

UiT

THE ARCTIC
UNIVERSITY
OF NORWAY

Лингвистические профили: квантитативный подход к теоретическим вопросам

Лора А. Янда

CLEAR (Когнитивная Лингвистика:
Эмпирические Подходы к Русскому Языку)

UiT (Арктический Университет Норвегии)



Традиционная проблема

Квантитативные исследования в теоретической лингвистике

Но:

В теориях не заложены квантитативные методы исследования.

Квантитативный анализ не обязательно релевантен для теории.

С точки зрения Больших Вопросов

Большие Вопросы

*Не ограничиваются рамками теории
Интересуют всех лингвистов*

Теория

*Помогает задать направление
Большим Вопросам*

Операционализация

*Способствует применению
квантитативных методов*

Общий обзор

1. Большие Вопросы
2. Теоретическая перспектива
3. Операционализация
4. **Примеры**
5. Заключение

1. Некоторые Большие Вопросы

- Каково соотношение между формой и значением?
- Как соотносятся лексика и грамматика?

2. Теоретическая перспектива: Когнитивная лингвистика

Минимальное предположение: язык можно описать в терминах общих когнитивных стратегий

- язык не является автономной способностью
- нет строгого разделения между грамматикой и лексикой
- априори не существуют универсалии

За основу берутся реальные употребления: обобщения вытекают из лингвистических данных

- нет строгого разделения между языком и речью
- нет глубинных структур

Значение является основой для всех лингвистических явлений

- нет семантически пустых форм
- различия в поведении мотивированы различиями в значении

Большие Вопросы, сформулированные Когнитивной Лингвистикой

Каково соотношение между формой и значением?

Как форма отражает значение и наоборот?

Можем ли мы использовать различия в форме для оценки значения?

Как соотносятся лексика и грамматика?

Как мы оцениваем значение в грамматике?

Можем ли мы использовать подобные модели для грамматических значений?

3. Операционализация: Лингвистические профили

Разновидности поведенческих профилей (Firth 1957, Harris 1970, Hanks 1996, Divjak & Gries 2006, Gries & Divjak 2009)

Грамматические профили: соотношение между частотным распределением форм и лингвистическими категориями

Семантические профили: соотношение между значениями (согласно семантической разметке) и формами

Конструкционные профили: соотношение между частотным распределением грамматических конструкций и значением

4. Примеры

- Грамматические профили: Время, вид, наклонение в русском языке
- Семантические профили: «пустые» приставки в русском языке
- Конструкционные профили: SADNESS в русском языке

Для каждого примера мы обозначим:

- Большие Вопросы
- Теоретические перспективы
- Операционализацию (Профили) и статистические методы

Грамматические профили: Время, вид, наклонение в русском языке

Janda, L. A. & Lyashevskaya, O. 2011.

“Grammatical profiles and the interaction of the lexicon with aspect, tense and mood in Russian”.

Cognitive Linguistics 22:4 (2011), 719-763.



Грамматические профили: Время, вид, наклонение в русском языке

Большие Вопросы:

Каково соотношение между формой и значением?

→ *между глагольной формой и грамматическим значением вида?*

Каково соотношение между лексикой и грамматикой?

→ *между лексическим значением глаголов и временем, видом и наклонением?*

Грамматические профили: Время, вид, наклонение в русском языке

Теоретическая перспектива:

Можем ли мы оценить выражение вида, опираясь на распределение глагольных форм?

Можем ли мы проследить различие между префиксацией и суффиксацией в качестве показателей вида?

Грамматические профили: Время, вид, наклонение в русском языке

Операционализация:

Грамматические профили: частотное распределение словоформ

- Распределение форм русского глагола по субпарадигмам
- Распределение русских глаголов по субпарадигмам

Данные:

Приблизительно 6 млн. глагольных форм из Национального корпуса русского языка (<http://ruscorpora.ru/>)

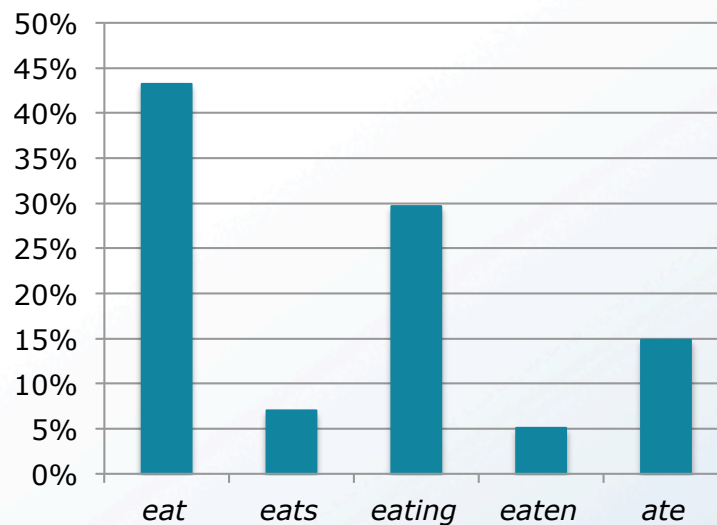
Статистика:

