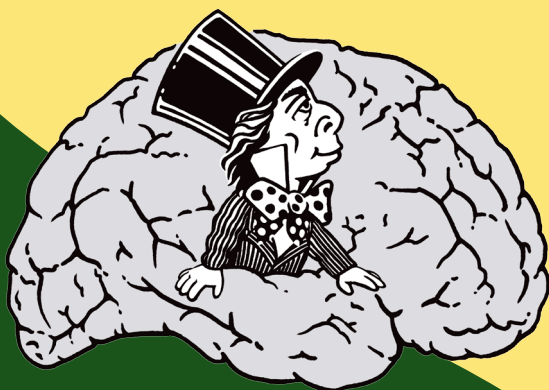


# КОГНИТИВНАЯ НАУКА

В МОСКВЕ



НОВЫЕ ИССЛЕДОВАНИЯ

МАТЕРИАЛЫ  
КОНФЕРЕНЦИИ  
2023

Под ред. Е.В. Печенковой, М.В. Фаликман, А.Я. Койфман

УДК 159.9  
ББК 88.25  
К57

Когнитивная наука в Москве: новые исследования. Материалы конференции 21 – 22 июня 2023 г. Под ред. Е.В. Печенковой, М.В. Фаликман, А.Я. Койфман. – М.: ООО «Буки Веди», Московский институт психоанализа. 2023 г. – 604 стр.

© Авторы статей, 2023

ISBN 978-5-4465-3880-5

УДК 159.9  
ББК 88.25

ISBN 978-5-4465-3880-5

© Авторы статей, 2023

## **ПОМОГАЕТ ЛИ ГРАФИЧЕСКОЕ ВЫДЕЛЕНИЕ И ВЕРСТКА ТЕКСТА ПОИСКУ ИНФОРМАЦИИ В СЛОВАРЕ: ИССЛЕДОВАНИЕ МЕТОДОМ АЙТРЕКИНГА**

А. Н. Пучкова\* (1, 2), А. А. Берлин Хенис (1), А. И. Ольховская (1)  
[puchkovaan@gmail.com](mailto:puchkovaan@gmail.com)

1 – Гос. ИРЯ им. А. С. Пушкина, Москва; 2 – ИВНД и НФ РАН, Москва

**Аннотация.** В современных словарях, особенно цифровых, используются различные методы лексикографической эмфазы – невербальных способов разметки и выделения определенных сведений в статье. При этом выбор конкретных способов не имеет экспериментально подтвержденных оснований. Метод айтрекинга дополняет традиционно используемые методы опроса и наблюдения и позволяет получить объективные и точные данные о параметрах поиска целевой информации в статье при различных вариантах ее разметки и оформления. В первом блоке эксперимента требовалось найти подходящие варианты употребления слова, которые в статье: а) выделяли курсивом, б) выделяли зеленым цветом шрифта, в) не выделяли. Во втором блоке нужно было найти верное значение многозначного слова, и подачу номерных значений давали: а) с простыми цифрами, б) с полужирными цифрами, в) с дополнительным разбиением на абзацы. Движения глаз при чтении регистрировали айтрекером EyeLink 1000+. Выделение цветом или курсивом не повлияло на скорость или правильность поиска, а фиксации были наиболее короткими на простом тексте. При поиске варианта значения слова оптимальной оказалась разбивка на абзацы, а выделение номеров полужирным провоцировало более быстрый, но менее правильный выбор значения.

**Ключевые слова:** словарь, чтение, поиск в тексте, выделение текста, айтрекинг, лексикография

Исследование поддержано государственным заданием FZNM-2020-0005.

### **Введение**

Лексикография как научное направление занимается составлением и изучением словарей и за последние десятилетия все более переносит свой фокус на пользователя словаря. «Очеловечивание» словарной продукции осуществляется благодаря подстройке содержательных и структурных параметров лексикографического произведения под параметры адресата (Каламбет, 2007), но такому осмыслению подвергались прежде всего содержательные компоненты словаря.

