

**Выявление семантической дистанции слов на основе данных ЭЭГ с помощью многослойного перцептрона**

**Шлычкова Елизавета Олеговна**

*Студент (бакалавр)*

Тюменский государственный университет, Институт социально-гуманитарных наук,

Тюмень, Россия

*E-mail: stud0000239916@study.utmn.ru*

Настоящая работа посвящена использованию многослойного перцептрона для предсказания семантической дистанции между словами, высчитанной через расстояние Левенштейна для закодированных определений слов, на основе данных электроэнцефалограммы, зарегистрированной у человека во время мысленного сравнения семантики представляемых слов (на примере слов испанского языка).

Алгоритмы декодирования внутренней речи, в том числе семантики слов, с помощью анализа электрической активности мозга играют большую роль в развитии систем нейропринтерфейсов. Одним из способов определения семантики лексических единиц является нахождение их семантической дистанции в рамках семантических полей, формируемых человеком на биологическом уровне. Последовательность этапов исследования представлена на рисунке 1.

В качестве основы был взят датасет, созданный Х. Кальво, Х. Л. Паредес и Х. Фигероа Назуно [Calvo, Paredes Paredes, Figueroa Nazuno 2018: электронный ресурс]. Однако для нашей работы мы переработали датасет, оставив только данные ЭЭГ при обработке слов во время третьего временного диапазона – после предъявления двух слов-стимулов.

М. А. Кронгауз выражает мысль о том, что «семантика языковых единиц описывается, как правило, не независимо, а в сопоставлении с другими языковыми единицами, как правило, близкими по значению» [Кронгауз 2005: 79]. Соответственно, мы можем представить каждое слово как набор основных сем, из которых оно состоит.

Всего участникам эксперимента предъявлялось 138 слов, каждое из которых мы сопоставили с определением для его эквивалента на английском языке. В качестве источника определений использовался Кембриджский испано-английский словарь [Cambridge Dictionary 2023: электронный ресурс]. Примеры определений для слов-стимулов представлены в таблице 2.

Из определений были исключены частотные слова [Wikipedia – Most common words in English 2023: электронный ресурс], а также редко встречающиеся (1-2 раза в рамках всех слов определений). В итоге каждому слову соответствовал набор из определенных лексических единиц.

Для нахождения семантической дистанции между словами мы закодировали наборы лексических единиц с помощью нулей и единиц, наподобие генетического метода (ноль – отсутствие данной единицы в наборе, единица – его наличие). Затем закодированные последовательности сравнивались между собой с помощью расстояния Левенштейна – полученные расстояния использовались в дальнейшем в качестве весов во взвешенной матрице, отражающей «координаты» каждого из слов, которые были представлены в виде графа (рисунок 3) с обозначениями на русском языке для удобства восприятия.

Чтобы выявить закономерности, мы обратились к многослойному перцептрону для предсказания значений семантической дистанции у пар слов на основе электрической активности мозга. Для получения точности предсказания были использованы метрики MSE и RMSE. В результате, MSE равнялось 47,607, а RMSE – 6,9. При этом минимальное значение дистанции равнялось 0, а максимальное – 38.

Полученные значения точности позволяют прийти к выводу, что многослойный перцептрон может использоваться для предсказания семантической дистанции между словами (высчитанной с помощью предлагаемого нами алгоритма) при анализе сигналов мозга. Однако в перспективе, для повышения результатов точности, возможно обращение к иным словарям для получения определений, а также к изменению весов при вычислении расстояния Левенштейна.

### Источники и литература

- 1) Кронгауз М.А. Семантика. / М.А. Кронгауз - М.: Издательский центр «Академия», 2005. 352 с.
- 2) Calvo H. Measuring Concept Semantic Relatedness through Common Spatial Pattern Feature Extraction on EEG Signals / H. Calvo, J. L. Paredes Paredes, J. Figueroa Nazuno // Mendeley Data. - 2018. URL: <https://data.mendeley.com/datasets/shzz5kbsgy/1> (дата обращения: 02.05.2023)
- 3) Cambridge Dictionary URL: <https://dictionary.cambridge.org/dictionary/spanish-english/> (дата обращения: 22.05.2023)
- 4) Wikipedia – Most common words in English URL: [https://en.wikipedia.org/wiki/Most\\_common\\_words\\_in\\_English](https://en.wikipedia.org/wiki/Most_common_words_in_English) (дата обращения: 22.05.2023)