Хи-квадрат, V Крамера, графики распределения

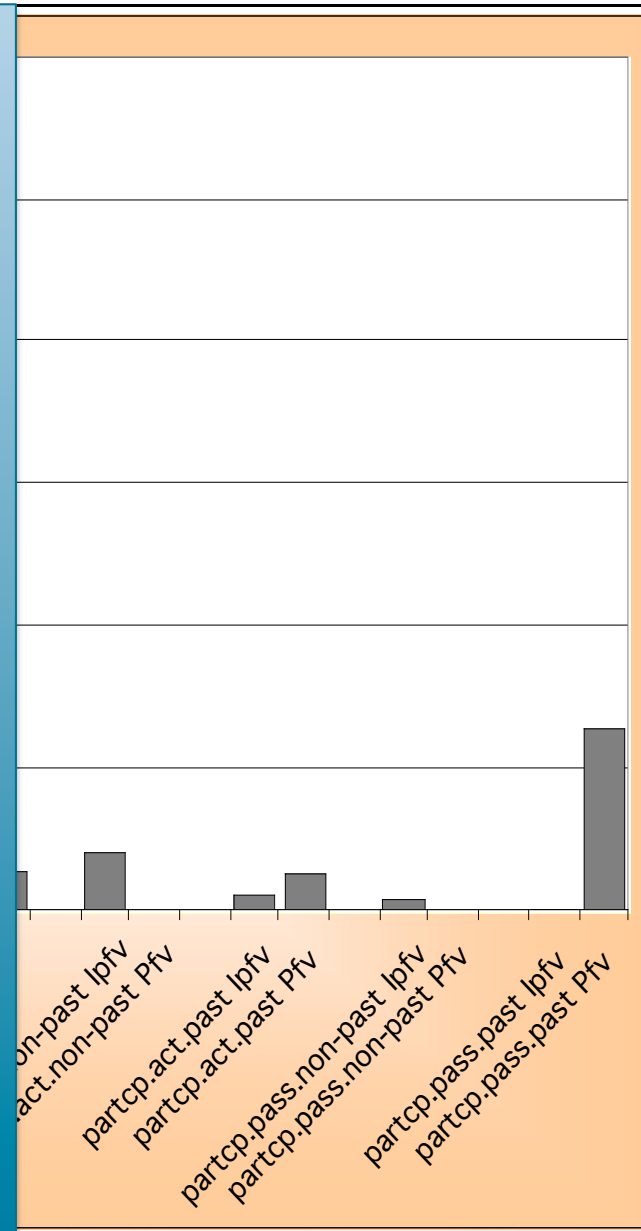
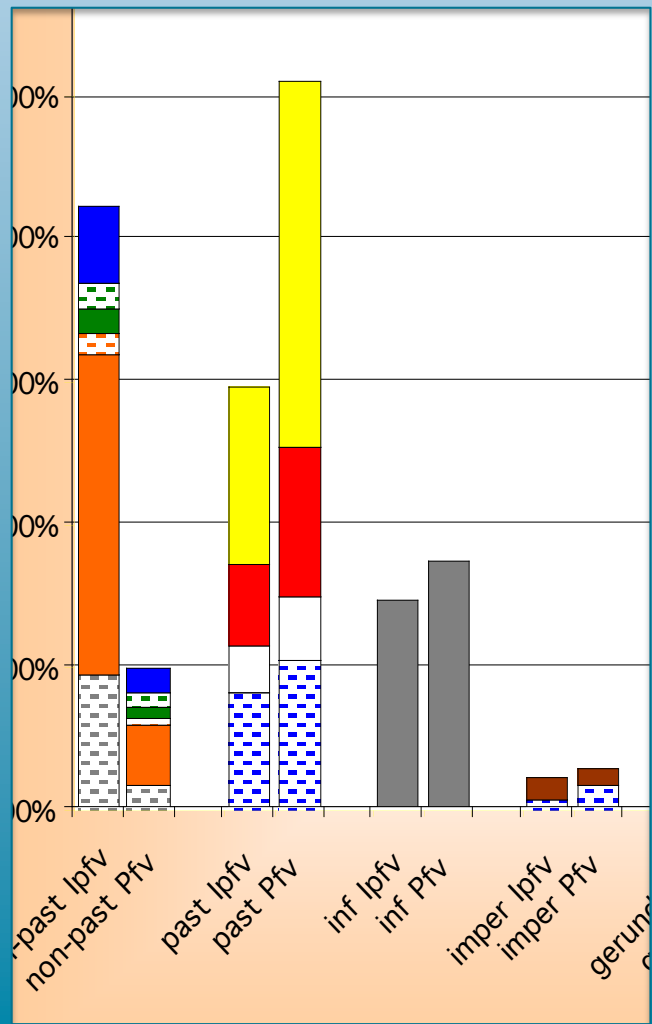
Что такое грамматический профиль?

