

ОБ ИСПОЛЬЗОВАНИИ КИТАЙСКИМИ УЧЕНЫМИ МЕТОДОВ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ ПРИ ОПРЕДЕЛЕНИИ НЕОЛОГИЗМОВ

M.V. Tokhta-Khodzhaieva

DEFINITION OF NEOLOGISMS WITH THE HELP OF COMPUTER LINGUISTIC METHODS BY THE CHINESE SCIENTISTS

Непрерывное появление в китайском языке большого количества новых лексических единиц, новых речевых оборотов и выражений является объективной закономерностью. Вслед за стремительным социально-экономическим развитием, постепенным увеличением международного обмена и, в особенности, всеобщим использованием Интернета, это явление стало еще более очевидным.

В соответствии с довольно устаревшими, на сегодняшний день, статистическими данными Рабочего комитета по национальному языку и письменности Китая, после начала реформ в китайском языке ежегодно появляется более 800 новых слов и словосочетаний.

Из-за размытости языковых понятий в китайском языке очень трудно дать точное определение новой лексической единицы – неологизма. Тем не менее, китайские лингвисты, опираясь на современные исследования, считают, что к определению неологизма нужно подходить с двух сторон.

1. Если придерживаться лексикографического критерия, то есть включенности лексической единицы в существующие словари, то неологизм – это возникшее различными путями, имеющее новую форму, новое значение или новое употребление слово или устойчивое словосочетание. Особенность неологизма заключается в слове «новый», и эта новизна и проявляется конкретно в форме, значении и употреблении неологизма. Если сопоставить эти три стороны неологизма с формой, значением и употреблением лексических единиц, относящихся к основному словарному фонду языка, закрепленному в словарях, то неологизм будет считаться таковым, если хотя бы по одному из этих признаков на-

блюдается несовпадение со словарными единицами. Эталонными словарями, отражающими словарный запас современного китайского языка, являются «Словарь современного китайского языка» («现代汉语词典») и «Большой словарь китайского языка» («汉语大词典»).

2. С точки зрения времени появления в языке (темпоральных характеристик), неологизм – это лексическая единица, появившаяся в определенный период времени и имеющая новую форму, значение и способ употребления. Лексическая единица является новой только для определенного временного периода. Новой – потому что ее употребление связано с конкретным периодом времени. В качестве примеров можно привести следующие словосочетания, появившиеся в языке после начала претворения в жизнь политики реформ и открытости в Китае в 1979 году, ставшие для того времени неологизмами:

经济特区 [jīngjì tèqū] специальная [особая] экономическая зона;

下海[xiàhǎi] «выйти в море» – бросить свое основное занятие и полностью отдаться коммерческой деятельности;

打工[dǎgōng] «работать; батрачить» – работать без зачисления в штат; устроиться на временную работу (о рабочих).

С точки зрения лингвистики, неологизмы в китайском языке можно подразделить на несколько групп в зависимости от источника их возникновения:

– имена собственные, включая имена людей, наименования мест, названия товаров, фирм, учреждений и т.д.;

– аббревиатуры, например: атипичная пневмония (非典[feidiǎn]), комитет по планированию рождаемости (计生委[jìshēngwěi]) и др.;

– диалектизмы, например: оплатить счет в ресторане (埋单[máidān]);

– вновь созданные слова, например: имейл

¹ Старший преподаватель факультета ГТ АНО ВО «Российский новый университет».

© Тохта-Ходжаева М.В., 2016.

(伊妹儿 [yīmèir]) (это, к тому же, фонетическое заимствование из английского языка);

– специальные термины, например: атипичная пневмония (非典型肺炎[fēi diǎnxíng fèiyán]), Blu-Ray Disc (蓝光光盘[lán guāng guāngpán]);

– транскрибирование слов, например: круто (酷[kù]), шоу (秀[xiù]), клон (克隆[kèlóng]);

– буквенные аббревиатуры, например: WTO (世界贸易组织 [shìjiè mào yì zǔzhī]), APEC (亚太经济合作组织 [yàtài jīngjì hézuò zǔzhī]);

– слова и выражения из основного словарного фонда языка, но изменившие значение и употребление, например:

下课 [xiàkè] – «конец занятий; уйти после занятий» стало употребляться в значении «шаг вниз», «коллапс», «увольняют», «отставка»;

充电 [chōngdiàn] – «заряжать, зарядка, подзарядка», в переносном значении – «набираться знаний, самосовершенствоваться, расширять кругозор; отдохнуть, «подзарядиться»;

– лексические единицы, которые долгое время не употреблялись в языке, но в последнее время опять появились с новым значением, например: «高就[gāojiù]» – «вступить в новую высокую должность, принять почетный пост», в настоящее время широко используется, чтобы спросить людей, где работа является более достойной.

