

STIMULSTAT: База данных,
охватывающая различные
психолингвистические
характеристики слов русского
языка

С. В. Алексеева*, Н. А. Слюсарь*¹, Д. А. Чернова*

*СПбГУ, Санкт-Петербург; ¹НИУ ВШЭ, Москва

Психолингвистика

- *Как люди понимают и обрабатывают языковые конструкции при **восприятии устной речи и при чтении**?*
- Решение: провести эксперимент (ай-треккер, энцефалограф, поведенческий эксперимент с измерением времени реакций и др.)
- Проблема: подбор языковых единиц с контролем одних параметров и варьированием других

Пример

Исследованием сегментации устной речи

- Задача: найти двусложные частотные существительные, прилагательные и глаголы с ударением на второй слог, длиной от пяти до семи букв и начинающиеся на буквы *в, р, н, л*.
- => как упростить выбор языковых единиц по многим параметрам?

StimulStat: база данных и веб-интерфейс

- Более 50 тыс лемм
- 97 параметров
 - Визуальные (длина, частотность)
 - Фонетические (место ударения, слоговая структура)
 - Грамматические (часть речи, индекс Зализняка, род, число, падеж ...)
 - Семантика (кол-во значений)
 - Омонимы/Омографы
 - Близкие по написанию слова (*сетка – секта; ток - кот*)

Постер

- Какие параметры Вам нужны? Приходите, обсудим