Родовая атрибуция несклоняемых неодушевленных существительных

Попытка осмысления

Илья Бызов, Виктория Васюкова, Никита Логин, Константин Филатов

Введение:

Русский язык постоянно пополняется новыми несклоняемыми существительными из других языков, и словари не всегда успевают фиксировать для них норму относительно родовой принадлежности.

Однако и словарная норма зачастую базируется не на объективных данных (e.g. корпусных исследованиях, опросах), а на интуиции составителя.

Методом интроспекции мы выяснили, что далеко не для всех лексем родовая принадлежность очевидна наблюдателю и что в отношении некоторых из них может наблюдаться (в основном, бинарная) вариативность. Для того, чтобы выяснить, что всё-таки определяет атрибуцию данных существительных и какие варианты атрибуции являются наиболее распространёнными, мы и решили обратиться к данному исследованию.

Основные понятия:

Атрибут - в синтаксисе второстепенный член предложения, обозначающий признак, качество, свойство предмета (Wikipedia). В русском языке обычно выражается прилагательным, прилагательным-местоимением или причастием.

Родовая атрибуция - приписывание существительному атрибута определенного грамматического рода (операциональное определение) Так как мы брали несклоняемые слова, то нам был важен именно род как синтаксический класс, а не склонение как класс словоизменительный.

Фонетический облик - звуковое выражение лексемы. Важен в данном исследовании, так как в русском языке для многих исконных слов существует очевидная связь между фонетическим обликом и родовой принадлежностью (последняя фонема номинатива: согласный для мужского рода, “-а” для женского и “-о/е” для среднего)

Обзор литературы

Зализняк А. А. "Русское именное словоизменение" с приложением избранных работ по современному русскому языку и общему языкознанию, 2002

Савчук С.О. “Корпусное исследование вариантов родовой принадлежности имен существительных в русском языке”, 2011

Атанасова И. “Несклоняемые существительные в современном русском языке”, 2013

Проблема. Цель. Задачи

Проблема: Недостаточная определенность существующих моделей родовой атрибуции несклоняемых неодушевленных существительных.

Цель: Уточнить имеющиеся модели родовой атрибуции несклоняемых неодушевленных существительных.

Задача 1: Разработать анкету для получения данных.
Задача 2: Сравнить эффективность методов, использованных при проектировании анкеты.

Почему именно анкета ?

- Корпусное исследование скорее отражает норму, чем узус (исключение: корпуса, посвященные именно речи, а не языку, такие как Araneum). Данные из корпуса гораздо сложнее обрабатывать, так как в представленном там массиве предложений атрибуция может происходить в любом месте. Также корпусные данные могут быть уже устаревшими на момент проведения исследования. *(а еще корпусное исследование уже проводилось)*

- Интервьюирование - требует слишком много ресурсов, в том числе времени и, возможно, денег. Есть преимущество в возможности вызвать спонтанную реакцию респондента, но ресурсоемкость делает данный метод нецелесообразным.

Объект и предмет

Объект: Род имен существительных

Предмет: Вариативность родовой атрибуции несклоняемых неодушевленных существительных

Гипотезы

Гипотеза 1: Атрибуция названий брендов зависит от написания. При написании кириллицей мозг человека воспринимает эти слова как русские, поэтому атрибутирует их по русскому склонению. При написании латиницей скорее ожидаема атрибуция по гипероним у.

Гипотеза 2: Атрибуция зависит от образованности и принадлежности к определенной культуре. Более образованный человек скорее выберет вариант атрибуции по гипероним у, тогда как менее образованный скорее будет исходить из русских типов склонения.

Гипотеза 3: Предположение об устройстве атрибутивного механизма. Предположительно, слово имеет несколько признаков, и вероятно часть из них представляет собой иерархически устроенную систему. Например, в качестве факторов атрибуции могут выступать:

- явное наличие “гиперонима”
- начальная форма слова
- состав семантических связей

Гипотеза 4: Характер атрибуции зависит от степени осознаваемости слова (например, как маркера образованности) или частотности его употребления.