Глаголы имеют разные формы:

<i>eat</i>	749 млн.
<i>eats</i>	121 млн.
<i>eating</i>	514 млн.
<i>eaten</i>	88.8 млн.
<i>ate</i>	258 млн.



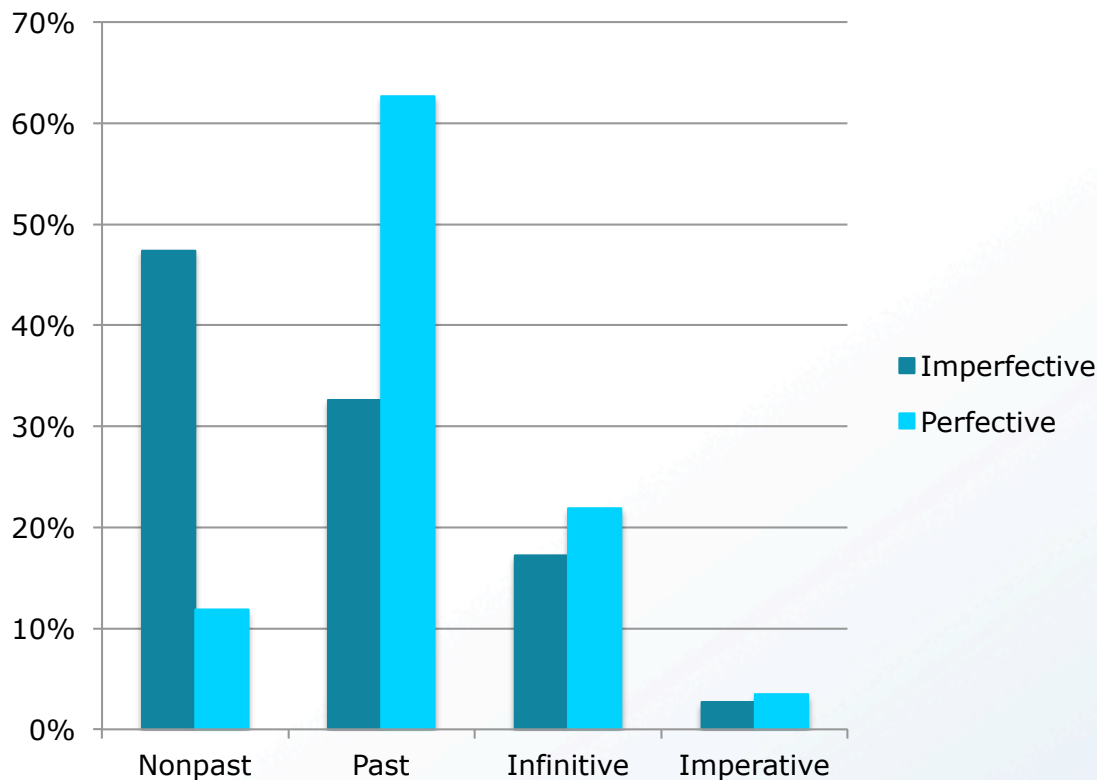
Грамматический
профиль глагола
eat



	lpfv	Pfv
non-past:	42.42%	9.83%
1 sg	5.58%	1.58%
1 pl	1.75%	1.05%
2 sg	1.71%	0.87%
2 pl	1.48%	0.38%
3 sg	22.50%	4.35%
3 pl	9.40%	1.60%
past:	29.51%	51.16%
sg.m	12.46%	25.75%
sg.f	5.63%	10.39%
sg.n	3.17%	4.72%
pl	8.24%	10.31%
inf	14.60%	17.50%
imper:	2.25%	2.96%
sg	1.45%	1.35%
pl	0.80%	1.61%
gerund	4.85%	3.03%
participles:		
act.non-	4.28%	0.00%
act.past	1.06%	2.59%
pass.nor	0.93%	0.00%
pass.pas	0.10%	12.93%

