

# Грамматические профили и формальная дифференциация русских двувидовых глаголов

Александр Пиперски  
РГГУ / РАНХиГС / ВШЭ

XII Конференция по типологии и грамматике  
Санкт-Петербург, 19.11.2015

# Состав двувидовых глаголов

*венчать, обещать*

*женить, казнить, [крестить,] молвить, ранить*

*велеть*

Некоторые глаголы на *-овать / -евать*:

*воздействовать, исследовать*

*гарантировать, гармонизировать,*

*линчевать*

[Виноградов и др. (ред.) 1952–1954: I, 456]

# Динамика двувидовых глаголов

Некоторые двувидовые глаголы на *-овать* сдвигаются в сторону нормальной видовой парности

*арестовать*

*организовать*

# Появление видовой парности

Суффиксальная имперфективация:

*арестовать — арестовывать*

*конфисковать — конфисковывать*

*организовать — организовывать*

Приставочная перфективация:

*информировать — проинформировать*

*организовать — сорганизовать*

*приветствовать — поприветствовать*

*рекомендовать — порекомендовать*

# Появление видовой парности

Суффиксальные и префиксальные дериваты имеют однозначную видовую принадлежность

Закрепление однозначной видовой принадлежности за исходным глаголом происходит постепенно (ср. *арестовать*)

# Появление видовой парности

Большинство работ ограничивается констатацией появления видовой парности [Зализняк, Шмелёв 1997; Петрухина 2009; ...] или анализом только одного из двух изменений ([Korba 2007] про перфективацию)

**Не объясняется направление сдвига!**

# Направление сдвига

Почему одни двувидовые глаголы имперфективируются, а другие — перфективируются?

1. Потому что они не могут пойти одним из этих путей по формальным признакам
2. Потому что они больше похожи на глаголы СВ или НСВ

# Направление сдвига

Почему одни двувидовые глаголы имперфективируются, а другие — перфективируются?

1. Потому что они не могут пойти одним из этих путей по формальным признакам
2. Потому что они больше похожи на глаголы СВ или НСВ

# Ударение

## и появление видовой парности

Суффиксальный имперфектив не образуется от глаголов на безударное *-ова-* [Мучник 1966; Jászay 1999]:

*информировать* — \**информировывать*

*приветствовать* — \**приветствовать*

Но могло бы быть

*рекомендовать* — *рекомендовывать*

# Направление сдвига

Почему одни двувидовые глаголы имперфективируются, а другие — перфективируются?

1. Потому что они не могут пойти одним из этих путей по формальным признакам
2. Потому что они больше похожи на глаголы СВ или НСВ

# «больше похожи»?

## Что значит «больше похожи»?

Задача классификации:

два заранее выделенных класса,  
по которым надо распределить  
некоторое количество сущностей

Нужно выделить признаки, на основе  
которых будет делаться классификация

# 0. Ручной способ

Просмотреть  $N$  контекстов для каждого двувидового глагола и узнать, в каких значениях он употребляется чаще: СВ или НСВ

Проблемы:

много ручной работы

сложные случаи  $\Rightarrow$

$\Rightarrow$  субъективность в решениях

# Примеры сложного случая

*Во всяком случае, говоря о морали, мы не можем **абстрагироваться** от религии — миллиарды людей находят в ней основу своей нравственности.*

[Александр Маркович. На пути к абсолютной морали (2003) // «Вестник США», 2003.07.23]

# 1. Диагностические формы

1. Глаголы СВ не встречаются в форме причастий непрошедшего времени [Крапивина 2009; Влахов 2010]

Слабо применимо к двувидовым глаголам: от них есть причастия непрошедшего времени, а иначе они не были бы двувидовыми

# 1. Диагностические формы

2. Глаголы НСВ образуют деепричастия на -я, а глаголы СВ — на -в:

*любить — любя*

*полюбить — полюбив*

Если от двувидового глагола образуется больше деепричастий на -я, он больше похож на НСВ, а иначе — на СВ

# Проблемы

Деепричастие — это только одна, причём довольно редкая форма

Возможно, необходимо привлечь к исследованию больше форм

## 2. Грамматические профили

[Janda, Lyashevskaya 2011; Lyashevskaya 2015; Kuznetsova 2015]

Грамматический профиль слова —  
относительная частота его различных форм

Каждому слову соответствует вектор  
значений  $\Rightarrow$  можно производить  
обычные статистические операции

# Грамматический профиль глагола и его вид

**Можно ли узнать вид глагола,  
обладая данными о частотности  
его форм в корпусе?**

Материал исследования:  
основной подкорпус НКРЯ, тексты  
с 1980 года по наст. вр. (точное  
вхождение), ~ 80 млн словоформ