Отбор слов

Отбор слов производился в соответствии с гипотезами:

Ауди – вариативность (ж//м) – гиперонимы: машина и автомобиль

Феррари – вариативность (ж//м) – гиперонимы: машина и автомобиль

Плейстейшн (PlayStation) – (ж//м) – разнородные семантические связи vs фонетический облик

Жиллет (Gillette) – (ж//м) – разнородные фонетические связи vs фонетический облик

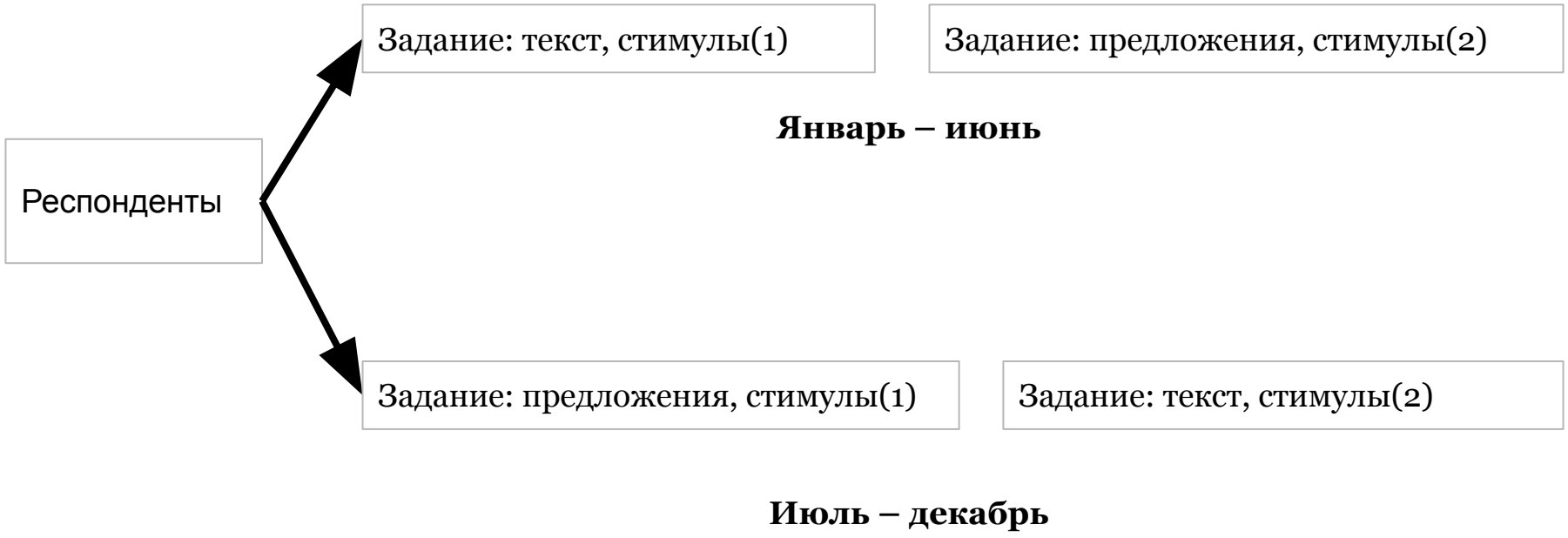
Евро – (м//с) – аналогия со словами рубль, доллар vs фонетический облик, высокая частотность, осознаваемость

Песо – (м//с) – аналогия со словами рубль, доллар vs фонетический облик, низкая частотность, осознаваемость

Проектирование анкеты

Стимулы (1)	Стимулы (2)
Ауди + машина	**Ауди + автомобиль**
Феррари + автомобиль	**Феррари + машина**
Playstation	**Gillette**
Жиллет	**Плейстейшн**
Евро	**Песо**

Проектирование анкеты



Зачем нам такая объемная анкета?

Родовая атрибуция – многофакторное явление и вариативность зависит от большего чем один количества факторов.

Потенциально невозможно изолировать один фактор от других в контексте.

Мы вынуждены помещать каждый фактор в каждый контекст, чтобы контролировать атрибуцию.

Почему нет “нейтрального” варианта?

