ОБРАЗОВАНИЕ И КУЛЬТУРА / EDUCATION AND CULTURE

Оригинальная статья / Original article

Modeling the Meaning of Individual Words Using Cultural Cartography and Keystroke Dynamics

T. A. Litvinova[a]✉, **O. V. Dekhnich**[b]
[a] *Voronezh State Pedagogical University, Voronezh, Russian Federation*
[b] *Belgorod State National Research University, Belgorod, Russian Federation*
✉ *centr_rus_yaz@mail.ru*

Abstract

Introduction. Revealing the psychologically real, individual meaning of the word as opposed to its dictionary meaning is the important task since such knowledge is crucial for effective communication. This is especially true for the words which denote key ideas and concepts of the culture. The word association experiment has been one of the most used methodologies to examine individual meaning of the word but it has been heavily criticized because of its subjectivity. In some of the recent works, data from language models and methods of vector semantics have been used to solve this problem. However, firstly, the very set of the features by which the meaning of the word is described is not uniform, which does not allow for a comparison of the results, and, secondly, some other types of data related to word production (i.e., behavioral data) are typically not taken into account. The aim of the present study is to reveal and systematically describe individual differences in the psychologically real meaning of the particular key words of the Russian culture using a new methodology which could be applied to any word association task. We propose to analyze data of different types (semantic features and keystroke dynamics markers) obtained during word association production to reveal individual differences in the word meaning.

Materials and Methods. The material of the study is a newly developed dataset containing associative reactions to the keywords of Russian culture, anonymized data about the informants, as well as the reaction time while producing associations measured using a program that records keystrokes. The proposed research methodology includes both the existing approaches (automatic extraction of relations from texts based on data from language models and methods of vector semantics, i.e., "cultural cartography using word embeddings") and a new list of features developed by the authors to describe individual differences in the meaning of a word based on the data from neurobiology about the meaning structure of word. A set of data analysis methods (linear mixed models, principal components analysis, hierarchical clustering on principal components) implemented in R packages is used to reveal individual differences in the word meaning in terms of the proposed list of features and association of the revealed differences with participants' characteristics.

Results. The cluster analysis showed the presence of two to three variants of psychologically real meanings for the 9 studied cue words which are listed among the key words of Russian culture. Systematic differences in the individual meanings of the words according to the proposed set of semantic features reflecting different aspects of semantic representations of word meaning in the human brain are described in detail, and a connection between specific features of the word meaning and the characteristics of the participants and markers of keyboard behavior are established for the first time.

Discussion and Conclusion. The specific scientific results related to the individual differences in the psychologically real meanings of the words, as well as fully reproducible methodology proposed in this paper (the dataset and code of this study are available on GitHub) can be used in the practice of effective teaching of Russian as a foreign language, in the study of the changes in semantics of the key words of the culture based on text data, for

designing effective political and advertising campaigns, etc. Among strands of the future research are the study into the effect of the different characteristics of the cue words on their semantic features and participants' keystroke behavior, the broadening of the list of the proposed characteristics, the use of new language models and text corpora for the further development of an important theoretical and applied problem of revealing and describing the psychologically real word meaning.

Моделирование индивидуального значения слова с использованием методов культурной картографии и данных клавиатурного почерка

Т. А. Литвинова[1 ✉], **О. В. Дехнич**[2]

[1] *Воронежский государственный педагогический университет, г. Воронеж, Россия*
[2] *Белгородский государственный национальный исследовательский университет, г. Белгород, Россия*
✉ *centr_rus_yaz@mail.ru*

Аннотация

Введение. Выявление и описание психологически реального значения слова актуально для ключевых слов, обозначающих важные и показательные для отдельно взятой культуры идеи, образы, представления. В работах последних лет для решения названной задачи эффективно привлекаются данные языковых моделей и методы векторной семантики. Однако набор описываемых значений слов, признаков не универсален, что не позволяет сопоставлять результаты; исследователями не учитываются другие типы характеристик, помимо семантических. Цель исследования – установление индивидуальных различий в психологически реальном значении слова с использованием автоматического инструментария на основе данных разных модальностей (семантических признаков нового типа и маркеров клавиатурного поведения).

Материалы и методы. Материалом исследования является авторский датасет, содержащий ассоциативные реакции к ключевым словам русского языкового сознания, обезличенные данные об информантах, а также признаки, описывающие клавиатурное поведение участников во время продуцирования ассоциаций. Методология включает современные подходы (опора на данные языковых моделей и использование методов векторной семантики – культурная картография), а также разработанный авторами список признаков для описания значения слова на основании данных нейробиологии. Методы анализа данных (линейные смешанные модели, метод главных компонент, кластеризация на главных компонентах), реализованные в пакетах на языке R, устанавливали индивидуальные различия в значении слов, описанных через предложенные группы признаков и их связи с характеристиками информантов.

Результаты исследования. Кластерный анализ показал наличие от двух до трех вариантов психологически реальных значений для девяти исследуемых слов-стимулов, входящих в ядро русского языкового сознания. Подробно описаны конкретные различия в значениях слов по системе обусловленных нейробиологических семантических признаков, а также установлена связь выделенных вариантов значений слов с характеристиками информантов и маркерами их клавиатурного поведения.

---

[1] Olga Dekhnich received no financial support for the research, writing, and publication of this article.

Обсуждение и заключение. Полученные результаты могут быть использованы в практике эффективного обучения русскому языку как иностранному, для исследования изменения семантики ключевых слов русской культуры на основании анализа текстов, проектирования эффективных избирательных и рекламных кампаний и др. Среди важнейших задач наших будущих исследований – установление связи различных характеристик ключевых слов, их семантических признаков и особенностей клавиатурного поведения информантов, расширение списка предложенных признаков, использование новых языковых моделей и корпусов текстов для дальнейшей разработки актуальной теоретико-прикладной проблемы исследования и системного описания психологически реального значения слова.

