

Классификация и преподавание: русских глаголов

Мигель Анхель Баэна Шотт

КазНУ — Ал-Фараби

(Научная презентация)

(Дата: 14 мая 2024 г.)

В настоящее время сфера преподавания языков быстро меняется, но не всегда эффективно используются технологии. Бессмысленное повторение форм спряжения глаголов лишь ограничивает навыки студентов. Следуя строгой системе, мы можем отойти от старого метода запоминания глаголов. Российский лингвист Андрей Зализняк уже сформировал критерии классификации русских глаголов, используя которую мы можем предложить материал для усвоения глаголов с учётом интересов каждого студента. В этом докладе предлагаю новую систему для того, чтобы диагностировать те недостатки, которые встречаются в речи студентов, изучающих русский язык, и оказать помощь в решении частных проблем.

I. ВВЕДЕНИЕ

"Главная гипотеза теории заключается в том, что усвоение языков происходит только одним способом: путем понимания сообщений".

— *Stephen D. Krashen*

Один из главных вопросов в изучении овладения языком заключается в том, что мы не знаем, что представляет собой процесс заимствования языка. Считается, что главным механизмом этого процесса является речь как исходный входящий сигнал. Другими словами, надо понимать эти сигналы, чтобы понять, как использовать грамматические структуры. Я утверждаю, что именно глаголы являются одной из сложных тем в усвоении языка. Например, чтобы понять правила использования глагола, надо обратить внимание на его управление, вид, спряжение, приставки и т. д. Согласно идее касательно входных сигналов, надо слышать все грамматические конструкции, чтобы студент принимал новые конструкции в свою речь. Однако, из-за сложностей с глаголами, студенты не всегда успевают слышать все вариантные формы каждого глагола, и в результате чего студентам трудно понимать, как они работают. В качестве примера, вспомним виды глагола. Зачастую нам сложно определить разницу между глаголами совершенного вида и несовершенного вида, и труднее, когда мы не уверены, есть ли у глагола оба вида. Например, глагол "твориться".

Я же предлагаю систематизированную систему, которая поможет выучить глаголы, обратить внимание на их использование в речи. Однако, чтобы применить эту систему необходимо диагностировать частотные недостатки в речи студентов, и только тогда мы сможем предложить им подходящие решения их проблем. Так что, этот доклад фокусируется на вопросе, как можно создать диагностирующую систему, выявляющую у студентов имеющиеся проблемы в усвоении русского языка и позволяющие их устранить.

II. ТЕОРИЯ

Чтобы начать описывать процесс усвоения русского языка, я предлагаю подход, который отличается от традиционных методов. Я считаю важным понимание того, что надо постоянно принимать сигналы в устной речи, чтобы усвоить любой язык. Также считаю немаловажным уметь выявлять погрешности в речи студентов, с целью их исправления, и не тратить попусту время на бесполезные задания. Подробнее поговорим об этом ниже, где речь пойдет о теории входных сигналов, и о том, как это работает с глаголами.

A. Входные сигналы

Теория входных сигналов связана с тем, что дети постоянно воспринимают на слух речь окружающего мира и повторяют те же грамматические структуры, которые извлекают из речи. Благодаря этой способности, все дети могут усваивать свой родной язык. Однако, я считаю, что похожий процесс может быть использован и взрослыми. Как считает Хауард Даурети, автор статьи «Отличие сигнала от шума: стратегии понимания на русском языке», студенты могут понимать тексты на русском языке даже без предварительной подготовки. Исследователь утверждает, что когда есть достаточно контекста, студенты могут догадаться, о чём идёт речь в любом тексте. Конечно, это не значит, что человек полностью понимает текст, но как минимум может извлекать главную информацию.