# Глаголы и формы

1001 самый частотный глагол из словаря [Ляшевская, Шаров 2009] (из них 979 охарактеризованных по виду глаголов)

4 формы:

INF	INDIC,(PRAES FUT)
INDIC,PRAET	IMPER

# Образец данных

Глагол	INF	INDIC, (PRAES FUT)	INDIC, PRAET	IMPER	Вид
<i>слышать</i>	0,12	0,31	0,54	0,03	НСВ
<i>смочь</i>	0,006	0,52	0,43	0,05	СВ
<i>заметить</i>	0,14	0,12	0,65	0,09	СВ
<i>использовать</i>	0,49	0,26	0,22	0,03	НА
<i>пытаться</i>	0,05	0,37	0,57	0,01	НСВ

# OneRule

[Holte 1993]

Доля форм презенса  $\geq 0,232 \Rightarrow$  НСВ

Доля форм презенса  $< 0,232 \Rightarrow$  СВ

93,5% качества

(10-проходная кросс-валидация)

# Более сложные методы

Никакие более сложные методы не дают существенного прироста качества

Логистическая регрессия (94,1%):

	Class
Variable	I
=====	
INF	-5.949
PRES	14.4693
PST	-10.5306
IMP	-11.0272
Intercept	3.0586

# Грамматический профиль глагола и его вид

Грамматический профиль позволяет  
довольно качественно определить вид  
глагола

**Можно применить протестирован-  
ные классификаторы и к глаголам,  
вид которых неизвестен =  
= к двувидовым глаголам**

# Двувидовые глаголы: состав

878 глаголов в «Грамматическом словаре» [Зализняк 2003]

*аблактировать — яровизировать*

Большинство глаголов редкие

304 глагола, от которых

4 рассматриваемых типа форм

суммарно встретились в

обследованном корпусе более 20 раз

# Проблема

Классификатор выдаёт какие-то результаты — но чему они соответствуют?

# 3. Эксперимент

## Лингвистический эксперимент

Спасибо вам за готовность поучаствовать в эксперименте! Он займёт у вас в общей сложности чуть меньше 5 минут. Пожалуйста, для начала заполните некоторые анкетные данные о себе. В случае, если вы не можете или не хотите отвечать на какие-либо из этих вопросов, оставьте соответствующие поля пустыми.

Пол:

Возраст:

Есть ли у вас лингвистическое образование?

Если да, укажите, какое именно:

Продолжить

Как известно, русские глаголы подразделяются на два вида — совершенный и несовершенный.

Глаголы совершенного вида отвечают на вопрос «что сделать?»: *прочитать, разделить*.

Глаголы несовершенного вида отвечают на вопрос «что делать?»: *рисовать, покупать*.

Сейчас вам будут предлагаться формы глаголов, и для каждой из них вы должны будете определять вид. Вы можете выбрать один из трёх ответов — «совершенный вид», «несовершенный вид» и «не знаю». Глаголы могут быть даны не только в начальной форме (например, приведённые выше глаголы могут попасться вам в формах *прочитай, разделят, рисовал, покупая*).

Ваша задача — за 3 минуты успеть определить вид как можно большего числа глаголов. За каждый правильный ответ вы получаете 1 балл, а за каждый неправильный ответ с вас снимается 1 балл; ответы «не знаю» не приносят вам баллов и не отнимают у вас баллы.

Когда 3 минуты истекнут, вам перестанут выдаваться глаголы, и мы попросим вас ответить ещё на несколько коротких вопросов.

Всё понятно. Начать эксперимент!

# гуляю

Совершенный вид

Несовершенный вид

Не знаю

# Дизайн эксперимента

40 случайных форм двувидовых глаголов  
240 случайных форм обычных глаголов

Сперва выдаются по крайней мере 6  
обычных глаголов, затем примерно  
каждый 10-й глагол двувидовой



## Младолингвисты (LingLing.ru)

Небольшой лингвистический эксперимент на LingLing'e.  
Помогите науке :)

<http://www.lingling.ru/exp/exp.php>

 Александр Пиперски

сегодня в 3:51

 2 Мне нравится  7

Скрыть комментарии



Мария Захарова

24 балла набрала, при том что 1 раз нечаянно не туда тыкнула((

сегодня в 7:17 | Ответить

 1



Варвара Поварова

63 балла

сегодня в 10:17 | Ответить



Мария Громова

50 баллов:)

сегодня в 12:57 | Ответить



Елизавета Лазарева

82 0\_0

сегодня в 13:58 | Ответить

# Хорошие люди

**— Как вы считаете, что изучалось в ходе эксперимента?**

— Умение людей правильно задавать вопросы к глаголам, различать виды глаголов.

**— Если у вас есть какие-то замечания по содержанию эксперимента, пожалуйста, выскажите их.**

— Все отлично, я бы все-таки убрал третий пункт) [«Не знаю» — А. П.]