Introduction

Do we mean the same thing when we use one word or another? The problem of revealing and describing the psychologically real meaning of a word, i.e., its functioning in the individual mental lexicon (as opposed to the word meaning represented in dictionaries – lexical meaning) is one of the actively researched problems of psycholinguistics. L. Vygotsky uses the concept of *sense* (smysl) to refer to the functioning of words in individual's system of meaning. Vygotsky describes *smysl* as an important component in the system of meaning and stresses the divergence between individual's sense of the word, common usage based on dictionary meanings and even sociocultural meaning ("meaning in a social context") which is considered to be an essential but subordinate part of sense: "Ultimately, the word's real sense is determined by everything in consciousness which is related to what the word expresses… [and] ultimately sense depends on one's understanding of the world as a whole and on the internal structure of personality"[2].

Identification and systematic description of individual differences in word meaning as well as the establishment of associations between such differences and various characteristics of individuals are important tasks which have not only theoretical but also practical implications in marketing, education, politics, etc., since this knowledge is very important for efficient communication. For decades, the main method for studying individual differences in word meaning has been word association experiment. However, such a methodology has disadvantages related, firstly, to the labor intensity of analysis, and secondly, to the subjectivity of the interpretation of its results and, therefore, the difficulty of comparing the findings obtained in different works.

In the last decades, distributional semantics which presents a usage-based model of meaning has become of the mainstream approaches to study the word meaning. Distributional semantic models (DSM) construct multi-dimensional (typically a few hundreds) graded word representations in the form of vectors (word embeddings) which capture

---

[2] Vygotsky L.S. Vol. 1. Problems of General Psychology, Including the Volume Thinking and Speech (Cognition and Language: A Series in Psycholinguistics). New York: Plenum; 1987.

many rich and nuanced aspects of the meaning, by extracting word co-occurrences from corpora [1; 2]. In constructed semantic space semantic relations are modeled as geometric relations, which is necessary since individual features lack the meaning. The resulting geometric relationships in DSM correspond to semantic relationships in language [2].

This methodology (DSM aligned with relational theories of meaning) has become widely used in many fields, including sociology and cultural studies. The meaning is central for cultural analysis, and formal analysis of a texts as the main source of meaning is a very important method. The authors demonstrated usefulness of word embeddings (WE) and a set of methods for relation extraction for cultural cartography, which is the process of revealing the meaning of a text "by the extent it references certain concepts or entities" [3]. This name was given to this methodology since "like a topographic map of terrain, it selectively simplified texts in useful ways" [4]. Typically, count-based approach is used for text analysis, but it has serious drawbacks: it is ill-suited for measuring magnitudes of conceptual engagement and similarity which are central to cultural analysis. On the contrary, WE preserve the graded, relational meanings of words and thus are ideal methodology for formal analysis of texts in cultural studies.

Specifically, the development of models of distributional semantics makes it possible to obtain estimates of texts that reflect their position on any antonymic scales (often referred to as cultural dimensions, i.e., generic binary oppositions that "individuals use in every-day life to classify agents and objects in the world" [5]. This methodology has also been successfully applied for the analysis of the results of word association experiment [6].

However, when describing the meaning of a word using these methods, binary oppositions are constructed based on various criteria that are suitable for a particular case, sometimes subjective ones. There is a need to create a universal and theoretically justified set of features which could be used for construction of word meaning throughout different tasks and texts.

In addition, existing works do not exploit the potential of analyzing data from some other modalities, such as typing data, which can be recorded using the keyboard behavior recorder programs that are widely used in modern writing research [7].

Keystroke dynamics captures keypress-related metadata (e.g., timing information of key down press and release time, inter-word and intra-word pause durations, etc.). Intuitively, typing on a keyboard utilizes multiple cognitive domains. It is widely used for the study of the writing process [8; 9], but has also been actively applied for different domain – from user identification [10] to early sclerosis detection [11] and lie detection [12]. However, the use of keystroke data for the study of the word meaning is very limited [13].

The purpose of the study is to establish individual differences in the psychologically real meaning of key words of Russian culture, i.e., words which denote ideas, concepts and representations most important for Russians, using data obtained in the course of a word association experiment with recorded keystroke dynamics and processed using methods of cultural cartography with word embeddings and a newly-developed set of neurobiologically justified features which reflect brain-based componential semantic representation.

The presented methodology is fully reproducible[3] and can be applied to the results of any word association experiments, as well as to the texts of any length to gain insight into the underlying semantic representation that different individuals have about specific words or concepts, which is important for the theory and practice of communication, for planning marketing and electoral campaigns, preparing new textbooks on lexical acquisition for second language (L2) learners, etc.

Literature Review

Distributional semantic models (DSM) make it possible to explore the problem of the word meaning in a new way and are undoubtedly among the most important achievements of modern linguistics. A. Utsumi showed that they encode concrete, abstract, spatial, temporal, perceptual, and emotional knowledge [14].

---

[3] Our code and dataset are available at: https://github.com/Litvinova1984/cultural-cartography.

One of the main theoretical results obtained with the help of such models was the statement about the systemic connection "between the knowledge that people acquire and the experience that they have with the natural language environment" [15]. However, language experience is inherently variable; its formation is influenced by different variables – demographic, cultural, etc. ones, as a result of which individual differences in the meaning of a word are observed [16]. The presence of differences in the word meanings among speakers of the same language from a relatively homogeneous cultural/educational background has been revealed based on both behavioral ratings and brain activation patterns [17] where it was shown that the magnitude of individual disagreements on the word meanings could be modeled on the basis of how much language or sensory experience is associated with a word and that this variation increases along with word abstractness.