Грамматические профили русских глаголов

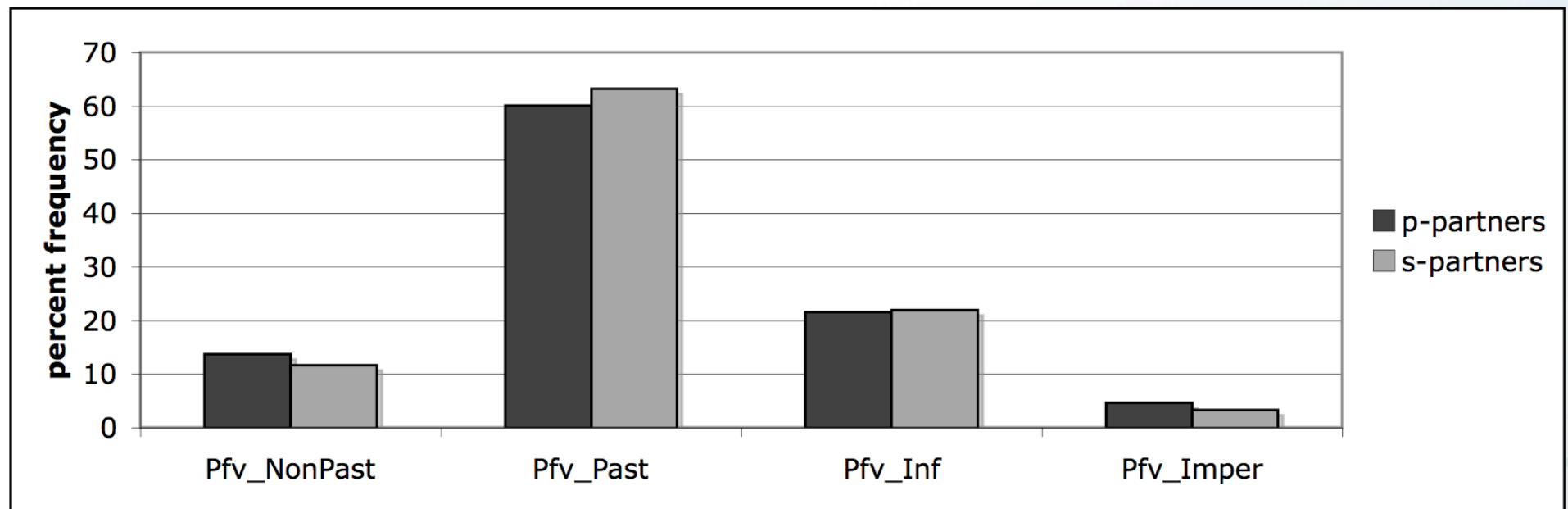
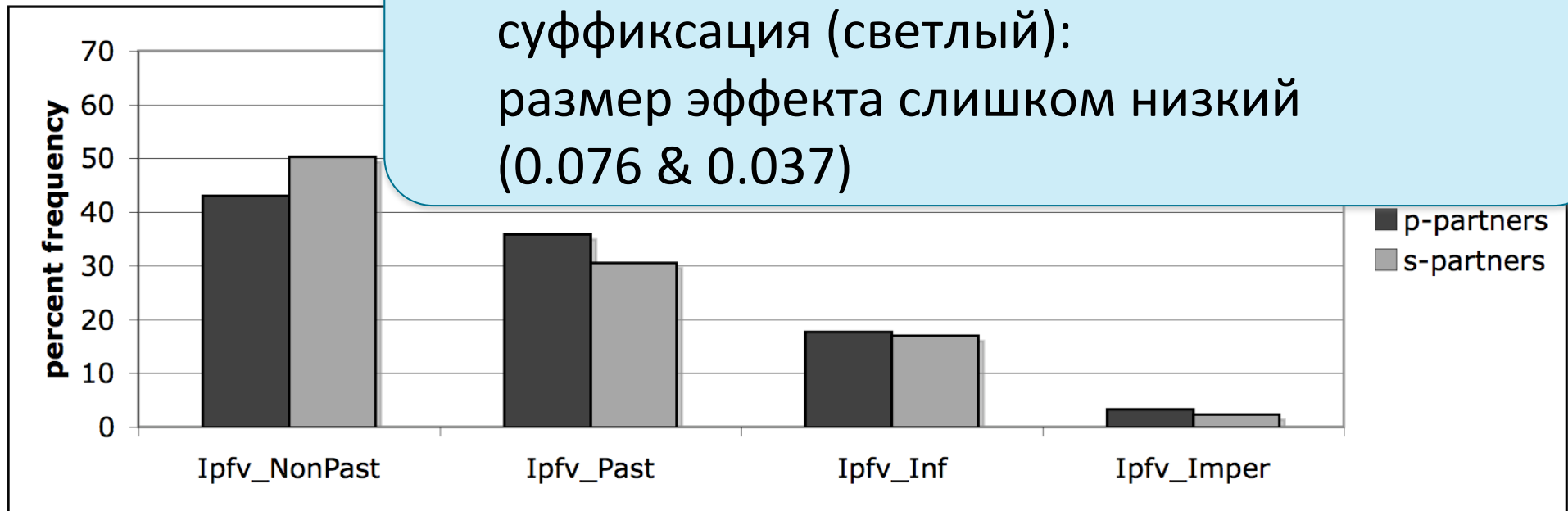
	Непрошедшее	Прошедшее	Инфинитив	Повелительное наклонение
Несовершенный вид	1,330,016	915,374	482,860	75,717
Совершенный вид	375,170	1,972,287	688,317	111,509