Однако помимо структуры и непосредственного содержания словаря, необходимо помнить и о его формальной организации – невербальных способах подачи словарной информации. Это стало особенно важно именно в наше время, когда все больше словарей переходят в цифровой формат, а экранный текст представляет значительно большие возможности модификации, чем пе-

чатный. К таким способам относятся: 1) шрифты (курсив, полужирный, разрядка, выделение путем изменения цвета); 2) принятые сокращения и условные знаки (пометы и символы); 3) выделение блока текста цветом (highlighting); 4) абзацные отступы; 5) иллюстрации и другие объекты мультимедиа; 6) расположение информации на листе или экране.

Совокупность перечисленных средств мы предлагаем называть лексикографической эмфазой на том основании, что все они выполняют функцию выделения определенных сведений. Лексикографическая эмфаза традиционно применяется в словарях для структурирования информации и навигации пользователя по тексту с целью облегчения задач поиска. Выбор в пользу конкретного невербального способа выделения текста (например, полужирным начертанием или цветом) осуществляется на основе опыта или вкуса составителя. Лексикографы традиционно опираются на анкетирование, интервьюирование и наблюдение для анализа стратегий словарного поиска (Atkins, Rudell, 2008). Исследования поиска внутри словарных статей с использованием айтрекинга немногочисленны, и не многие рассматривают невербальные компоненты текста, особенно в контексте работы со словарем на своем родном языке (например, Müller-Spitzer, 2014). Цель лексикографической эмфазы — помочь пользователю в поиске релевантной информации; ближайшими по содержанию можно считать исследования выделения гиперссылок (Gagl, 2016) или использования методов невербальной сигнализации для оптимизации обучения (Schneider et al., 2018).

Цель данного эксперимента состоит в применении метода айтрекинга для оценки возможности влияния на скорость и качество поиска информации в словарной статье с помощью манипуляции лексикографической эмфазой. Эта работа направлена на экспериментальную проверку потенциальной эффективности и наличия реальных различий между распространенными сейчас на практике методами лексикографической эмфазы. Для нашего пилотного исследования мы отобрали два варианта манипуляции текстом и две типовых задачи поиска информации внутри статьи. Было отобрано два параметра: 1) текстовые иллюстрации (речения); 2) подача полисемии (номерных значений многозначного слова) и задачи поиска примера употребления и подходящего значения слова.

## Методика

В исследовании приняли участие 45 студентов (41 женщина, 4 мужчин) бакалавриата и магистратуры Института русского языка им. А.С. Пушкина. Средний возраст составил 21.2 года (18 – 41 год). Информанты отмечали небольшой опыт работы со словарями, что должно повлиять на их стратегии работы со статьями в силу отсутствия специфических знаний о структуре словарных статей и правилах лексикографической эмфазы.

Эксперимент был разделен на два блока по типу выполняемого задания и изменяемым графическим параметрам текста, и для каждого блока было отобрано по три словарных статьи, максимально близких по структуре, объему и уровню сложности.

В первом блоке перед участниками была поставлена задача найти в словарной статье примеры употребления слова. Варьировалось оформление текстовых иллюстраций: курсив, прямое написание и цветовое выделение (шрифт зеленого цвета). Выделены были примеры употребления слова, а также производных этого слова других частей речи (например, «непомерный» – «непомерно»). Во втором блоке нужно было выбрать толкование многозначного слова из нескольких возможных вариантов, соотносящееся с приведенным в задании контекстом. Вариантами подачи начала определений многозначных слов были: полужирная цифра, обычная цифра и сочетание абзацного отступа и цифры. Каждая из подготовленных для эксперимента статей существовала во всех вариантах оформления, и псевдослучайное предъявление было организовано таким образом, чтобы каждый испытуемый увидел по одному разу каждую статью и каждый вариант графического выделения или верстки текста, а в целом в группе были в равной степени предъявлены все варианты комбинаций «статья – графическое выделение». В итоге было получено по 9 стимулов для каждого из двух блоков, а также 2 отдельные тренировочные статьи для демонстрации в начале блоков.