В настоящее время, в связи с огромным наплывом неологизмов в различные сферы деятельности людей, китайские лингвисты, как ученые-теоретики, так и практики, столкнулись с серьезной проблемой определения места неологизмов в системе языка и включения их в словари и справочные пособия, предназначенные для различных категорий пользователей языка – литераторов, журналистов, редакторов, преподавателей, переводчиков, студентов, школьников и т.д. В первую очередь здесь возникает проблема выявления неологизмов в громадном количестве текстов. Как же определить, что данное слово или устойчивое словосочетание является неологизмом? Как при этом избежать субъективного фактора в восприятии какой-либо лексической единицы как неологизма, как не оторваться от объективных научных критериев при его определении? Насколько нам известно из китайских источников, китайские ученые с успехом применяют новейшие компьютерные методы, специально разработанные программы и системы для проведения этой работы. Мы попытаемся вкратце изложить алгоритм работы с использованием таких методов.

Сначала необходимо обратить внимание на основные причины трудности обнаружения неологизмов

в китайском языке, отмеченных китайскими учеными.

1. В китайском языке нет интервалов между словами.

2. Кроме наименований объектов и буквенных сочетаний, нет какой-либо специфической словообразовательной модели, по которой можно было бы определить, что данное слово является неологизмом.

3. Сложно идентифицировать новые слова и формы слов с низкой частотностью употребления.

4. Еще труднее идентифицировать старую лексику, в значении и употреблении которой произошли изменения.

По данным современных компьютерных исследований, большинство неологизмов составляет группа имен собственных, в которой самый высокий процент (свыше 90%) компьютерной идентификации неологизмов имеют имена людей (антропонимы) и географические названия (топонимы). Что же касается названий различных учреждений, то там аналогичный процент гораздо ниже из-за сложности их словообразовательной структуры. Исследований по другим группам имен собственных очень мало, но, как отмечается, в целом процент распознавания еще ниже. Так как некоторые группы имен собственных уже достаточно исследованы, мы подробнее рассмотрим другие группы неологизмов – лексических единиц с новой формой.

В настоящее время в Китае существуют два метода компьютерного исследования таких неологизмов:

1) метод разработки свода правил, при котором идентификация неологизмов происходит путем создания специальных лексических глоссариев и списка словообразовательных моделей для последующей обработки языкового материала компьютером;

2) статистический метод: в обследуемых текстах выделяются лексические цепочки с высокой частотностью употребления, затем удаляются те из них (так называемый мусор), которые заведомо нельзя отнести к неологизмам. При этом дополнительно используется и метод сочетаемости, выделяются наиболее частотные сочетания иероглифов.

Главным недостатком первого метода является ограниченность какой-либо сферой, к тому же требуется сначала разработать сам свод правил. Целью способа автоматического распознавания новой лексики, о котором идет речь в данной статье, является максимальная обработка веб-страниц с целью нахождения новых лексических

единиц, не ограниченных определенной сферой употребления или количеством входящих в их состав иероглифов. Этот способ уже активно используется при создании электронной версии «Словаря неологизмов современного китайского языка» и в значительной степени облегчает работу по идентификации неологизмов. Система, основанная на данном методе, имеет степень точности от 30% до 40%.

Неологизм представляет собой комбинацию уже существующих символов или имеющихся в языке лексических единиц. Но конкретная комбинация только тогда может стать устойчивой, быть признанной в качестве неологизма, а не временного или случайного сочетания смыслов или иероглифов, если она регулярно повторяется в речи. Регулярная повторяемость является непременным условием признания слова неологизмом. Другое условие – давность появления. Новизна лексической единицы должна ощущаться носителями языка. Когда она стирается, слово выходит из разряда неологизмов. Например, в истории китайского языка второй половины XX века это произошло с таким устойчивым словосочетанием, как «особая экономическая зона», которое возникло сразу после начала проведения политики реформ и открытости в конце 1970-х гг., еще раньше этот процесс прошли «продовольственные карточки», «талоны на ткань» и другие, появившиеся после образования КНР.