хи-квадрат
 = 947756
 df = 3 (степени
 свободы)
 р-значение < 2.2e-16
 V Крамера
 (размер эффекта)
 = 0.399
 (средний-большой)

Дистрибуция р

Префиксация (темный цвет) vs. суффиксация (светлый):
размер эффекта слишком низкий
(0.076 & 0.037)



Распределение русских глаголов по субпарадигмам: Глаголы несовершенного вида и их аттракция к императиву



Proportion of Imperfective Imper

Императив несовершенного вида

“делайте X!”

- **Вежливость: гость знает, чего ожидать:** *раздевайтесь, садитесь*
- **Настойчивость: слушатель находится в нерешительности:** *ступайте, глядите, забирайте*
- **Настойчивость: слушатель вел себя недолжным образом:** *проваливай, кончай, не перебивай*
- **Вежливые просьбы:** *выручайте*
- **Добрые пожелания:** *выздоровливайте*
- **Идиомы:** *давайте посмотрим, прощайте, соединяйтесь (слоган), запевай (в армии)*

Грамматические профили: Результаты

- Глаголы совершенного вида ведут себя иначе, чем глаголы несовершенного вида
- “Глагольные пары” ведут себя одинаково независимо от того, какая морфема (префикс или суффикс) используется для указания на вид
- Мы можем точно идентифицировать глаголы, которые тяготеют к определенным сочетаниям время-вид-наклонение
- Педагогика: Мы можем разработать материалы, стратегически сосредоточенные на релевантные глаголы и сочетания время-вид-наклонение

Семантические профили: «Пустые» приставки в русском языке

Janda, L. A. & Lyashevskaya, O. 2013. "Semantic Profiles of Five Russian Prefixes: *po-*, *s-*, *za-*, *na-*, *pro-*". *Journal of Slavic Linguistics* 21:2, 211-258.



Семантические профили: «Пустые» приставки в русском языке

Большие Вопросы:

Каково соотношение между формой и значением?

→ ...между префиксами и значением глаголов?

Существуют ли «пустые» формы?

→ Так ли «пусты» приставки, как принято считать?

Основа несоверш. вида	Ф-ма соверш. вида с префиксом
<i>советовать</i>	<i>посоветовать</i>
<i>варить</i>	<i>сварить</i>
<i>писать</i>	<i>написать</i>
<i>твердеть</i>	<i>затвердеть</i>
<i>греметь</i>	<i>прогреметь</i>

Семантические профили: «Пустые» приставки в русском языке

Теоретическая перспектива:

Можем ли мы оценить соотношение между приставками и значениями глаголов?

→ *Распределение приставок vs. семантические группы глаголов*

Как мы можем показать, что «пустые» формы на самом деле не пусты?

→ *Показать, что приставки демонстрируют различное семантическое поведение*

Семантические профили: «Пустые» приставки в русском языке

Операционализация:

Семантические профили: соотношение между значениями (согласно семантической разметке) и формами

→ Распределение русских глагольных приставок vs. семантическая разметка

Данные:

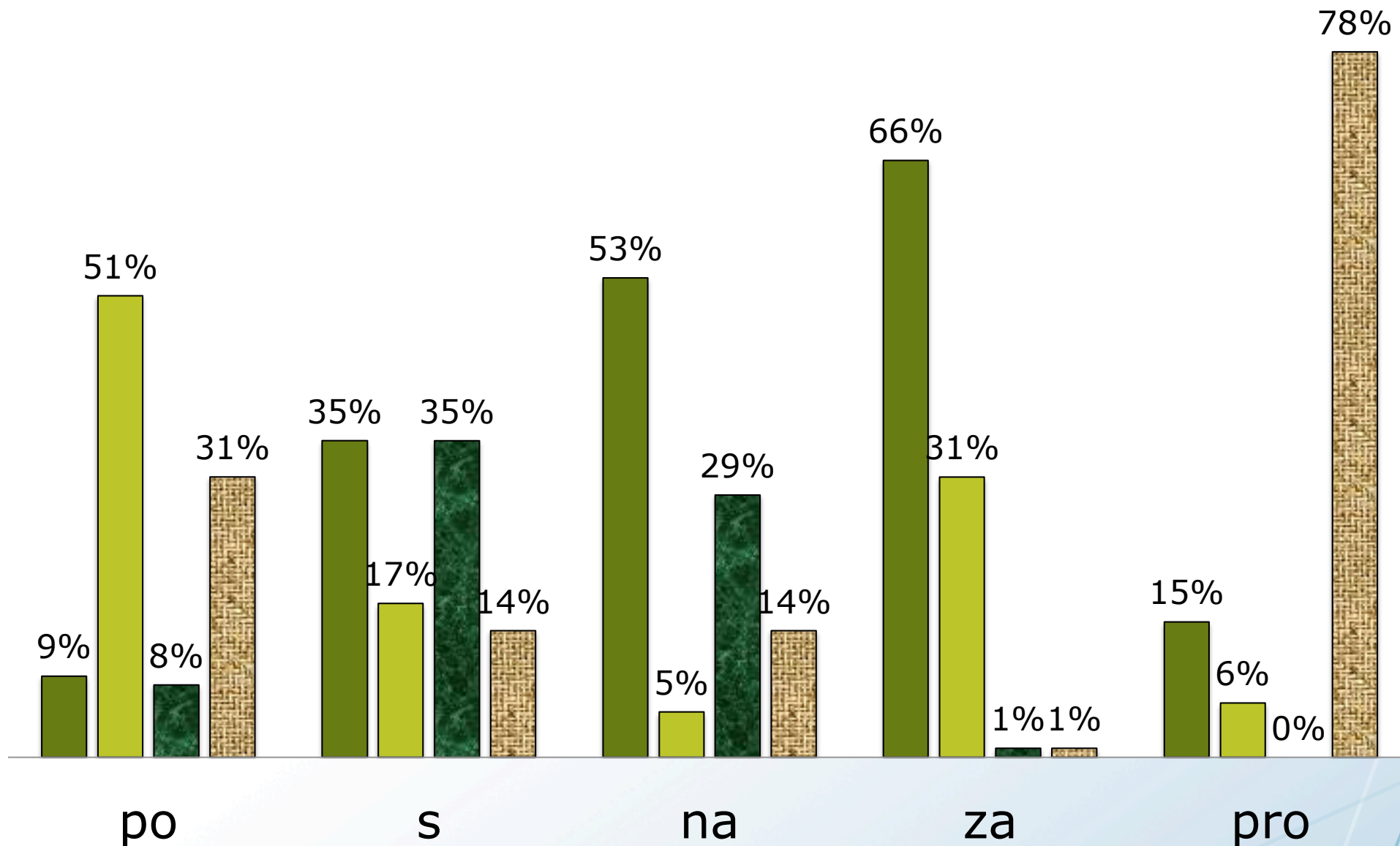
382 глагола с «пустыми» приставками из базы данных Exploring Emptiness (<http://emptyprefixes.uit.no/index.php>), независимо приписанная семантическая разметка в Национальном корпусе русского языка (<http://ruscorpora.ru/>)

Статистика:

Хи-квадрат, V Крамера, тест Фишера

■ IMPACT ■ CHANGEST ■ BEHAV ■ S&S

хи-квадрат = 248, df = 12, p = 2.2e-16; V Крамера = 0.8



Семантические профили: Результаты

- Каждая приставка обладает уникальным семантическим профилем
- Каждая приставка демонстрирует аттракцию и отторжение к различному набору семантических классов глаголов
- Возможно установить значения приставок и ожидания того, как приставки комбинируются с глаголами
- Педагогика: Мы можем разработать материалы, облегчающие процесс запоминания ≈ 2000 верных комбинаций приставка+глагол

Конструкционные профили: SADNESS в русском языке

Janda, L. A. & Solovyev, V. 2009. “What Constructional Profiles Reveal About Synonymy: A Case Study of Russian Words for sadness and happiness”. *Cognitive Linguistics* 20:2, 367-393.



Ускоренный курс по теме «падеж & SADNESS»



- Все падежи могут употребляться с предлогом
- Все падежи, кроме предложного, могут употребляться без предлога
- 70 конструкций [(предлог) [СУЩЕСТВИТЕЛЬНОЕ]_{падеж}]

SADNESS:

6 близких синонимов (нет обобщающего понятия)

грусть, меланхолия, печаль, тоска, уныние, хандра

Конструкционные профили: SADNESS в русском языке

Большие Вопросы:

Каково соотношение между формой и значением?

→ *Каково соотношение между синонимами и грамматическими конструкциями?*

→ *Каково соотношение между синонимами?*

Конструкционные профили: SADNESS в русском языке

Теоретическая перспектива:

Можем ли мы оценить различие между синонимами, опираясь на распределение в грамматических конструкциях?