Каждый выбранный для исследования вариант реализации параметра имеет реальный словарный прототип, поэтому выбранные варианты можно считать представителями разных типов эмфазы в практической лексикографии. Так, курсивная подача примеров используется в большинстве традиционных словарей, прямое написание характерно для многих электронных версий бумажных словарей, цветовое выделение речений наблюдается в основном в онлайн-словарях, в том числе в Викисловаре. Различные варианты подачи нумерации пунктов в тексте также распространены как в цифровых, так и в печатных словарях.

Задание с примерами располагалось в верхней части экрана, отделенное от текста статьи чертой. После обнаружения целевой части статьи нужно было навести на нее курсор и щелкнуть левой кнопкой мыши. На выполнение задания респондентам давали неограниченное время.

Для регистрации движений глаз использовали айтрекер Eyelink 1000+ с частотой записи 500 Гц и калибровкой по 13 точкам. Перед началом каждого блока демонстрировали тренировочное задание, данные которого не учитывались.

После проверки качества регистрации данных в анализ параметров движений глаз были включены записи 44 участников. Статистический анализ проводили в программе Statistica 10.

## **Результаты**

Был проведен двухфакторный дисперсионный анализ влияния типа эмфазы и варианта стимула (текста конкретной статьи) на время выполнения задания, время обнаружения целевой части статьи, долю правильных ответов, длительность фиксации при чтении. Поскольку внутри блока задание было одним и тем же, также было проверено влияние фактора порядка предъявления стимула на эти параметры.

В первом блоке (поиск примера употребления слова) ни оформление текстовых иллюстраций, ни вариант стимула не повлияли на скорость работы, время обнаружения целевого текста или долю правильных ответов. Ошибкой испытуемого считался выбор примера употребления не целевого слова, а однокоренного другой части речи. При этом был очень выражен эффект обучения: от первого к третьему стимулу среднее время выполнения задания снизилось с  $10.446 \pm 6.099$  секунд до  $6.530 \pm 3.165$  секунд ( $F(2, 132) = 7.97$ ,  $p < .001$ ). Оформление текста значимо повлияло на длительность фиксаций при чтении статьи, в то время как сам текст и порядок стимулов влияния не оказали: для прямого текста средняя длительность фиксации  $214 \pm 88$  мс, для курсива —  $227 \pm 98$  мс, для цветного текста —  $225 \pm 104$  мс ( $F(2, 2688) = 3.67$ ,  $p = .026$ ). Парные сравнения в тесте Стьюдента показали значимые отличия длительности фиксаций на прямом тексте от двух других вариантов ( $t = 2.850$ ,  $p = .009$  для курсива;  $t = 2.469$ ,  $p = .03$  для цветного; сделана поправка Бонферони для множественных сравнений).

Во втором блоке (поиск подходящего к контексту толкования многозначного слова) ни оформление, ни сам текст, ни порядок стимула не повлияли на общее время выполнения задания. При этом целевую часть текста обнаруживали быстрее всего при выделении номеров значений полужирной цифрой ( $10.22 \pm 4.28$  секунд); вторым по скорости был поиск при разделении значений по абзацам ( $12.10 \pm 5.21$  секунд); наконец, выделение значений простой цифрой давало самый длительный поиск ( $12.96 \pm 5.16$  секунд),  $F(2, 97) = 3.13$ ,  $p = .048$  для оформления, текст и порядок стимулов значимо не влияли. Парное сравнение в тесте Стьюдента показало значимые различия только в паре полужирный-простой текст ( $p = .015$ ). При этом доля правильных ответов снижалась в случае оформления нумерации полужирным шрифтом: 67.3% для полужирной нумерации, 82.5% для простой нумерации, 94.1% для абзачного разделения (различия значимы в точном тесте Фишера  $2 \times 3$ ,  $p = .0098$ ). Ошибкой испытуемого считался выбор неправильного значения слова для предложенного контекста. При анализе средней длительности фиксаций на тексте статьи на нее значимо повлиял фактор оформления эмфазы ( $F(2, 5403) = 3.38$ ,  $p = .034$ ), но не остальные факторы. Эффект достигался за счет более высокой длительности фиксаций при разбивке на абзацы ( $219 \pm 89$  мс) по сравнению с простым ( $212 \pm 79$  мс) или полужирным ( $213 \pm 84$ ) текстом.