# Плохие люди

— **Есть ли у вас лингвистическое образование?**

— Да.

— **Если да, укажите, какое именно.**

— ОТиПЛ; кфн

...

# Плохие люди

...

— Как вы считаете, что изучалось в ходе эксперимента?

— Восприятие двувидовых форм.

— Если у вас есть какие-то замечания по содержанию эксперимента, пожалуйста, выскажите их.

— Кажется, какой-то глагол повторялся. Не знаю, не смутит ли участников, что в случае двувидовых глаголов нужно давать "пораженческий" ответ "не знаю" — тогда как они, наоборот, вроде бы "знают" лучше.

# Результаты эксперимента

120 опрошенных, не понявших суть эксперимента

415 ответов про двувидовые глаголы

Существенных различий между формами не наблюдаются

Различия между отдельными глаголами велики

# Различия между формами

	СВ	НСВ
INF	44%	56%
INDIC,(PRAES FUT)	41%	59%
INDIC,PRAET	54%	46%
IMPER	44%	56%

# Различия между глаголами

	СВ	НСВ
<i>реабилитировали</i>	100%	0%
<i>опробовали</i>	93%	7%
<i>переквалифицируешь</i>	91%	9%
...		
<i>постулировать</i>	10%	90%
<i>пикировала</i>	9%	91%

# Результаты применения различных методов

	GER	OneRule	LogReg	Exp
НСВ	184 (75%)	175 (58%)	183 (60%)	15 (42%)
СВ	60 (25%)	129 (42%)	121 (40%)	21 (58%)
NA	60	0	0	268

# Эксперимент как образец

	GER	OneRule	LogReg
Верно	22 (81%)	23 (64%)	27 (75%)
Неверно	5 (19%)	13 (36%)	9 (25%)
NA	9	0	0

# Выводы (1)

Грамматические профили (частоты употребления различных форм) позволяют довольно надёжно определять вид одновидовых глаголов даже при помощи простого классификатора OneRule

# Выводы (2)

Если считать эксперимент с бесконтекстным определением вида надёжным источником информации о том, на что похож двувидовой глагол, то:

лучше всего ему соответствует классификатор по формам деепричастия

OneRule по частотности презенса справляется заметно хуже более сложных моделей

# Вопросы к самому себе

Является ли эксперимент истиной в последней инстанции?

Не ли проверить всё вышесказанное на исторических данных?

Какие другие модели можно применить?

**Спасибо за внимание!**

# Библиография (1)

Виноградов В. В., Истрина Е. С., Бархударов С. Г. (ред.). 1952–1954. *Грамматика русского языка*. В 2-х тт. Москва: Издательство АН СССР.

Влахов А. В.. 2010. *Причастия будущего времени в русском языке*. Санкт-Петербург: СПбГУ. Выпускная квалификационная работа бакалавра филологии.

Зализняк А. А. 2003. *Грамматический словарь русского языка*. 4-е изд. Москва: Русские словари.

Зализняк Анна А., Шмелёв А. Д. 1997. *Лекции по русской аспектологии*. München: Otto Sagner.

Крапивина К. А. 2009. *Причастный таксис в русском языке*. Санкт-Петербург: СПбГУ. Выпускная квалификационная работа бакалавра филологии.

# Библиография (2)

- Ляшевская О. Н., Шаров С. А. 2009. *Частотный словарь современного русского языка*. Москва: Азбуковник.
- Мучник И. П. 1966. Развитие системы двувидовых глаголов в современном русском языке. *Вопросы языкознания* 1. 61–75.
- Петрухина Е. В. 2009. *Русский глагол: категории вида и времени*. Москва: МАКС Пресс.
- Holte, Robert C. 1993. Very simple classification rules perform well on most commonly used datasets. *Machine Learning* 11(1). 63–90.
- Janda, Laura A. & Olga Lyashevskaya. 2011. Grammatical profiles and the interaction of the lexicon with aspect, tense, and mood in Russian. *Cognitive Linguistics* 22(4). 719–763.

# Библиография (3)

Jászay, Laszlo. 1999. Видовые корреляты при двувидовых глаголах. *Studia Russica XVII*. 169–177.

Korba, John Joseph. 2007. *The development of overt aspectual marking among Russian biaspectual verbs*. Chapel Hill: The University of North Carolina at Chapel Hill.

Kuznetsova, Julia. 2015. *Linguistic profiles: Going from form to meaning via statistics*. Berlin: de Gruyter Mouton.

Lyashevskaya, Olga. 2015. *Corpus-based profiles of Russian nouns: from grammatical number to lexical semantics*. (Working Papers by NRU HSE. Series WP BRP “Linguistics”).