способ обновления сленга: *обезьянник* (скамейка для задержанных в милиции)*; корочки (диплом); травка (наркотики)* и др.

Обновляется и сленговая фразеология: *звездюлей надавать (наказать); как трусы без резинки* (быть свободным)*; шнурки в стакане* (родители дома)*; ясен перец* (понятно).

Бесспорно, увлечение сленгом свидетельствует о низкой речевой культуре говорящего, о примитивном восприятии окружающего мира. Для большинства выбор сленга в качестве средства общения оказывается временным, «подростковым». Но процессы функционирования сленга, его формирования и изменения всегда интересны как объект исследования филолога.

А. С. Мартынович (Минск, Беларусь)

ЛЕКСИЧЕСКИЙ ПОДХОД К ОБЕСПЕЧЕНИЮ КОРРЕКТНОСТИ АВТОМАТИЧЕСКОГО АНАЛИЗА ОЦЕНОЧНЫХ ИНТЕРНЕТ-ОТЗЫВОВ

Формирование общественного мнения – многофакторный процесс, в котором оценочные высказывания играют едва ли не ключевую роль. Они раскрывают предмет речи с разных сторон, организуют при необходимости разновекторную общественную полемику.

Огромное количество мнений пользователей о различных объектах, явлениях, событиях размещается на интернет-ресурсах. Обобщение и систематизация оценочных высказываний, а также вывод о целостном общественном мнении относительного того или иного объекта, явления, события становятся возможными благодаря автоматическому анализу. В нашем исследовании анализу подвергнуты оценочные высказывания о предприятиях гостиничного или ресторанного бизнеса (о его персонале, услугах, ассортименте, качестве и стоимости блюд, интерьере, месте расположения и т. д.).

Лингвистическая база данных позволяет создать формальную модель оценки, в основу которой закладывается правильное выделение и анализ оценочных лексических единиц из текста отзыва о предприятии гостиничного или ресторанного бизнеса. Для проверки работоспособности модели был создан компьютерный анализатор, написанный на языке программирования Python.

Источниками отзывов послужили специализированные сайты *http://www.tripadvisor.com* и *http://www.booking.com*,особенностью которых является субъективный характер представленной информации (тексты отзывов содержат эмоционально-оценочную составляющую).

В лингвистическую базу данных вошли лексические единицы, отражающие оценку объекта в целом, аспектные категории (*service, room, cuisine*) и аспектные термины

(*atmosphere, location, noise, quality, speed of service, staff politeness, size, comfort, design*), а также дополнительные списки лексических единиц, семантически связанных с оценкой: список слов и/или словосочетаний-интенсификаторов (*absolutely, always, awfully, better, by far, completely, deeply, especially, extremely, highly, less, more*) и список слов-инверторов (*aren't, isn't, least, never, no, non, not, nothing, wasn't, weren't, without*). К графическим способам усиления оценки в материале исследования были отнесены прописные (заглавные) буквы, восклицательный знак, правая круглая скобка, условно обозначающая улыбку (положительная оценка), левая круглая скобка, условно обозначающая неодобрение (отрицательная оценка).

В списке оценочной лексики каждая словоформа была наделена определенным семантическим весом в диапазоне от +3 (сверхположительное оценочное значение) до –3 (сверхотрицательное оценочное значение). Например: *The service is excellent (+3); We had a nice (+1) dinner at this restaurant; The waiters were slow (–1); Overall we were disappointed (–3).* Положительное слово-интенсификатор, стоящее перед положительным оценочным словом, увеличивает вес данного слова на 1. Например: *The portions are quite big (+2).* Этот же интенсификатор, стоящий перед отрицательным оценочным словом, увеличивает вес данного слова на –1. Слово-инвертор, стоящее перед словом с положительным семантическим весом, меняет направление его оценки с положительной на отрицательную (и наоборот).

В ходе автоматического анализа из массива текстов выбирается очередной текст отзыва о конкретном предприятии. Каждое выделенное слово в тексте отзыва сравнивается с единицами лингвистической базы данных. Если выделенное слово является аспектной категорией, аспектным термином либо ключевым словом, то его ближайший контекст проверяется на наличие в списке оценочной лексики. Согласно списку оценочной лексики, в счетчики добавляются весá для каждого аспекта и связанных с ним ключевых слов. Далее происходит проверка наличия перед оценочными единицами слов-интенсификаторов и слов-инверторов. Аналогично проводится проверка на наличие в предложении текста отзыва графических способов усиления оценки. Затем компьютер суммирует показатели счетчиков для каждой аспектной категории в общий счетчик веса ее положительной или отрицательной оценки для отдельного текста отзыва о конкретной гостинице или ресторане. Сравнивая общие счетчики положительной и отрицательной оценки, система делает вывод о направлении оценки данного аспекта. После анализа всех текстов отзывов о конкретном предприятии компьютер суммирует показатели всех счетчиков в единый счетчик веса положительной оценки и единый счетчик веса отрицательной оценки каждой аспектной категории. Сравнивая счетчики положительных и отрицательных весов, система делает

окончательный вывод об общественном мнении, касающемся конкретного предприятия гостиничного и ресторанного бизнеса Республики Беларусь.

Проделанная работа показала, что положенный в основу лексический подход позволяет, в целом, правильно извлекать оценочные суждения пользователей о гостинице или ресторане и корректно формулировать вывод об общественном мнении. В то же время точное определение компьютером синтаксических структур, входящих в текст отзыва, может представлять проблему, поскольку язык анализируемых высказываний – это язык преимущественно разговорный в письменной форме. Авторы не всегда соблюдают расстановку знаков препинания, допускают орфографические ошибки. Наличие неправильно написанных слов препятствует корректному анализу текста, выделению как оценочных слов, так и объектов оценки. Частично решить указанную проблему поможет процедура нормализации лексических единиц.

На правильное выделение оценочных слов значительное влияние оказывает омонимия, например: *like* – глагол (оценочный) и *like* – союз (не содержит оценку), *little* – прилагательное (оценочное) и *little* – наречие (интенсификатор). Для решения данной проблемы необходимо проводить дополнительную процедуру тегирования текста по частям речи.

Стоит также упомянуть о специфике употребления некоторых оценочных слов. Так, положительное слово *high (+1)* в сочетании с названием аспектной категории *cost*меняет направление оценки на противоположное. Необходимо выявить все подобные случаи и разработать дополнительные правила лексической сочетаемости, влияющие на определение веса оценки.

Компьютер, основываясь только на правилах поиска оценочных слов в ближайшем контексте ключевого слова конкретной аспектной категории, не может правильно извлекать информацию из сложных синтаксических конструкций. К примеру, при обработке предложения *The breakfast had everything that could be wished* система распознаёт аспектную категорию *cuisine* по ключевому слову *breakfast*, однако не может правильно определить ее оценочный характер, так как в данном случае она выражена не лексической единицей, а синтаксической конструкцией *had every thing that could be wished.* В силу этого предложенная нами формальная модель может быть дополнена рядом синтаксических правил, учитывающих грамматическую структуру оценочных конструкций.

Таким образом, несмотря на обозначенные сложности формальной модели оценки, лексический подход позволяет более или менее правильно извлекать оценочные суждения пользователей о предприятиях гостиничного или ресторанного бизнеса и на этой основе корректно формулировать вывод об общественном мнении относительно объекта оценки.