### Обсуждение и выводы

Результаты анализа первого блока говорят о том, что не всегда лексикографическая эмфаза в виде той или иной модификации шрифта (цвет, курсив) может значимо повлиять на результативность зрительного поиска релевантной информации в словарной статье, хотя и может спровоцировать более медленное чтение выделенного фрагмента текста. При этом для скорости выполнения задания в первую очередь оказался важен эффект обучения, когда в ряду однотипных заданий участники запоминают расположение целевой области. В результате анализа второго блока мы можем отметить существование баланса скорость-точность. Некоторые способы выделения нумерации вариантов

значения, видимо, провоцируют испытуемых к спешке и выбору неверного, но расположенного раньше варианта. При этом разделение вариантов абзацными отступами оказалось оптимальным для поиска правильного варианта значения слова. Подобные закономерности мы обнаружили у неопытных пользователей словарей, не знакомых с общими принципами строения статей и лексикографической эмфазы; в случае наличия экспертизы можно ожидать более эффективное использование невербальных средств разметки текста.

На основе данной работы мы можем сказать, что неопытные пользователи словарей ограниченно пользуются лексикографической эмфазой для решения задач поиска: курсивное или цветное выделение целевого текста не помогает при отсутствии знаний о структуре статьи, а организация нумерованного списка отдельными абзацами предпочтительна по сравнению с подачей в одном абзаце при поиске нужного варианта определения.

## Литература

*Каламбет Е.В.* Проявления антропоцентризма в современной отечественной лексикографии (на материале лингвистических словарей): автореф. дис. ... канд. филол. наук. Кубанский государственный университет, Краснодар, 2007.

*Atkins B.T., Rudell M.* The Oxford guide to practical lexicography. Oxford: Oxford University Press, 2008.

*Gagl B.* Blue hypertext is a good design decision: No perceptual disadvantage in reading and successful highlighting of relevant information // PeerJ. 2016. Vol. 4. P. e2467. <https://doi.org/10.7717/peerj.2467>

*Müller-Spitzer C., Michaelis F., Kopenig A.* Evaluation of a new web design for the dictionary portal OWID. An attempt at using eye-tracking technology // Using Online Dictionaries / C. Müller-Spitzer (Ed.). Berlin: De Gruyter, 2014. P. 207 – 228. <https://doi.org/10.1515/9783110341287.207>

*Schneider S., Beege M., Nebel S., Rey G.D.* A meta-analysis of how signaling affects learning with media // Educational Research Review. 2018. Vol. 23. P. 1 – 24. <https://doi.org/10.1016/j.edurev.2017.11.001>

## DOES TEXT FORMATTING AND LAYOUT HELP IN A SEARCH FOR INFORMATION IN THE DICTIONARY?: AN EYE TRACKING STUDY

A. N. Puchkova\* (1, 2), A. A. Berlin Khenis (1), A. I. Olchovskaya (1)  
[puchkovaan@gmail.com](mailto:puchkovaan@gmail.com)

1 – Pushkin State Russian Language Institute, Moscow;

2 – Institute of Higher Nervous Activity and Neurophysiology of RAS, Moscow

**Abstract.** Modern dictionaries, especially digital ones, use various methods of lexicographic emphasis – nonverbal ways of marking up and highlighting certain information in an article. The choice of specific methods is up to the compilers' preferences. Eye tracking complements the traditionally used methods of questioning and observation and allows us to obtain objective and accurate data on the parameters of the search for target information in an article with different variants of markup and design. For this study, we selected the following font and layout parameters: italicized, colored or plain text variants of the word usage and presentation of polysemous word's meaning numbers in simple, bold figures or

with additional breakdown into paragraphs. The task was to find an example of the target word's usage or its appropriate meaning. Eye movements during reading were recorded with an Eyelink 1000+ eye tracker. Color and italics did not affect the speed or accuracy of the search, and fixations were shortest on simple text. When searching for a meaning variant, paragraphing was optimal, and bold numbers provoked faster but less correct choices.

**Keywords:** dictionary, reading, text search, text highlighting, eye tracking, lexicography

The study was supported by the state assignment FZNM-2020-0005.