The authors of the above cited works make an important conclusion about the need for further research of individual differences in word meaning, since it is clear that the causes of communication failures, especially in the areas such as politics, sociology or legal domains where there are many terms without external referents lie not only in the contextual use of the words, but in their different understanding among different people.

The presence of such differences in the word meaning, associated with the individual nature of linguistic experience, has also been proven in the works using the methodology of constructing individual DSM with their subsequent alignment [15; 18] as well as DSM constructed on texts written by peoples from different cultural, political groups, etc. [19].

Thus, the latest advances in the field of computer semantics and neurobiology indicate the differences in the understanding and the use of the words even among speakers with a common background, but there is no systematic description of such differences, despite its extreme importance, which is due to a number of methodological and theoretical reasons. Thus, the construction of individual semantic models requires a large number of texts from each author (e.g., in [15; 18], several million tokens from each author were analyzed), as well as the use of special methods for aligning such models, while the problem of choosing the optimal method of alignment is still open [20].

In view of these problems, a methodology which combines the use of easily available pretrained DSM and the set of relation extraction techniques was introduced and successfully applied for the range of tasks [21]. This methodology could be applied for texts of any length and to any number of texts (even to one text) and could extract the location of any text on the semantic pole defined by any juxtaposing terms irrespective of their presence in the text (they should be in DSM). With this approach, it is possible to arrange objects by size, gender, dangerousness, intelligence, temperature, speed, and so on [22]. This methodology is especially useful for revealing stereotypes and understanding social identities (white – non-white, rich – poor and so on) [3]. The set of such juxtaposing terms varies from task to task, however, there is a need to construct universal set of features to reveal the difference in the meaning of any word among people from different social, demographical groups, etc. We argue that this could be done using data from neurobiology. It is known that the nature of an individual variation in the word meaning are related to the general principles of its representation in the human brain as well as variables affecting this variation [17]. A basic set of approximately 65 experiential attributes of semantic representation based on neurobiological considerations, comprising sensory, motor, spatial, temporal, affective, social, and cognitive experience was introduced in [23]. It was shown that these features are encoded in WE and could be predicted with a fairly high accuracy (while some features are predicted more efficiently than others) [14; 24].

It seems promising to supplement the studies of the individual meaning of a word with the behavior data, in particular with that about the keystroke dynamics during word association production, especially with that on the duration of pauses between the cue and associates. Pauses are considered as behavioral correlates of cognitive processes. Studies using pause data in examining word meaning are rare while it is claimed that

keystroke logging could be a "breakthrough in WA methodology which can unlock its undoubted potential" [25].

Reaction time – "oral" analogue of pause duration measured by keystroke logging software – has been used as measure of stimulus affectivity. Rapaport[4] showed that reactions to traumatic stimulus words had longer delays than those to neutral stimuli. More recent works have shown that emotional words typically evoke longer reaction time than neutral stimuli (usually, the number of prolonged reactions is calculated; different threshold is used, but as usual pauses above 3 seconds have been considered as prolonged [26]).

The influence of linguistic properties of cue words on reaction time has also been studied. In it was shown that a cognitive workload is manifested in reaction times, and abstractness could be responsible for associative difficulty; while emotionality does not[5].

For typed associations, such works are rare, as we mentioned earlier. M. Aldridge et al. have shown that pause duration is related to the strength of links in lexical selection processes [25]. Using the pause data and word frequency information, the authors of [13] proved the presence of semantic drift over the short time (25 seconds) of a free word association task. They observed a notable decrease in the diversity of terms generated earlier in the task, while more unique terms (a greater diversity and relative uniqueness) were generated in the 4th time quartile. The authors argue that revealed semantic drift might serve as a scalable indicator of the invocation of language versus simulation systems. To the best of our knowledge, works which combine semantic attributes of the word meaning and keystroke data are absent.

Thus, in the present work we attempt to probe a complex methodology for systematic description of the individual differences in word meaning using data and methods from distributional semantics, cultural cartography, keystroke logging research, neurobiology of semantics. The proposed methodology is fully reproducible and could be applied for different units of analysis (text, data from word association experiments) and different tasks.

Materials and Methods

Material. The material for this study was the RuPersWordAssociation dataset[6] which contains associative reactions to 50 carefully selected cue words which are listed among the key words of Russian culture [6]. There are many definitions of key words of culture [27]; following O.V. Zagorovskaya, we consider as key words of Russian culture those words which denote most important ideas, concepts and representations of traditional Russian culture; reflect the most essential features of the worldview (mentality) of the Russian people and are the "key" to understanding of the most important fragments of the Russian culture; preserve the collective experience of the Russian people, Russian spiritual and moral values in their meanings [27].

RuPersWordAssociation dataset is, to the best of our knowledge, the largest existing (at least from publicly available) word association database in terms of breadth of metadata about the informants (demographics, personality traits) and about association data per se (pause duration, semantic similarity metrics between the cue words and reactions) and in terms of linguistic annotation (more than 22 000 "cue word – associate reaction" pairs was manually annotated for the type of a relation from carefully constructed list).

The uniqueness of the dataset for the study of an individual word meaning is that it contains associative responses in a non-aggregated form since aggregating word association across the participants makes it difficult to determine the mechanisms of association and the characteristics of an individual meaning.

We asked our participants ($n = 49$) to produce five responses based on the studies of category production, where recent responses remain active in working memory and can influence up to five subsequent responses [28]. Therefore, we can consider the resulting

---

[4] Rapaport D., Schafer R., Gill M. Diagnostic Psychological Testing: The Theory, Statistical Evaluation, and and Diagnostic Application of a Battery of Test. Chicago: Yearbook Publishers; 1946. 516 p.