Стивен Крашен, один из главных пропагандистов теории входных сигналов, отмечает, что человек может усваивать язык, когда он принимает особенно *понятные сигналы*. Для взрослых этот процесс обучения отличается тем, что нет достаточно ресурсов или времени, чтобы им постоянно приводить примеры *хорошей речи*, зато взрослые могут понимать грамматику. Если человек понимает грамматику, то это понимание представляет собой контекст, благодаря которому можно усваивать язык. То есть, чтобы создать

понятные сигналы для студентов, мы должны ориентироваться на грамматический уровень каждого человека. В исследовании Крашена это сделано разными глагольными структурами на английском языке. Но для обучения русскому языку, я утверждаю, что стоит обратить внимание на то, что русские глаголы повторяют разные скрытые шаблоны. Например, что касается спряжения глаголов, я считаю, что ограничение контекста, который можно разбивать на понятные сигналы, отрицательно сказывается для студентов. Когда студенты, опираясь на традиционную грамматику, думают, что существует только первое и второе спряжение, у них нет достаточно контекста для понимания того, откуда возникают все исключения, вспомним смену ударения в глаголах.

Б. Классификация глаголов согласно Зализняку

Советский лингвист Андрей Анатольевич Зализняк описал систему классификации русских глаголов в «Грамматическом словаре русского языка». В его классификации все глаголы разделяются на 16 групп (и некоторые являются исключениями). Каждая группа объединяет глаголы согласно типу спряжения. Например, на первой классификации видно, что все глаголы на *-ать*, *-ять*, *-еть* в форме первого лица ед. числа и третьего лица ед. числа настоящего времени оканчиваются на *-аю*, *-ает*, *-яю*, *-яет* и *-ею*, *-еет*. Обратите внимание на рис. 1 для сравнения первых четырёх классификаций. В контексте данного проекта, стоит разделить все глаголы на разные классификации, потому что тем больше классификации, тем меньше количество исключений слов.

III. МЕТОДИКА

Для того, чтобы систематизировать подход к данному вопросу надо разделить методику на две части. Первая часть заключается в поиске наиболее частотных недостатков в примерах, написанных студентами, в сравнении с *правильной речью*. В контексте данного проекта, носители русского языка в Алматы владеют *правильной речью*.

Вторая часть предполагает исправление выявленных недостатков. В этой части предлагаются разные тактики, которые имеют смысл для каждого индивидуального студента. В следующих пунктах я детально объясню, что представляет собой данный процесс и как именно он осуществляется.

А. Диагностика

Первая часть заключается в том, что необходимо ограничить нашу систему педагогической цели в преподавании глаголов. Именно, надо задать следующий

Номер типа	Инфинитив оканчивается на:	1 ед. и 3 ед. наст. (буд.) оканчивается соответственно на:
1	-ать -ять -еть	— -аю, -ает — -яю, -яет — -ею, -еет
2	-овать -евать (после шипящей или ц) -евать (в прочих случаях)	— -ую, -ует — -ую, -ует — -ую, -уют
3	-нуть	— -ну, -нет
4	-ить	— -ю (но -у после шипящей), -ит

Рис. 1: Первые четыре классификации глаголов, согласно Зализняку. Обратите внимание на то, что спряжения зависят от первого и третьего лица настоящего времени. Таблица взята из «Грамматического словаря русского языка» А. А. Зализняка.

вопрос, *когда студент знает глагол?*. Хотя вопрос звучит просто, мы сразу же сталкиваемся с некоторыми проблемами, когда решаем проверить, какие глаголы знает студент. В процессе этого, Я отметил следующие недостатки:

1. Студент не умеет правильно спрягать глаголы.
2. Глагол не существует в активном лексиконе студента.
3. Студент не использует правильное управление глагола (семантика глагола не понятна студенту).
4. Студент слишком часто употребляет одни и те же глаголы.