Нейтральная атрибуция -- атрибуция в свободном контексте, на которую не оказывается влияние

Доподлинно неизвестно, сколько факторов влияют на атрибуцию

Оставив респондентам “естественную” возможность присвоить род, мы не смогли бы сделать никаких выводов касательно состава и характера факторов, повлиявших на атрибуцию

Добавление нейтрального варианта к уже существующим увеличило бы анкету текущего дизайна в 3 раза

Выборка

Малое количество времени

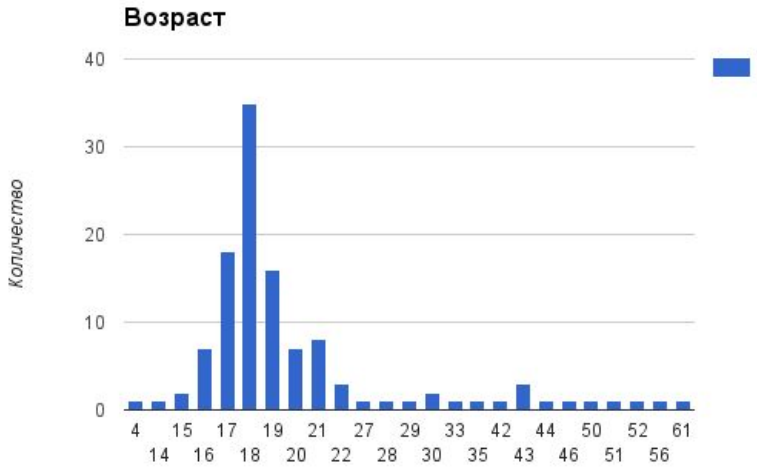
Отсутствие ресурсов

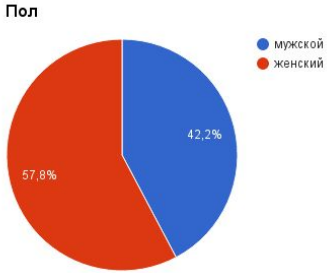
Было решено отказаться от рассмотрения

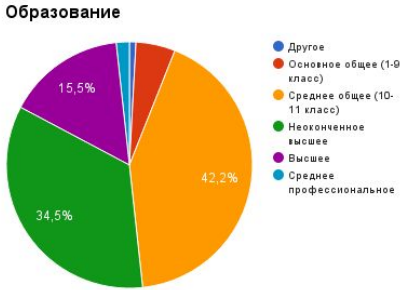
Гипотезы 2

Выборка – конформная

Выборка


Возраст
Количество
40
30
20
10
0
4 15 17 19 21 27 29 33 42 44 50 52 61
14 16 18 20 22 28 30 35 43 46 51 56


Пол
мужской
женский
42,2%
57,8%

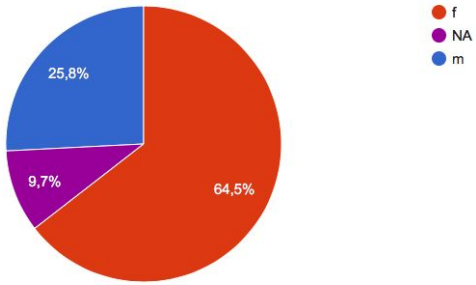

Образование
Другое
Основное общее (1-9 класс)
Среднее общее (10-11 класс)
Неоконченное высшее
Высшее
Среднее профессиональное
15,5%
42,2%
34,5%

Сложности:

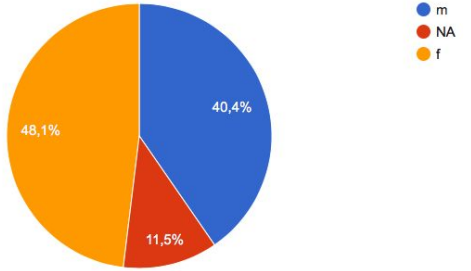
1) Пропуски в тексте
    a) много мусора
2) Составление предложений
    a) начальная форма - мужской род местоимений
    b) некоторые предложения слишком сложные и не даются респондентам
3) Многие респонденты жаловались на размер анкеты -> много незаполненных полей
4) Филлеры недостаточно хорошо маскируют предмет исследования (количество догадавшихся более 66%)
5) В ряде диагностических контекстов происходит нейтрализация вариантов

