

РАСПОЗНАВАНИЕ ОМОНИМОВ И ЛЕКСИКО-СЕМАНТИЧЕСКИХ ВАРИАНТОВ СЛОВ ЕСТЕСТВЕННОГО ЯЗЫКА

Д.С. Яровая, Д.В. Личаргин

научный руководитель канд. тех. наук Личаргин Д.В.

Сибирский федеральный университет

На сегодняшний день широко распространены и разрабатываются разнообразные системы перевода, нуждающиеся в распознавании омонимов и лексико-семантических вариантов.

Проблема полного понимания текста является актуальной в связи с тем, что качество перевода напрямую зависит от четкого определения значения слов, в частности, омонимов и лексико-семантических вариантов слов (ЛСВ).

Проблема омонимии давно и широко исследовалась различными авторами, в частности Л.А. Булаховским, С.О. Карцевсим, Л.В. Щербой.

Цель данной работы состоит в изучении контекстуального поведения омонимов и лексико-семантических вариантов слов в тексте с учетом семантических шаблонов.

Задачи данной работы заключаются:

1. В описание существующих результатов в области исследования значений слов;
2. В рассмотрении понятий, таких как омонимы и лексико-семантические варианты с точки зрения их позиции в семантических шаблонах, спроецированных на текст.

Основная идея работы состоит в построении модели распознавания слов на основе имеющихся таблиц, а также возможности применения этой модели в учебном процессе при обучении принципам перевода.

Предлагаемая модель – состоит в использовании таблиц семантических шаблонов с учетом простейших алгоритмов, распознающих омонимы и ЛСВ в виде вариантов контекстуального употребления слов в соответствии с шаблонами.

Новизна работы состоит в анализе понятий омонимии и ЛСВ с точки зрения выделенных из пространства семантической классификации шаблонов.

При изучении иностранного языка существует проблема частичного непонимания устной речи. Кроме того, при достаточно высоком уровне знания грамматики и наличия лексического запаса слов и терминов, выявляются серьезные ошибки при использовании иностранного языка. Источником этой проблемы часто являются омонимы.

Омонимы вызывают определенные трудности у изучающих английский язык, так как они звучат одинаково, но имеют различное написание и значение. Например, deer (олень) – dear (дорогой). И эта схожесть в словах иногда приводит к замешательству, у тех людей, для кого английский язык не является родным.

Основная идея работы состоит в построении модели, с помощью которой можно будет понимать разницу между омонимами и ЛСВ, а также научить понимать в каком смысле слово используется в широком или узком контексте.

Для начала необходимо дать определение омонимам:

Омонимы (от греч. *homos* – одинаковый и *опута* – имя) – это разные по значению, но одинаково звучащие и/или пишущиеся единицы языка, в частности слов, морфем и др. (см. Таблицу 1, 2).

Homophones (омофоны) – это слова, которые произносятся одинаково, но пишутся по-разному и имеют разные значения (buy- купить; bye – до свидания). Как мы

могли заметить основная сложность для изучающих английский язык заключается в том, чтобы на слух правильно определить, какое слово используется в том или ином предложении. Интересно, что носителям языка, омофоны не причиняют такого дискомфорта.

Как в русском, так и в английском языках одно слово может иметь несколько значений. Также слова одинаковые по звучанию, но разные по написанию имеют совершенно разные значения. Очень часто предложения, содержащие омонимы, вызывают недоумение у не подготовленного собеседника. И на начальных этапах изучения языка возникают большие трудности. Человек просто не сможет правильно перевести предложение, и тем самым понять смысл.

I cannot open this can without your help. – Я не могу открыть эту консервную банку без твоей помощи.

В этом предложении слово CAN встречается два раза. В первом случае это модальный глагол, переводится как «мочь, иметь возможность». Во втором случае слово переводится как «консервная банка». Не многие люди знают о втором варианте, следовательно, они не смогли бы правильно перевести это предложение.

Еще пример:

Cat plays with a mouse. Ann has broken her mouse.

Слово mouse в первом предложении переводится как животное, мышь, во втором как компьютерная мышь. Также можно заметить, что если неправильно перевести слово, то это может перерасти в шутку.

Таких примеров можно привести как угодно много. Как уже можно было заметить, что для успешного изучения языка, нужно сразу или последовательно учить все значения каждого слова. Помимо омонимов, существуют такие понятия как омофоны и омографы.

Омофоны – слова одинаковые по произношению, но разные по значению, и написанию. Омографы – одинаковые по написанию, но разные по значению, и произношению.

Самый действенный способ изучения употребления омонимов – это практика. Чтение помогает различать такие слова и понимать их возможный смысл. В то время как письменная речь позволяет закрепить полученные навыки. При этом аудирование тренирует улавливание омонимов в устной речи.

Еще несколько примеров слов, где имеет место одинаковое произношение, но разное значение:

- BANK – BANK (насыпь/берег – банк): the bank of the river; the Bank of England; a bank account.
- CARAT – CARROT (карат – морковь): a two-carat diamond; grated carrots.
- GILT – GUILT (позолота/позолоченный – вина): to cover with gilt; we have no proof of his guilt.
- RACE – RACE (паса – скачки/гонка): the human race; horse racing; a race horse; race cars.

