

СЕМАНТИЧЕСКИЙ АНАЛИЗАТОР ГЛАГОЛЬНО-ПРЕДЛОЖНЫХ КОНСТРУКЦИЙ СО ЗНАЧЕНИЕМ МАТЕМАТИЧЕСКИХ ДЕЙСТВИЙ

При обработке естественного языка возникает проблема распознавания анализатором глагольно-предложных конструкций со значением математических действий, особенно при анализе узкоспециальных текстов, где такие конструкции составляют значительную часть всей лексики.

Глаголы, которые встречаются внутри этих конструкций, могут иметь два значения: для ясности назовём их «бытовое» и «математическое». В первом случае глагол сохраняет свой основной смысл, во втором – его семантика встраивается в контекст математического термина и может ощутимо измениться.

Специфика конструкций со значением математических действий состоит в том, что основной глагол, использованный в них, как правило, встречается гораздо чаще сам по себе, в своём достаточно абстрактном бытовом значении: в конструкциях «касаться окружности», «лежать на прямой», «приводить к общему знаменателю», например, глаголы «касаться», «лежать» и «приводить», будучи взяты отдельно от контекста, прямо не указывают на возможность их математической семантики.

Анализаторы не всегда справляются с задачей правильного разбора таких конструкций, и порой из-за неверно определённой семантики страдает и синтаксис. Однако главная проблема – это перевод основного глагола на любой иностранный язык, в котором данная конструкция отличается от своего аналога в русском языке. Например, словосочетание “выделить полный квадрат” на английском языке будет звучать как ‘to complete the square’, то есть, при дословном переводе, “заполнить квадрат”. В некоторых онлайн-переводчиках встречаются неверные переводы – например, ‘to select a full square’, что совершенно не соответствует семантике конструкции.

В настоящий момент разрабатывается семантический анализатор, способный определить, в каком из значений, бытовом или математическом, был употреблён глагол внутри глагольно-предложных конструкций со значением математических действий. Были выбраны 100 глаголов, охватывающие практически весь массив употребляемых глаголов со значением математических действий, собраны в словарь, являющийся базой анализатора. Для каждого глагола, входящего в него, определены все возможные бытовые и все возможные математические значения, каждому из которых дано определение. Далее в словарь были добавлены валентности – потенциально присоединяемые к глаголу зависимые слова или классы слов, образующие вместе с ним устойчивую конструкцию. Помимо зависимых слов, полезно добавить в словарь фразеологизмы с использованием конкретного глагола. Определение валентности требует масштабного исследования статистики сочетаний с помощью корпуса, составленного из текстов, содержащих в том числе глаголы в их математическом значении. Пример записи глагола в словаре демонстрируется в табл. 1. Курсивом выделены некоторые валентности; можно заметить, что, как и в большинстве случаев, бытовые значения преобладают над математическими. Поэтому ведётся работа по выявлению валентностей из корпуса математических текстов.

¹ Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Таблица 1. Пример словарной статьи

ВЫКОЛОТЬ	
Математическое значение	Бытовое значение
<p><i>(точку)</i> исключить точку из промежутка</p>	<p>1) <i>(глаз)</i> проткнуть острием</p> <p>2) <i>(татуировку)</i> прокалывая чем-либо острым, нанести какой-либо узор, рисунок и т.п.</p> <p>3) <i>(глыбу)</i> вырубить из массы чего-либо</p> <p>4) <i>хоть глаз выколи</i></p>

Опираясь на словарь с размеченными конструкциями, можно создать семантический анализатор, который будет определять тип значения глагола, встретившегося в тексте. Рассматривается несколько возможностей построения анализатора: разбор на основе правил, машинное обучение, комбинация правил и машинного обучения, база данных. Наиболее целесообразным на данный момент представляется построить анализатор, использующий словарь с валентностями как базу данных.

Потенциальные проблемы, связанные с работой анализатора, могут заключаться в том, что существуют глаголы, математическое значение которых практически совпадает с бытовым: например, в словосочетаниях “выполнять преобразование” и “выполнять поручение” семантика глагола остаётся практически неизменной. В таких случаях решение о смысле глагола должно приниматься анализатором на основе валентностей.

Возможное применение анализатора состоит в использовании его при переводе узкоспециальных текстов с русского языка на английский или любой другой иностранный язык. В анализатор легко добавить модуль перевода конструкции, который будет основываться на базе данных с правильными эквивалентами глагола на других языках. Тогда анализатор может быть задействован при переводе глагольно-предложных конструкций, чья семантика осталась неоднозначной после разбора.

Список литературы:

1. Тузов В.А. *Компьютерная семантика русского языка.* – СПб. Изд-во С.-Петербур. ун-та, 2004. – 400 с.