Результаты

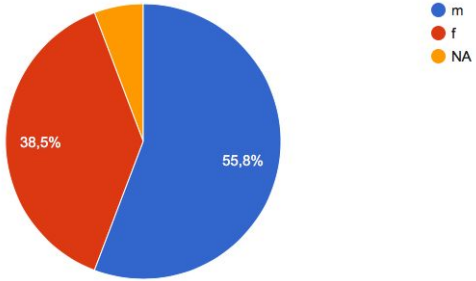
Ауди+машина(т)



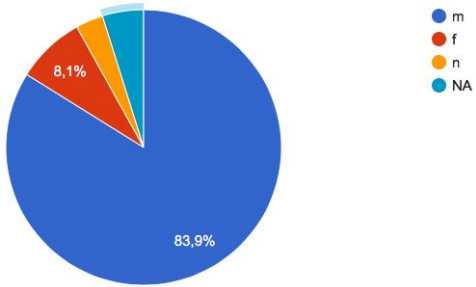
Ауди+авт(т)



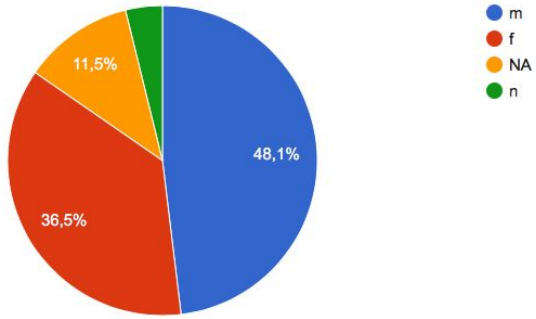
Ауди+машина(пр)



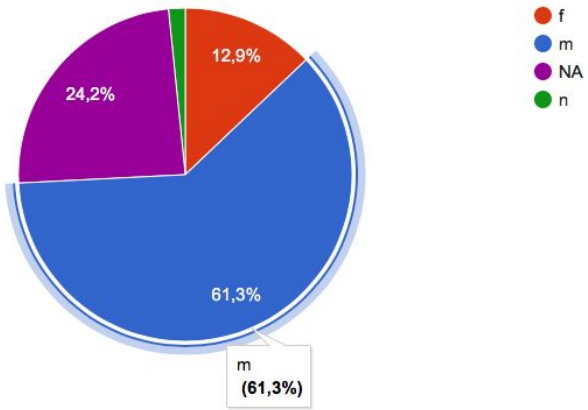
Ауди+авт(пр)



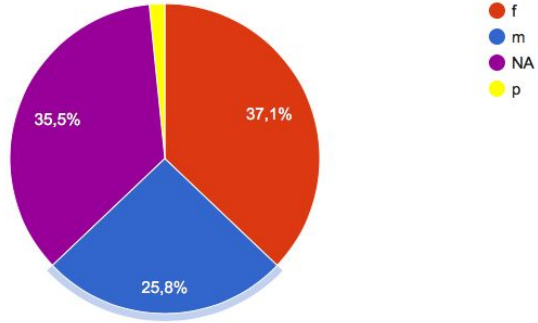
Феррари+маш(т)



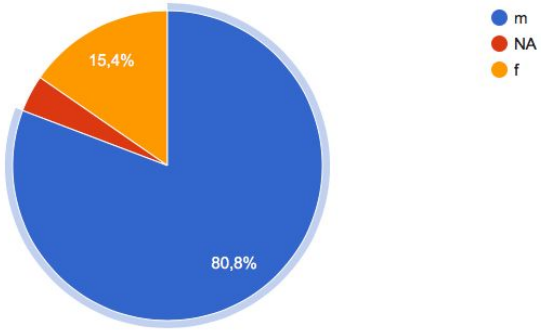
Феррари+авт(т)



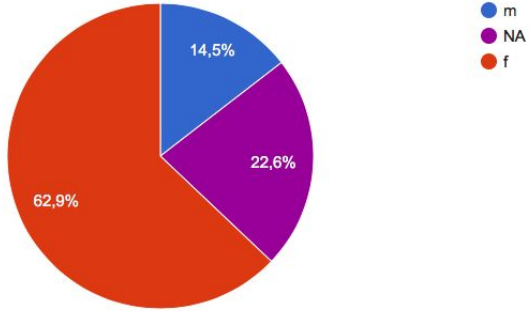
Феррари+маш(пр)