В соответствии с формой омонимы подразделяются на омофоны, омографы и абсолютные омонимы. Таким образом, повторим, что омофоны (homophones < Gk homos – same, phono – sound), как следует из их названия, обладают одинаковой звуковой формой, но различаются не только семантически, но и графически (bare/bear, road/rode, etc.). Омографы (homographs < Gk homos – same, grapho – write), напротив, совпадают графически, но читаются различно (row, read, bow, etc.).

Абсолютные омонимы обладают общей звуковой и графической формой, различаясь семантикой и нередко принадлежностью к частям речи (bear – carry, bear – animal; match – contest, match – person; match – fit).

Таблица 1

Пример употребления одного из омонимов

I я	need to обожать	set поставить	the ... этот...	on the ... на этом...
we мы	love to любить	place установить	box коробка	shelf полка
we мы	like to нравиться	lay подложить	package пачка	window sill подоконник
you вы	tend to быть склонным	put положить	pack пачка	table стол
they они	dislike to не любить	lay > out разложить	bag сумка	basis базис
they они	hate to ненавидеть	spread стелить	basket корзина	stand стенд
he..(e)s он..ет/ит	desire to страстно желать	locate разместить	dish блюдо	stand вешалка
she..(e)s она..ет/ит	wish to желать		container контейнер	closet шкаф

Полисемия — многозначность, много вариантность, то есть наличие у слова двух и более значений, исторически обусловленных или взаимосвязанных по смыслу и происхождению.

Под полисемией понимается многозначность слов как единиц лексики. Но правильнее её назвать лексическая полисемия — соответствие одному слову нескольких, во-первых, явлений действительности, во-вторых, смысловых понятий, семантически связанных друг с другом. Наличие общего семантического признака отличает полисемию от омонимии и омофонии.

Приведем некоторые примеры:

- полисемия (разные ЛСВ): keep – держать, сохранять, хранить; cold – холодный, насморк; back – обратно, спина;
- омонимия: lean – наклон, худой; toast – гренки, тост;
- омофония: buy (покупать) – by (около) – bye (пока); coarse (грубый) – course (курс);

Полисемия может быть, как грамматической, так и лексической. Примером первой может служить полисемия 2-го лица единственного числа русских глаголов: «Ты этого не поймешь» и «Тут ничего не поймешь», или для артикля «the» в английском языке, выполняющего как уточняющую (The tiger was old), так и обобщающую функцию (The tiger is a cat-like animal).

Далее дадим определение понятию лексико-семантический вариант (ЛСВ). ЛСВ — одно из понятий, которое соответствует словоформе рассматриваемого слова. Причем, значения этих понятий пересекаются в виде множеств семантических элементов, в отличие от омонимов, где разные значения случайно схожих по форме слов практически не пересекаются, или пересекаются на уровне отдаленных ассоциаций. Мы выяснили, что слово само по себе не имеет, как правило, одного конкретного значения. Слово получает каждое конкретное значение только под влиянием других слов, т.е. в контексте. По выражению Шехтмана Н.А. «в тексте всегда есть знаки, функция которых состоит в том, чтобы указывать на элементы значения других

знаков». Рассмотрим еще один шаблон, как пример подмножества векторизованной классификации для распознавания варианта значения употребляемого в тексте.

Таблица 2

Пример употребления альтернативного омонима со своим значением

It is ... это (является) ...	for the ... для этого ...	to ... в ...	the ... этот ...	of the ... этого ...
easy for легко	scientist ученый	scan просматривать	rough copy черновой вариант	book книга
hard for тяжело	professor профессор	read читать	draft черновик	newspaper газета
interesting for интересно	doctor доктор	write писать	typescript машинописная копия	paper газета
boring for неинтересно	bachelor бакалавр	print напечатать	table (of from) таблица	magazine журнал
funny for забавно	student студент	publish опубликовать	copy копия	journal журнал
ridiculous of нелепо	philosopher философ	understand понимать	edition издание	report доклад
important for важно	psychologist психолог	tear рвать	column (of in) колонка	piece of writing произведение
necessary for важно	physicist физик	type печатать	cell ячейка	article статья
unimportant for неважно	philologist филолог	correct исправлять	page страница	story рассказ
clever of умно	biologist биолог	edit редактировать	double page разворот	literature литература

С помощью таблицы мы составим предложение «It is easy for scientist to understand the table (информация) of the report».

В противоположность таблице 1, задающей фразы типа: «I need to put the bag on the table».

В семантической классификации, каждая точка, это - отдельный омоним или ЛСВ, а в действительности словам соответствуют кластеры точек, таким образом, нам надо из кластера выбрать только одно понятие – точку смыслового пространства на основе функциональных отношений к другим точкам – словам из того же шаблона как подмножества многомерного пространства слов. Эту «точку» можно выбрать на основе таблиц, что может применяться как иллюстрация при изучении языка в школе и вузе.

Выводы. В работе выполнен анализ омонимов и ЛСВ с точки зрения их позиции в структуре семантических шаблонов для анализа текста. Подчеркивается важность продолжения исследований по алгоритмизации процесса распознавания позиции ЛСВ и омонимов в семантических шаблонах. Делается вывод о необходимости учета семантически структурируемого и распознанного контекста на уровне синтагм и структур фраз и повествований.