### Иллюстрации

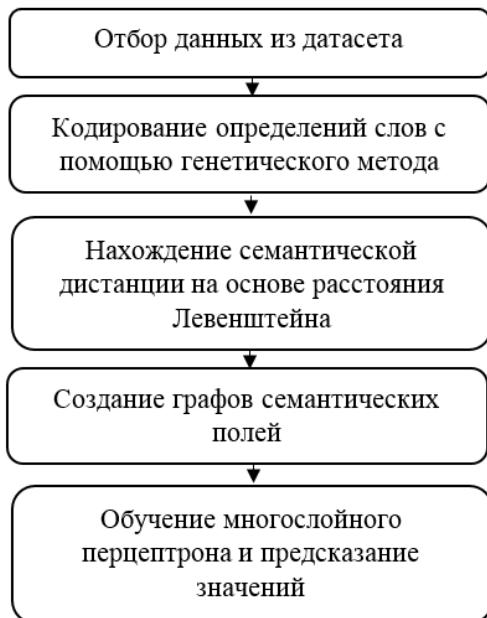


Рис. : 1. Этапы проведения исследования

Слово-стимул	Определение
escudo	a broad piece of metal, wood carried as a protection against weapons something or someone that protects a design which is used as the symbol of the town, family see also coat of arms, a family badge or crest
plátano	the long curved fruit, yellow-skinned when ripe, of a type of very large tropical tree a type of tree with broad leaves.
galleta	a crisp, sweet piece of dough baked in small flat cakes cookie a biscuit

Рис. : 2. Примеры определений слов-стимулов

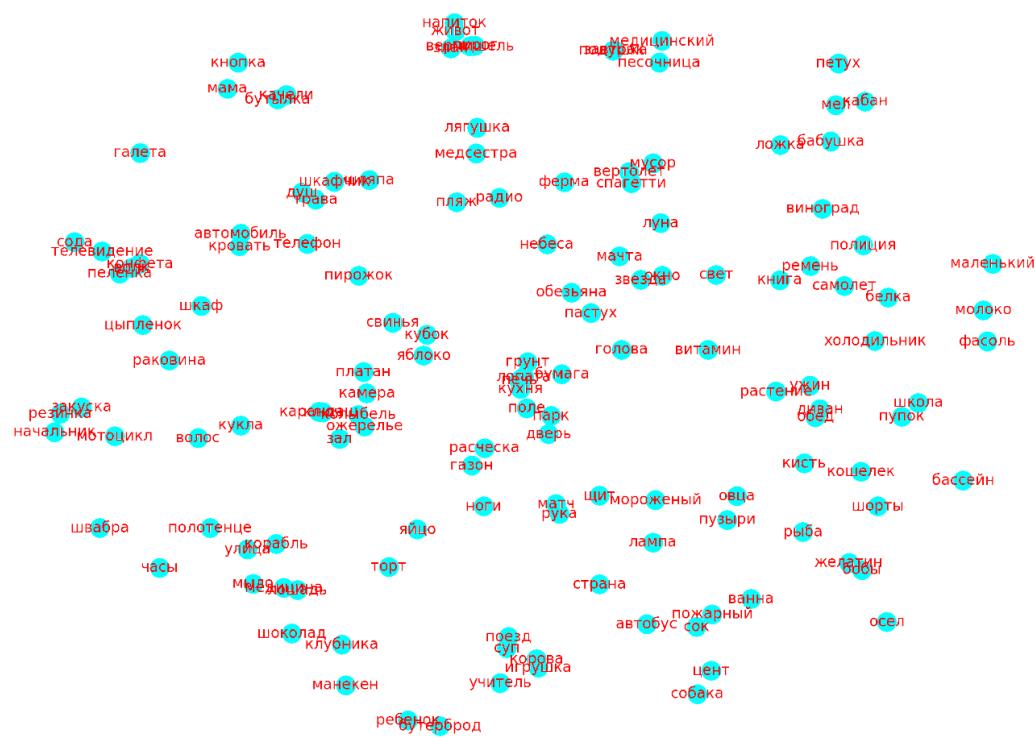


Рис. : 3. Граф семантических полей