Чаще всего, любой студент находится в своем процессе обучения с комбинацией прежних недостатков. Иными словами, студент может уметь правильно спрягать глаголы из первых четырёх категорий (где располагаются большинство распространённых глаголов), но не правильно спрягает глаголы из двенадцатой категории, как *дуть* или *мыть*, а задуть свечу и вымыть руки являются важными действиями[1]. Однако, нужно вспомнить, что речь является двухсторонним процессом. Так что, надо сравнивать речь каждого студента с тем, как говорят именно носители

русского языка. То есть понимать слово и уметь повторить слово, чтобы другие поняли сообщение. Следовательно, если можно сказать, что студент умеет использовать глагол, когда тот глагол появляется в его речи, и это совпадает с тем, как говорит большинство людей на русском языке, то имеет смысл искать следующие положительные и отрицательные случаи в речи студентов:

1. Глагол использован и студентами и носителями с той же частотой (✓)
2. Студенты используют глагол, который не использован носителями (×)
3. Студенты не используют глагол, который использован носителями (×)
4. Глагол не использован ни студентами ни носителями с той же частотой (✓)

Чтобы выявить эти ситуации, необходимо получить примеры письменной формы языка студентов, и сравнить их с тем, как говорят носители языка. Я запустил опрос, чтобы принять разные тексты, которые были написаны этими двумя группами. Затем написал скрипт в Питоне, который соединил все полученные тексты и извлёк все глаголы и просчитал, сколько раз глаголы встречались в сборе материала в каждой группе, студентов и носителей языка. Надо заметить, что все формы каждого глагола были считаны при инфинитивной форме глагола. Например, формы «была», «буду», «будь», считаются при глаголе «Быть». Однако, не имеет смысла бессистемно извлекать все глаголы, если мы уже знаем, что один из глаголов не является частотным в речи большинства людей. Чтобы ограничить поиск, я решил использовать как базу «Частотный словарь русского языка», и искал только те глаголы, которые показываются в списке самых распространённых слов. Согласно *Wiktionary*, я мог бы найти 11995 разных глаголов, но я ограничил процесс только на 1397 глаголах, согласно частотному словарю, где я нашёл ≈ 500 глаголов, использованных двумя группами.

Однако, одна большая проблема заключается в том, что иногда слова имеют смысл только в отдельно взятых контекстах, и некорректно общую языковую ситуацию. Например, если в опрос попало эссе студента факультета геологии, то такие слова как «минерал» или «отложение» являются распространёнными, в то время, как в речи обычного человека они встречаются крайне редко. Так что, надо было найти другую метрику, которая может определить важность каждого слова. В связи с этим, надо было найти такой фактор, с которым мы можем сравнивать сборы текстов друг с другом, без уклона на носителей или на студентов. Мы вернёмся к этой дилемме, когда обсудим анализ данных.

a. Откуда происходит информация этих глаголов?

Необходимость собирать глаголы для этого проекта возникает от того, что главной программой для заучивания глаголов будет программа *Anki*, позволяющая запоминать информацию интервальным повторением. Благодаря эффективности этой программы, я решил использовать карты *Anki* в работе со всей значимой информацией. Однако, существуют тысячи разных глаголов в русском языке, а вручную записывать карты *Anki* потребовало бы слишком много времени. Поэтому я использовал библиотеку (library) *Python*, которая называется *Genanki* для того, чтобы автоматически заполнять карты в то время, как я решаю, какие глаголы важны. Но проблема в том, что *Genanki* создаёт карты только с уже существующей информацией. Когда 1397 глаголов являются частотными согласно корпусу русского языка, вручную писать эту информацию для *Genanki* всё же занимает много времени. Тогда я решил обратиться к интернет-словарю *Wiktionary*.