Итак, новое слово должно отвечать двум условиям:

- 1) регулярная повторяемость;
- 2) появление и распространение в относительно недавний момент времени.

Стоит отметить, что в момент своего возникновения в языке неологизмы часто, если не в подавляющем большинстве случаев, являются непонятными для абсолютного большинства носителей языка, так как обозначают новые понятия, это термины-неологизмы. Для того чтобы правильно понять значения такого термина, требуется провести его словообразовательный и семантический анализ. Например, медицинский термин 非典 [fēidiǎn] «атипичная пневмония», являясь неологизмом-аббревиацией, без подобного анализа не понятен, следует сначала осознать, что его компонент 典 [diǎn] выступает здесь как заместитель слова 典型 [diǎnxíng] «типичный; классический», из которого он получился в результате сокращения (морфемной контракции). Также следует понимать, что полностью опущен третий компонент полного состава термина-неологизма – 肺炎 [fèiyán] «пневмония». Лишь

после достаточно многократного употребления в текстах СМИ аббревиация становится понятной языковому коллективу без специального пояснения или анализа (впрочем, это происходит за весьма короткий период времени).

Таким образом, автоматический поиск неологизмов происходит следующим образом: в первых, проводится глобальная обработка веб-сайтов, затем в веб-контенте выделяются повторяющиеся цепочки, а в них уже выявляются лексические единицы – кандидаты в неологизмы. Затем проводится фильтрация – и в итоге высвечивается неологизм.

Цель поиска повторяющихся цепочек – выделить все цепочки, повторяемость которых в одном тексте выше определенной частоты. В процессе поиска за единицу можно брать как иероглиф, так и слово. Метод поиска повторяющихся цепочек, основанный на слове, считается более эффективным, чем иероглифический метод, так как позволяет значительно сократить число так называемых мусорных цепочек. Например, возьмем отрезок речи «...中国人民...中国人...». Если анализировать его с помощью иероглифического метода, то мы получим повторяющуюся цепочку «中国人». Однако при использовании словесного метода при анализе отрезка речи «...中国人民...中国人...», «中国人» уже не будет выделено как повторяющаяся цепочка.

Непрерывное появление большого количества новых слов в китайском языке является объективной закономерностью, причем вслед за всеобщим использованием Интернета это явление становится еще более очевидным. В связи с этим изучение появления новых слов в китайском языке обладает важным практическим значением, является актуальной потребностью. В данной статье предложен метод автоматического обнаружения новой лексики, когда путем крупномасштабного анализа веб-сайтов устанавливается огромное множество слов и словосочетаний, среди которых выявляются повторяющиеся цепочки. После автоматической проверки их в соответствии с правилами словообразования проводится дальнейшая фильтрация и, наконец, из собранного языкового материала извлекается новая лексика. Результат экспериментальных исследований материалов газет «人民日报» [rénmín rìbào], «华东新闻» [huádōngxīnwén] и «江南日报» [jiāngnán rìbào] свидетельствует о том, что процент точности обнаруженной новой лексики достигает примерно 30. Эксперимент показал, что метод, описанный в данной статье, дает возможность обнаружить большую часть неологизмов в обследованных текстах. Системой актив-

но пользуются при составлении «Современного электронного словаря неологизмов китайского языка», что сильно облегчает эту работу.

Литература

1. 基于构词法的网络新词自动识别初探/郑家恒, 李文花. – 山西大学学报(自然科学版). – 2012, 25 (2). – С. 115–119 // Чжэн Цзяхэн, Ли Вэньхуа. Автоматическая идентификация

интернет-неологизмов на основе принципов словообразования. – Шаньси: «Цзыжань кэсюэбань», 2012. – С. 115–119.

2. 我的新生词语规范观/张德鑫, 水至清, 则无鱼. – 北京大学学报(哲社版), 2000, 2005. – С. 106–119 // Чжан Дэцин, Шуй Чжицин, Цзэ Уюй. Мой взгляд на новые речевые обороты. – Пекинский университет, «Чжэшэбань», 2000. – С. 106–119. – [Электронный ресурс]. – <http://www.hubce.edu.cn/cbb/qwjs/lib/33118.html>