[5] Brown W.P. A Retrospective Study of Stimulus Variables in Word Association. *Journal of Verbal Learning and Verbal Behavior.* 1971;10(4):355–366. https://doi.org/10.1016/S0022-5371(71)80034-8

[6] The dataset is freely available at: https://github.com/Litvinova1984/word_association_dataset

associative rows as the representative units for the cue word meaning analysis. More details on the RuPersWordAssociation dataset could be found in the conference paper [29]. All respondents were informed about their participation in the study.

For this particular study we selected 9 cue words (ДОБРО "good", ДОМ "home, house", ДРУГ "friend", ЖИЗНЬ "life", МИР "world", НАСТОЯЩИЙ "real", СЕМЬЯ "family", СЧАСТЬЕ "happiness", ХОТЕТЬ "want"). These particular words were selected for the current study as they were presented twice (in a random order) in each questionnaire (other words were presented once) to examine their meaning using methodology presented in this paper as comprehensively as possible.

When we constructed the questionnaire for RuPersWordAssociation, we selected these particular words to be presented twice as they are listed in available word association resources (dictionaries and databases) annotated for respondent demographics since RuPersWordAssociation was created in the course of the larger project aimed at revealing the characteristics of word meaning in their relation to informants' demographics.

The final dataset contained 740 associative rows. The data on the respondents' gender was considered as a categorical variable in the subsequent analysis, data on the psychological characteristics (Big-5 scores and scores on Differential Emotions Scale) as quantitative variables [6; 28].

Methods. Semantic Feature Set Construction. To construct our feature set, we used CMDist function from text2map package [30] which takes the word counts from a document-term matrix (DTM) as the input, a matrix of word embedding vectors, and a set of concept words (or vectors). The "cost" of transporting all the words in a document to a single vector or a few ones (denoting a concept of interest) is the measure of engagement with the concepts [31; 32].

It is possible to use the offset of several juxtaposing words using a function get_direction from text2map package to extract the engagement of a text with this or that pole of the scale (higher numbers indicate the closeness to the first member of opposition).

We propose a novel approach for the construction of the list of terms which constitute semantic directions. Specifically, we used the set of brain-based components of semantic presentation form [23]. Each of these components – for which "there are likely to be corresponding distinguishable neural processors, drawing on evidence from animal physiology, brain imaging, and neurological studies" [23] – belongs to one of 13 different domains – "aspects of mental experience". In this work, an example of the words rated high on these features was proposed. We used both translations of these words to Russian to construct our semantic oppositions and also data from psycholinguistic databases and dictionaries. E.g., for constructing dimensions which are related to the visual modality we used the database with a human rating of different words for this modality described in [33] (we selected top-50 words with the highest and lowest rank on the visual modality and constructed oppositions), words from the group "Vision" from Russian version of the LIWC thesaurus for creating a semantic region related to visual words [34], etc. We believe that using different sources for constructing dimensions is necessary to obtain reliable results.

We also constructed the list of basic semantic oppositions based on the data presented in [23], different lists of oppositions which are used in semantic differential methodology [6].

The resulting feature set is presented on GitHub[7].

We used a pretrained model ruwikiruscorpora_upos_cbow_300_10_2021 which was trained on Wikipedia and National corpus of the Russian language. The model released in December 2021 is the most recent one available for download and is the closest to the date of the creation of our dataset (October – December 2022).

Keystroke Dynamics Feature Set Construction. We processed the pause data as follows. First, we extracted outliers using function boxplot.stats from the grDevices package (the values that are beyond the "whiskers", i.e. those above 8 753 milliseconds (ms). We added the number of such

---

[7] https://github.com/Litvinova1984/cultural-cartography

pauses as a separate feature (Long_pauses). Then we inspected the remaining values. A minimal pause duration was 114 ms, median = 1 759 ms, mean = 2 422 ms. Based on this data as well as on the previous literature [35], we set a threshold of 2 seconds for cognitive pauses (i.e., we counted the number of pauses longer than 2 seconds and considered them as cognitive ones). This threshold has been used by researchers for many reasons, including that it is twice as long as the mean typing rate and for the ease of comparison of the results in different works.

Data Analysis Methods. We applied a set of modern data analysis methods including linear regressions and linear-mixed effect models, principal component analysis (PCA), hierarchical clustering on principal components (HCPC). All the stages of the analysis were performed in R. To build the linear regressions, we used lme4 package [36]. We performed PCA for exploratory analysis of our feature set using FactoMineR package [37] and its functions (dimdesc and catdes) allowing us to establish the connections between the components and the qualitative (gender) and quantitative characteristics (scores on the psychological tests and age) of our informants taken as supplementary variables. Supplementary variables have no influence on the PCA. They are used to interpret the results of the analysis.

A workflow of the study is presented in Fig. 1.

Results

Pause Behavior during Word Association Production. As Fig. 2 shows, there are differences in pauses duration depending on the pause location (i.e., between a cue word and the first reaction – PAUSE 1, the second reaction – PAUSE 2, etc.). To test these differences for their significance, we performed the Kruskal–Wallis test which proved the difference between the durations of pauses depending on their positions (Kruskal–Wallis chi-squared = 242.96; df = 4; *p-value* < 2.2e- 16). Pairwise Wilcoxon signed rank comparisons with Bonferroni correction shown the differences between the duration of the PAUSE 1 and other pauses, PAUSE 5 and other pauses (there were no differences in duration between PAUSE 2 and PAUSE 3, PAUSE 3 and PAUSE 4, PAUSE 2 and PAUSE 4), with the first pause (i.e., between a cue word and the first associate) being the longest one.

Eta-squared estimate (the measure of the Kruskal–Wallis effect size) calculated using kruskal_effsize function from rstatix R package is 0.0704, which corresponds to the moderate effect [38].