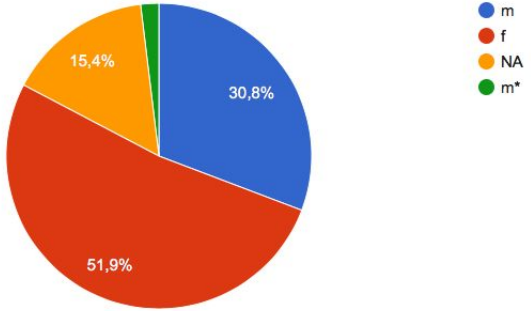
Феррари+авт(пр)



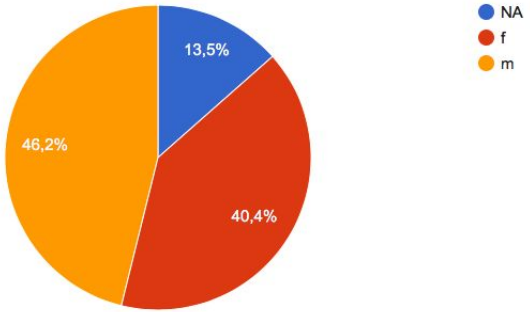
PlayStation(т)



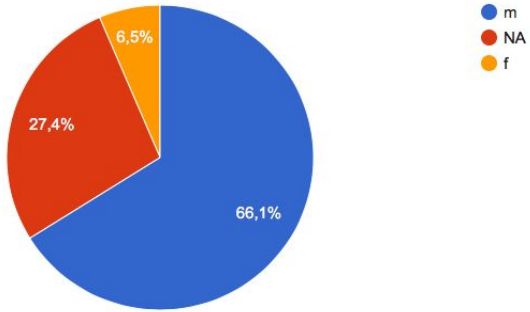
Плейстейшн(т)



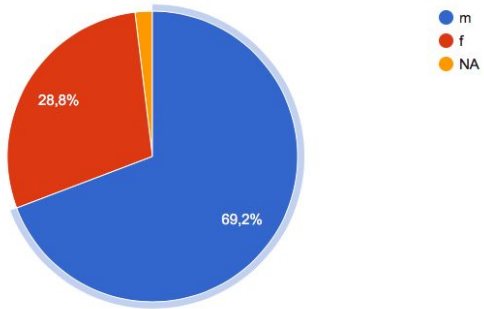
PlayStation(пр)



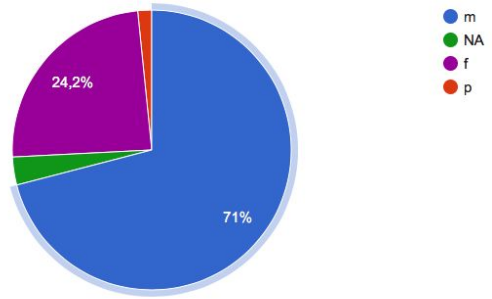
Плейстейшн(пр)



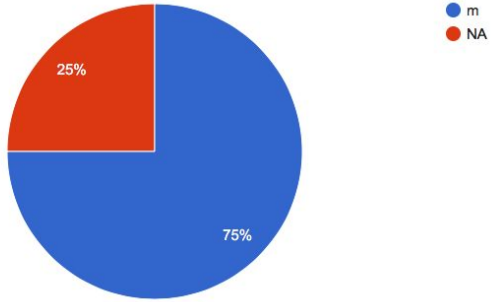
Gillette(т)



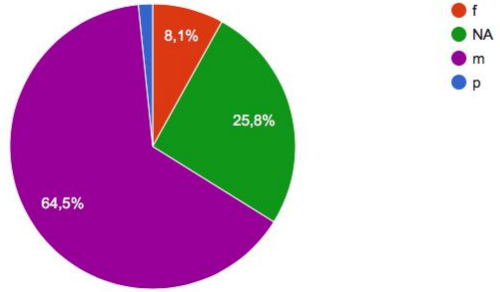
Жиллет(т)



