос. ун-т. — Л., 1986.

4. Учеб. слов.-справ. Лексико-семантические группы русских глаголов / Под общ. ред. Т. В. Матвеевой. — Свердловск, 1988.

5. Филолингвия. Машинный перевод: правила против статистики [Электронный ресурс]. — Режим доступа: <http://filolingvia.com/publ/25-1-0-898>. — Дата доступа: 9.03.2012.

6. Яндекс. Машинный перевод [Электронный ресурс]. — Режим доступа: <http://company.yandex.ru/technologies/translation/>. — Дата доступа: 10.03.2012.