This finding corresponds to the results obtained by S. MacNiven and R. Tench about the existence of miniclusters in association rows related to a semantic shift [13]. In our data, based on the pause behavior, the first associate constitutes the first cluster, associates from 2nd to 4th – the second cluster, and 5th associate makes up the third cluster. In future studies,
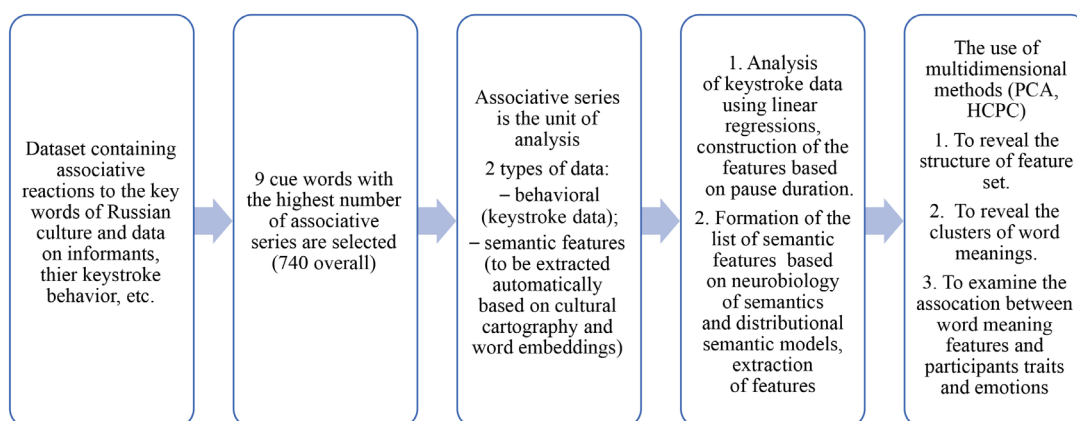

F i g. 1. Workflow of the study
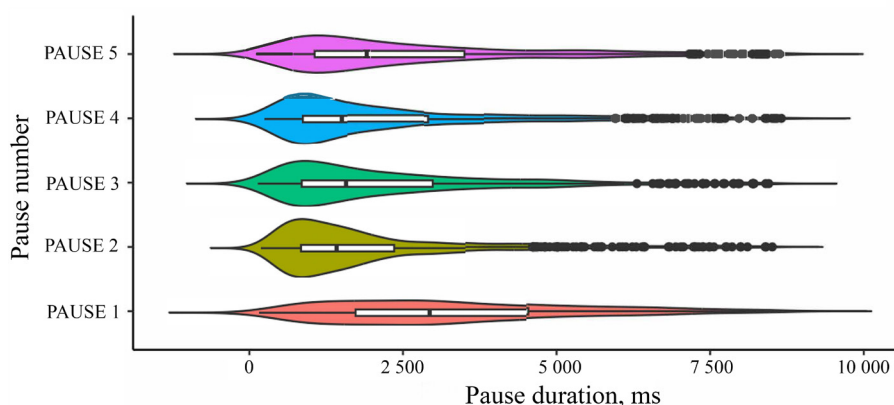
Source: Compiled by the authors.

F i g. 2. Duration of pauses between stimulus and reactions depending on positions

Source: Hereinafter in the article all figures are made by the authors in the R environment using the ggplot2 library.

it would be interesting to compare the pause duration and semantic similarity indices to further inspect the presence of a semantic shift in the word association production based on the data from two modalities.

Further, we aimed to inspect if there were any differences between the numbers of cognitive pauses depending on the stimulus word and informants. Kruskal–Wallis rank sum test showed the absence of differences between the number of cognitive pauses for different stimulus words (Kruskal–Wallis chi-squared = 3.6546; df = 8; *p-value* = 0.8869) but confirmed them for informants (Kruskal–Wallis chi-squared = 9.7125; df = 1; *p-value* = 0.00183).

We further examined the effect of the stimulus word and the informant on the number of cognitive pauses using a linear regression apparatus. We first built a basic model using glm function from stats package (the number of cognitive pauses was log-transformed), then, using lmer function from lme4 package, the following models were built: 1) a model with the author as a random effect; 2) a model with the stimulus word as a random effect. A comparison of the basic model with two models with random effects using the AIC criterion showed that a decrease in this criterion is observed only for the model with the informant as a random effect.

Therefore, this analysis has revealed (for the first time for the Russian language)

both general trends of keystroke behavior while producing word associations (existence of three "blocks" with borders in certain positions) and the absence of the general effect of the stimulus word on the number of cognitive pauses (which were previously shown to be related to the emotional or cognitive state of the participants) and the presence of the effect of the participant, which could indicate a different meaning of the same words for different individuals.

Multidimensional Analysis of Semantic and Keystroke Data. For each stimulus word, we performed PCA on our feature set (46 semantic features and 2 features describing keystroke dynamics – the number of cognitive pauses and the number of long pauses) with the calculation of the correlation between the main components (using FactoMineR function dimdesc) and the qualitative and quantitative variables (as a rule, the first two components which explained most of the variation – more than 50% – are considered). Then we performed hierarchical clustering on principal components (HCPC). The number of clusters was determined based on the relative gain of inertia (visual inspection was also performed). Then we inspected which variables have the largest contribution to the cluster division (using eta2 criterion implemented in FactoMiner) and characterize the clusters.

PCA performed on the features describing the meaning of stimulus word ДОМ "House, home"[8] with a follow-up description of the

---

[8] Visualization is presented on Github: https://clck.ru/3EsmpU.

dimension has shown that the following features have positive correlations with the first component (PC1) (only the features with the highest correlation coefficients are discussed): CognitionImage, SomatNorms, MotorPractice, VisNorms, OlfacNorms (i.e., features related to sensory domains), emotional states of participants Fear, Anxious-Depressive, Guilt, Long_pauses. Negative correlations are observed for PC1 and the number of cognitive pauses, CognitionAbstract, SocialSelf, VisIntens, Causal, SocialLIWC, CognitionLIWC.