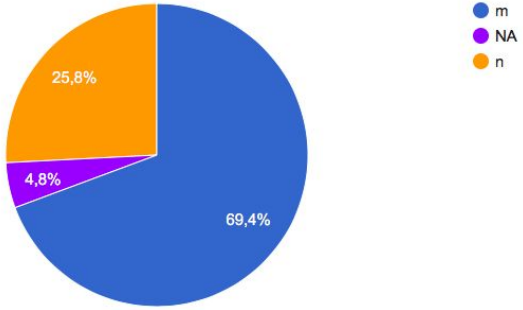
Жиллет(пр)



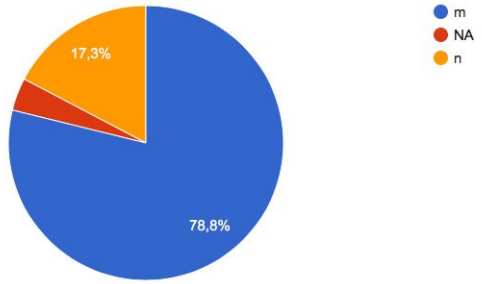
Gillette(пр)



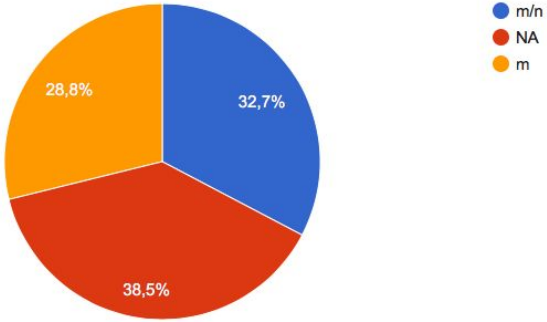
Песо(т)



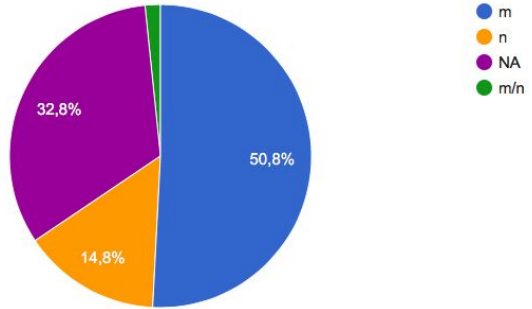
Евро(т)



Песо(пр)



Евро(пр)



Проблемы анализа данных

Из-за сложности выборки мы не смогли воспользоваться методами математической статистики

А значит, не можем говорить ничего относительно корреляций и статистической значимости полученных результатов

(1) **Латиница/Кирилица:**
1. PS vs. ПС

i) текст (PS) — предложение (ПС):
т: м — 7, ж — 29
п: м — 32, ж — 4
p-value = 0.00000001

ii) текст (ПС) — предложение (PS):
т: м — 16, ж — 22
п: м — 20, ж — 18
p-value = 0.5

2. Gillette vs. Жиллет

i) текст (Жиллет) — предложение (Gillette):
т: м — 34, ж — 10
п: м — 39, ж — 5
p-value = 0.3

ii) текст (Gillette) — предложение (Жиллет):
т: м — 28, ж — 11
п: м — 39, ж — 0
p-value = 0.001 (на это не смотреть, со значениями меньше 5 хи-квадрат не работает)

(2) **Гипероним в контексте:**

Как это можно интерпретировать?

- На атрибуцию можно влиять
- Атрибуция контекстно зависима
- Некоторые факторы действительно существуют и, похоже, что они работают так, как предсказывают гипотезы
- Некоторых факторов, возможно, не существует

Дальнейшие гипотезы

Гипотеза 1: Слова со схожими факторами атрибуции, ведут себя в разных контекстах аналогично

Гипотеза 3: (Остается)

Что еще можно сделать? Теория.

- ➔ Разработать методику оценки эффективности контекста
- ➔ Разработать методику выявления факторов атрибуции

Что еще можно сделать? Практика.

- ➔ Сделать выборку вероятностной/систематической
- ➔ Возможно, увеличить выборку
- ➔ Учесть такие переменные, как возраст/образование
- ➔ Перемешивать контексты внутри каждого задания, чтобы получилось более ровное распределение не ответивших
- ➔ Сделать предложения, в которых мы проверяем гипотезу про написание, одинаковыми
- ➔ Возможно, рассматривать только тех, кто сделал все задания, а потом сравнять количество людей в группах

Спасибо за внимание!