Конструкционные профили: SADNESS в русском языке

Операционализация:

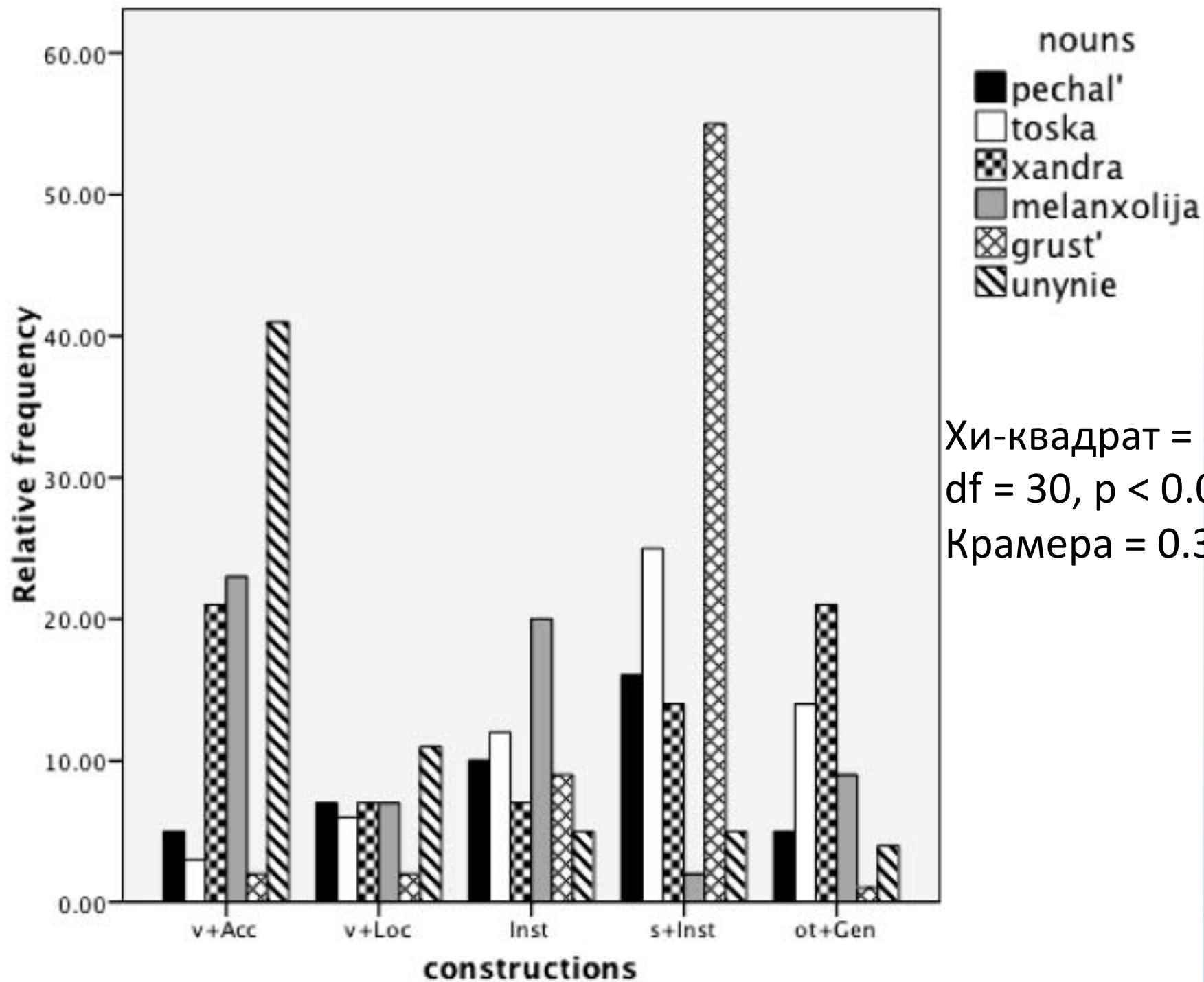
Конструкционный профиль: соотношение между частотным распределением грамматических конструкций и значением

→ Слова семантического поля SADNESS vs. распределение в конструкциях типа [(предлог) [СУЩЕСТВИТЕЛЬНОЕ]_{падеж}]

Данные: 500 предложений на каждое слово из Национального корпуса русского языка и Библиотеки Максима Мошкова