PC1 is related to gender ($R^2 = 0.1527$; $p = 0.0003$).

For the second component, high values of EmoPleasant, AudLIWC, VisLIWC, VisFace are contrasted with GustTaste, EmoDisgust, EmoAngry, EmoSentiment, AudIntens, SocialGender.

At the second stage, we performed clustering on the components (Fig. 3). The criterion "gain of inertia" suggests two-cluster solution. Cluster division is associated with gender (*p-value* of the chi-square test = 0.0012).

The first – "female" – cluster is characterized by the high values of semantic features related to cognitive and social domains (CognitionAbstract, CognitionLIWC, SocialSelf, SocialLIWC, Causal) spatial-temporal (TempDuration, TempAge, SpatialUpDown), visual (VisIntens) domains, emotions with a positive connotation (EmoHappy) and a large number of cognitive pauses.

The second – "male" – cluster is characterized by the high values of the features related to the sensory and motor domains (CognitionImage, MotorPractice, SomatNorms, VisNorms, SomatTexture, VisColor, OlfacNorms), numbers (SpatialNumber), male pole of the "male-female" opposition (SocialGender), negative emotionality (EmoSentiment, EmoAngry), a large number of long pauses, high values of the scores on Guilt scale (one of the emotional states of the informants).

Thus, there are two clusters of the meanings of the word ДОМ which are related to gender: the first meaning is based primarily on the features from the cognition and social domains and is associated with positive emotions, the second – one – on sensory-motor ones, which are different systems for representing knowledge [35].

PCA for the cue word СЕМЬЯ "Family" shows that PC1 positively correlates with a lot of sensory and motor components in the word meaning (CognitionImage, SomatNorms, OlfacNorms, VisNorms, GustNorms, SomatTexture, MotorPractice, MotorBinder, SomatLIWC), as well as SpatialNumber and EmoPleasant. It is interesting to note that PC1 also positively correlates with scores on Agreeableness (one of Big-5 traits). A negative correlation is observed between PC1 and CognitionAbstract, CognitionLIWC, Causal, VisSize, TempAge, EmoSurprised. PC2 positively correlates with the emotional components in the word meaning (EmoPleasant, EmoHappy,
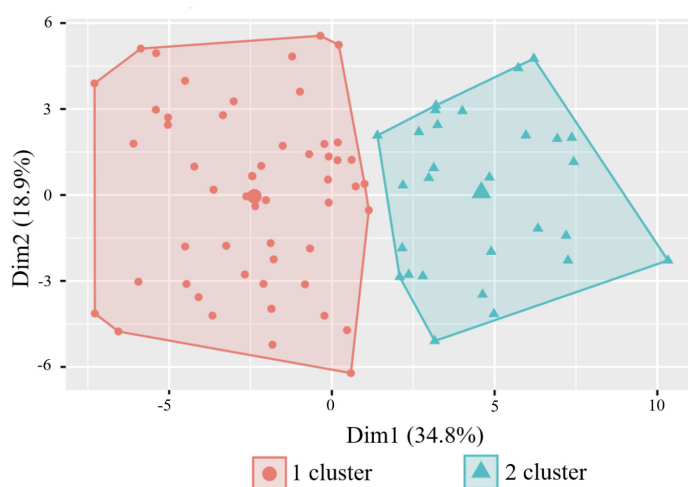

F i g. 3. The cluster analysis on principal components for the stimulus word ДОМ "House, home"

EmoBenefit) as well as visual (VisIntens, Vis-Face) and social ones (SocialSelf). It is also correlated with the scores on Joy (emotional state). Negative correlations are observed for the negative emotional components of semantics (EmoDisgust, EmoAngry, EmoSentiment), Drive, AudIntens and negative emotional state of participants (Contempt and Sorrow). It is notable that individual individual semantics of this word is strongly connected to the traits and emotions of the informants.

HCPC shows a three-cluster solution. The first cluster in the meaning of this word is described by the high values of CognitionAbstract, CognitionLIWC, Causal, Emo-Benefit, EmoHappy, DriveNeeds, TempAge, SocialSelf, and the low values of CognitionImage, MotorPractice, SomatNorms, VisNorms, MotorBinder. This cluster could be named as "Cognition and Positive Emotion".

The second cluster – "Negative emotions and sounds" – is characterized by the high values of the negative semantic components (EmoAngry, EmoDisgust, EmoSentiment), sensory-motor components (MotorPractice, AudIntens, AudNorms, CognitionImage), AttentionArousal, high scores on Sorrow, Contempt, AnxiousDepressive, the low values of EmoPleasant, EmoHappy, VisIntens, EmoBenefit, SomatProprioception, VisFace, CognitionAbstract, Contiousness.

The third cluster "Sensory, no sound and pleasant" has high values on SomatTexture, VisColor, EmoPleasant, VisFace, GustNorms, OlfacNorms, VisNorms, CognitionImage, low values on AudIntens, TempAge, EmoAngry, EmoDisgust, VisMotion, GustTaste, CognitionLIWC.

PCA for the cue word НАСТОЯЩИЙ "Real" shows the opposition of sensory components (CognitionImage, Somat-Norms, OlfacNorms, VisNorms, SomatTexture, GustNorms) and mostly the cognitive components (CognitionAbstract, CognitionLIWC, TempAge, Causal, DriveNeeds, GustTaste) on PC1. A positive correlation is also observed for PC1 and Neurotism, scores on the integral emotional scale PositiveIntegral, scores on Interest and Joy. PC2 is positively correlated with VisIntens, EmoPleasant, SocialSelf, EmoHappy, SomatSurface and Neurotism, negatively

with EmoAngry, EmoSentiment, Drive, EmoDisgust, GustTaste.