В словарной статье *Wiktionary* можно найти всю информацию по каждому слову (в принципе) на любом языке. Этимология, определение, схожие слова, а главное для нас, спряжение — всё можно найти. Помимо этого, все статьи, организованные в той же форме, можно трансформировать одним скриптом *Python*, чтобы сделать *webscrapping*. Кратко говоря, это — техника, чтобы автоматически и систематизировано получить данные из интернета. В рисовании 2 можно заметить структуру спряжения глагола *ахать*, а в рисовании 3 видно, как полученная информация вглядеться в после *webscrapping*.

	present tense	future tense
1st singular (я)	áхаю áxaju	бúду áхать búdu áxat'
2nd singular (ты)	áхаешь áxaješ'	бúдешь áхать búdeš' áxat'
3rd singular (он/она/оно)	áхаёт áxajet	бúдет áхать búdet áxat'
1st plural (мы)	áхаем áxajem	бúдем áхать búdem áxat'
2nd plural (вы)	áхаёте áxajete	бúдете áхать búdete áxat'
3rd plural (они)	áхают áxajut	бúдут áхать búdut áxat'

Рис. 2: Таблица спряжений глагола *ахать*. Остальная информация следует похожую структуру.

Можно менять детали согласно задаче. Например, если студенту важно знать только деепричастные формы, тогда можно переработать скрипт для того, чтобы он получил только информацию о деепричастиях. В каждом конкретном случае предлагаю индивидуальные решения для разных ситуаций. То есть, после того, как определяем важные глаголы анализом данных, мы получаем информацию об этих глаголах через *webscrapping*. А потом готовим карты *anki*, что-

А согласно критериям, написанным в секции III A, мы сразу решаем, то ли убрать данный глагол, то ли добавить его в список важных глаголов. Однако, в этом также заключается и проблема. Чтобы сумма **TF-IDF** имела смысл, надо описать нулевую гипотезу. Нулевая гипотеза может быть следующей:

*„Сумма терминов **TF-IDF** в обеих сборах является одинаковой”.*

Тогда следующая задача совпадает с испытанием обоих сборов, потому что если нулевая гипотеза не права, это обозначает, что существует разница между суммами. Тогда смотря на наклон суммы, мы можем решить, что сделать с глаголом, согласно остальным критериям из секции III A. Задача звучит несколько тривиально, но это важный процесс в статистических вопросах. Например, если суммы **TF-IDF** слишком близкие друг к другу, можно говорить, что статистически возможно те суммы являются одинаковыми. И в этом случае можем ли мы говорить, что числа 0.073645 и 0.073534 появились не случайно?

IV. РЕЗУЛЬТАТЫ

Благодаря ответам на опрос, 11686 слов, написанные студентами русского языка, были получены. Из-за того, что только 6134 слова происходили из текстов носителей русского языка, я решил собрать большинство текстов[2], которые студенты программы «флагман» читали в течение академического года. То есть 18737 были получены из академического материала. Для трёх групп следующие количества разных глаголов были найдены, 486 в текстах студентов, 545 из носителей языка, и 893 происходили из академического материала. В результате этого, 59 глаголов существовали в текстах носителей языка, но их нет в речи студентов. Помимо этого, материал с урока является самым разнообразным, где 407 разных глаголов существуют в материале, но их нет в речи студентов, и 348 разных глаголов были найдены в материале, но их нет в речи носителей русского языка.

После того, что все коэффициенты *TF-IDF* были подсчитаны, я сравнивал три группы друг с другом, и нашёл, что визуально можно заметить разницы. Например, обратите внимание на следующие графики, где сравниваются несколько примеров из каждой группы. Интересно, что глагол **быть** всегда является самым распространённым глаголом.

После того, что я сравнивал все глаголы, я нашёл следующие количества интересных глаголов.

1. Из группы в Рис. 4, надо использовать реже 68 разных глаголов
2. Из группы в Рис. 5, надо выучить 57 глаголов и использовать реже 117 разных глаголов.



Рис. 4: Сравнение между студентами русского языка и носителями. Например, здесь сразу видно, что глагол *стремиться* является интересным. Из большого количества слов, только каждый четвертый глагол из ста самых распространённых видны, но все числа сохранены.



Рис. 5: Сравнение между студентами русского языка и материалам с урока. Например, здесь сразу видно, что глагол *выходить* является интересным.