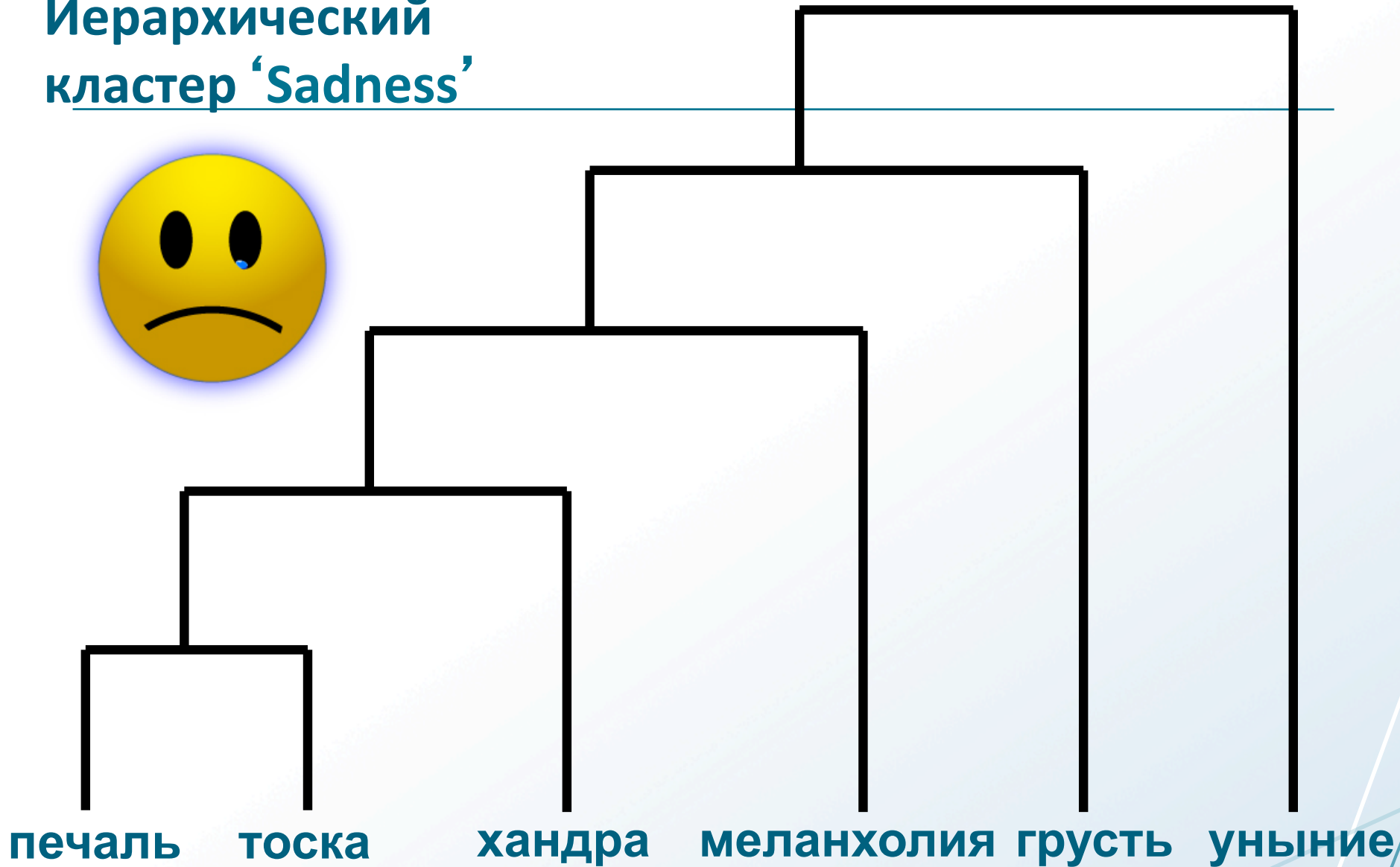
Статистика:

Хи-квадрат, V Крамера, Иерархическая кластеризация (квадрат Евклидовой метрики)



Хи-квадрат = 730.35,
df = 30, $p < 0.0001$, V
Крамера = 0.305

Иерархический кластер 'Sadness'



Конструкционные профили:

Результаты

- Каждый синоним имеет уникальный конструкционный профиль
- Одни синонимы ближе по значению, другие дальше
- Педагогика: Обучение релевантным конструкциям для каждого из синонимов
- Машинный перевод: Выбор лексики с учетом конструкционных профилей

Заключение

- **Лингвистические профили** дают нам возможность:
 - соединить теоретическую лингвистику с квантитативными исследованиями
 - определить операционализацию для применения квантитативных методов
 - ответить на Большие Вопросы, и одновременно получить полезные результаты

Заключение

- **Большие Вопросы:**
 - Каково соотношение между формой и значением? *Сложное.*
 - Как соотносятся лексика и грамматика? *Тесно.*
- **Результаты:**
 - Глаголы совершенного вида ведут себя по-другому, чем глаголы несовершенного вида, независимо от морфологии.
 - Мы можем определить, какие глаголы тяготеют к каким ячейкам в парадигме, и какие существительные тяготеют к каким конструкциям.
 - Пустых приставок нет; у каждой приставки уникальный семантический профиль.
 - Мы можем разработать материалы для стратегического обучения виду и синонимии.