There are three clusters in the individual semantics of this cue word. The first cluster is described by the high values of Cognition-Abstract, CognitionLIWC, Causal, SpatialUp-Down, TempAge, EmoSurprised, SocialLIWC, low values of CognitionImage, MotorPractice, SocialGender, SomatNorms, OlfacNorms. The second cluster is characterised by the high values of SomatSurface, VisFace, Emo-Pleasant, VisIntens, SocialSelf, VisMotion, EmoHappy, SomatLIWC, Interest, low values of EmoDisgust, EmoFear, Causal, EmoAngry, AttentionArousal, Drive. The third cluster is characterized by the high values of EmoAngry, EmoSentiment, MotorPractice, SomatTexture, VisNorms, CognitionImage, the low values of SocialSelf, VisIntens, CognitionAbstract, EmoHappy, EmoPleasant, SomatSurface, CognitionLIWC.

As for the cue word ДОБРО "Good/Kindness", PCA shows an opposition between the sensory components (SomatTexture, GustNorms, VisColor, SomatLIWC, VisLIWC, VisNorms), scores on Contiousness and Joy, on the one hand, and GustTaste, AudiIntens, EmoAngry, TempAge, EmoDisgust, CognitionAbstract, on the other. PC2 positively correlates with SocialSelf, VisIntens, CognitionAbstract, EmoHappy, SpatialUpDown, CognitionLIWC, EmoFear, Contempt and negatively with MotorPractice, CognitionImage, EmoSentiment, EmoAngry, SocialGender.

Two clusters are allocated, the first one is characterized by the high values of Gust-Taste, EmoDisgust, AudiIntens, EmoAngry, TempAge, CognitionAbstract, Drive, Emo-Surprised, EmoSentiment, the second cluster is described by the high values of EmoHappy, EmoPleasant, SomatTexture, SomatLIWC, GustNorms, VisLIWC and the high scores on Joy, Contiousness, Extraversion.

PCA for the stimulus word СЧАСТЬЕ "Happiness" shows a positive correlation of PC1 and VisNorms, SomatTexture, Vis-LIWC, SomatLIWC, VisColor on very high level (higher than 0.85), as well as with the number of long pauses. Negative correlations are revealed between PC1 and Cognition-Abstract, GustTaste, TempDuration, TempAge,

the number of cognitive pauses. PC2 is positively correlated with SocialSelf, EmoFear, VisIntens, SocialLIWC, SomatNociception, negatively correlated with MotorPractice, Drive, EmoAngry, EmoSentiment, CognitionImage.

A two-cluster solution is revealed, with the first cluster being denser. The first cluster is characterized by the high values of TempDuration, EmoAngry, GustTaste, CognitionAbstract, AudiIntens, TempAge, EmoBenefit, the number of cognitive pauses, the second cluster is described by the high values of VisLIWC, SomatLIWC, VisColor, VisBody, MotorBinder, AudLIWC, SomatTexture, VisNorms, the number of long pauses, EmoPleasant, EmoHappy.

PCA for the cue word ДРУГ "Friend" shows positive correlation of PC1 and OlfacNorms, GustNorms, CognitionImage, SomatTexture, VisNorms, SomatNorms, MotorBinder, number of long pauses, Interest, Openness, and a negative correlation with CognitionAbstract, EmoSurprised, GustTaste, VisSize, DriveNeeds, TempAge, the number of cognitive pauses. An opposition of VisIntens, SpatialUpDown, EmoHappy, SocialSelf, EmoFear, EmoPleasant and EmoAngry, EmoSentiment, MotorPractice, Drive, EmoDisgust is observed on PC2.

Three clusters are revealed. The first cluster is characterized by the high values of GustTaste, EmoDisgust, EmoSentiment, VisMotion, EmoAngry, DriveNeeds, AudIntens, low values of EmoPleasant, GustNorms, SomatTexture, AudLIWC, VisFace, OlfacNorms, EmoHappy, EmoFear, SocialLIWC, VisNorms. The second cluster is described by the high values of VisIntens, SpatialUpDown, SocialSelf, EmoHappy, CognitionAbstract, SocialLIWC, EmoFear, EmoPleasant, SomatSurface, TempDuration, CognitionLIWC, the low values of EmoSentiment, MotorPractice, SocialGender, EmoAngry, Drive, SomatNorms, CognitionImage. The third cluster is described by the high values of CognitionImage, OlfacNorms, SomatTexture, GustNorms, SomatNorms, MotorBinder, EmoPleasant, Long_pauses, Openness, the low values of CognitionAbstract, VisSize, EmoSurprised, DriveNeeds, CognitionLIWC, TempAge, SocialSelf, EmoBenefit, GustTaste, the number of cognitive pauses.

PCA for the cue word ЖИЗНЬ "Life" shows that PC1 is positively correlated with the sensory components (VisColor, VisLIWC, SomatLIWC, CognitionImage, SomatTexture, GustNorms, VisNorms, OlfacNorms), the positive emotion components (EmoPleasant, EmoHappy), as well as the emotional states of the participants (Joy, PositiveIntegral), negatively with CognitionAbstract, Causal, TempAge, GustTaste, EmoAngry, EmoDisgust. PC2 positively correlates with SocialSelf, VisIntens, SocialLIWC, SpatialUpDown, EmoPleasant, SpatialProx, CognitionLIWC, Agreeableness, Extraversion, negatively with MotorPractice, EmoSentiment, GustTaste, SocialGender, EmoAngry, CognitionImage, EmoDisgust, Contempt, Shame.

A two-cluster solution is revealed, the first one being denser, with the high values on CognitionAbstract, Causal, CognitionLIWC, TempAge, TempDuration, GustTaste, SocialLIWC, EmoAngry, EmoDisgust. The second cluster is described by the high values of the sensory domain and concreteness (CognitionImage, VisColor, SomatLIWC) and positive emotions (EmoPleasant, EmoHappy), Interes, Joy, PositiveIntegral.