3. Из группы в Рис. 6, надо выучить 56 глаголов и использовать реже 1 глагол.

Так что, обязательно надо рассмотреть эти глаголы, чтобы решить, как можно их усваивать. Одно заключение, которое возникает от этих результатов, предлагает, что все выученные глаголы с урока использованы студентами в правильной частности. Однако, также надо спросить себя, почему студенты не усваивают все глаголы. Существуют около 400 глаголов, которых вообще нет в активной речи студентов, зато это не обозначает, что их нет в пассивной речи этой группы.

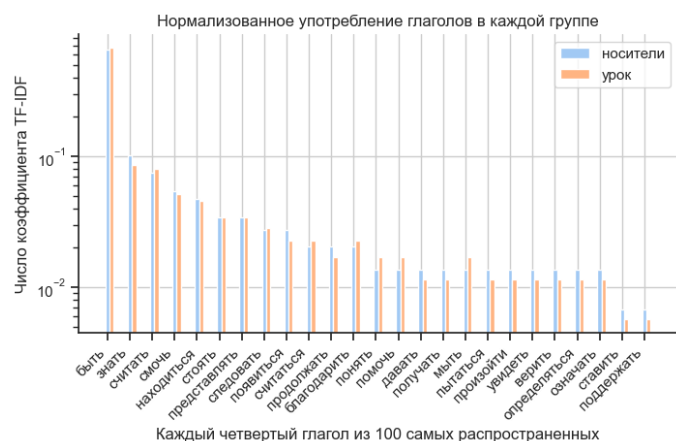


Рис. 6: Сравнение между носителями русского языка и материалами с урока. Например, здесь употребление глаголов более пропорционально

А. Исправление

Для того, чтобы начать исправить частотные недостатки в речи студентов, надо рассмотреть обстоятельства, подсказанные прежним процессом. Вспомогательная часть III A и используя результаты из части IV, мы можем определить каким образом является подход для каждого недостатка.

Ссылаемся на обстоятельства, написанные в части III A. Для каждого студента в программе «Флагман» интересно получить каждый список глаголов из секции IV, чтобы решить с репетитором, каких глаголов нет в речи данного студента. Если студент вообще не умеет правильно использовать глагола, просто надо

выучить глагол. Если студент заметит, что он использует любой глагол чаще, чем надо, просто им будет надо убрать данный глагол из своей речи.

Если у студента проблемы касательно спряжения глаголов, то используя полученные списки, то студент может проверить, правильно ли он умеет спрягать глаголы. В каждом списке есть информация о спряжении, так что это не является трудным вопросом.

V. ЗАКЛЮЧЕНИЕ

Благодаря технологии, можно не только рассматривать разные глаголы в контексте того, как они использованы в настоящей речи, но также в контексте каждого студента в программе «Флагман». Через предлагающий процесс, можно решить какие слова являются пунктами интереса. Если мы уже знаем куда обратить внимание, то мы сможем анализировать где именно располагаются недостатки студентов, и информировать академический план.

Однако, мы сталкиваемся с некоторыми недостатками. Например, чтобы работать с более значительными числами, необходимо собирать больше тексты. Если все флагманские студенты отправили бы свои тексты, можно было лучше сравнивать результаты. Я считаю, что нужны около 100 тысяч слов, чтобы получить лучшие данные. Кроме того, лучше было бы собирать все тексты с уроков письма, академического письма, языка СМИ, и межкультурной коммуникации. Ещё недостаток заключается в том, что структуры данных, которые я придумал в своём скрипте *Python*, требуют много памяти, и в случае мы будем расширять проект, эти структуры должны быть оптимизированы.

[1] Вернёмся к этим примерам из-за того, что отмечаемый процесс в следующей части использует академическую речь, чтобы найти недостатки в речи студентов, и труд-

но верит в то, что студент используют эти конструкции в своей речи.

[2] Только те тексты, которые были доступными в платформе *Canvas*.