PC1 for the cue word ХОТЕТЬ "Want" is positively correlated with CognitionImage, GustNorms, SomatTexture, VisColor, VisLIWC, AudLIWC, SpatialNumber, SomatLIWC, negatively with TempAge, CognitionAbstract, VisSize, VisMotion, DriveNeeds, EmoDisgust. PC2 positively correlates with SocialSelf, SocialLIWC, VisIntens, TemporalLIWC, EmoPleasant, EmoFear, EmoHappy, negatively with EmoSentiment, SocialGender, Drive, EmoAngry, MotorPractice, GustTaste.

The first cluster is characterized by the high values of CognitionAbstract, TempAge, VisSize, CognitionLIWC, DriveNeeds, Causal, VisMotion, SocialSelf, EmoHappy, VisIntens, Contiousness and Agreeableness. The second cluster is characterized by the high values of CognitionImage, SomatNorms, SomatTexture, GustNorms, SpatialNumber, MotorPractice, VisLIWC.

PCA of the cue word МИР "World" shows positive correlation between PC1 and VisNorms, SomatNorms, CognitionImage, OlfacNorms, SomatTexture, VisColor, negative correlations between PC1 and CognitionAbstract,

TempDuration, TempAge, SocialLIWC, SocialSelf, scores on Astonishment. PC2 correlates positively with AudLIWC, VisIntens, EmoPleasant, EmoHappy, SocialSelf, SocialLIWC, Contiousness, negatively with EmoAngry, EmoSentiment, Drive, GustTaste, MotorPractice, EmoDisgust, Anger, Contempt.

Three clusters are revealed. The first cluster "Bitter, loud and negative" is described by the high values of GustTaste, AudIntens, TempDuration, Drive, EmoDisgust, EmoAngry, EmoSentiment, low values of EmoPleasant, AudLIWC, VisLIWC, VisColor, SomatTexture, EmoHappy, VisNorms. The second cluster – "Visual, social and positive" – is characterized by the high values of VisIntens, EmoHappy, SocialSelf, EmoPleasant, SocialLIWC, AudLIWC, CognitionAbstract, low values of EmoAngry, MotorPractice, EmoSentiment, SocialGender, Drive, CognitionImage, GustTaste. The third cluster – "Sensory" – is characterized by the high values of VisNorms, SomatNorms, OlfacNorms, CognitionImage, VisColor, SomatLIWC, low values of TempDuration, SomatSurface, GustTaste, CognitionAbstract, SocialLIWC, SocialSelf.

Discussion and Conclusion

This brief analysis shows that the word meaning is highly variable among individual, but these differences in individual semantics are systematic.

First, the main plane of PCA typically reflects the opposition between cognition (social) and sensory domains, which reflect basic "language vs sensory" distinction in the word meaning structure and – more broadly – two forms of knowledge representations in the human brain [39]. Plausibly, there are different ratios of the language/sensory components even in semantics of the same words in different people. The second plane in PCA for the most considered cue words typically reflects the distinctions between the negative and positive poles of the emotional component of the word meaning. This finding is in line with previously reported findings regarding the existence of emotional components in the meaning of all the words, not only emotional ones [40].

Second, we found the correlations between the emotional components of the word meaning and the components of the other domains, mostly sensory ones (i.e., EmoPleasant positively correlates with $r > 0.5$; $p < 0.005$ with VisIntens, VisFace, SomatSurface, that is brightness, beauty and warmness are positively associated with pleasantness; negative emotions (EmoSentiment) are strongly correlated with the high values of GustTaste (i.e., bitter taste), EmoAngry is associated with AudIntens, i.e., "angry" and "loud" are related, etc.). These findings are in line with the results of works which consider synesthesia in word semantics [41; 42]. It is interesting to note that the male pole of the opposition "Male – Female" (SocialGender) is positively correlated with negative emotions in general (EmoSentiment), EmoAngry, negatively with EmoHappy and SocialSelf.

Our research has also confirmed the effect of the emotional states and stable personality traits on individual meaning of words. This line of research has been actively developing, but controversial results are reported [6; 43]. Our results show the existence of the correlations between the values of particular semantic features and stable personality traits, but correlation coefficients are low (although correlation is significant, $p < 0.0001$): the highest ones are at the range of $0.1...0.2$ $(-0.2...-0.1)$. Extraversion is positively correlated with AudNorms, SomatProprioception, SocialLIWC, EmoHappy, negatively with GustTaste, EmoSentiment, Drive. Agreeableness is positively correlated with SocialLIWC, EmoSentiment, negatively with GustTaste. Contiousness positively correlates with VisIntens, AudNorms, AudLIWC, SomatSurface, TemporalLIWC, SocialSelf, SocialLIWC, EmoPleasant, EmoHappy, negatively with GustTaste, SocialGender, EmoSentiment, EmoAngry, EmoDisgust, EmoSurprised, Drive. Neurotism correlates positively with VisLIWC, MotorBinder. Openness positively correlates with AudNorms, MotorBinder, negatively with DriveNeeds. However, our detailed analysis revealed that specific words are more related to the informants' characteristics than others.

We revealed that the pausing behavior is highly individual and does not depend on the stimulus word. The number of cognitive pauses is positively correlated with

Neurotism, negatively with Extraversion, i.e., the number of cognitive pauses is related to stable personality characteristics (all the correlation coefficient values are low – at the range of 0.1…0.2 (–0.2…–0.1) – but significant, $p < 0.001$). The number of long pauses positively correlates with Interes, Joy, Astonishment, Shame, PositiveIntegral, i.e., the number of long pauses is related to emotional states of the informants.

Correlations between the number of pauses (both cognitive and long) and semantic features were found for some particular cue words. We were able to find associations between the number of the cognitive pauses and the cognitive components of meanings, long pauses and emotional components for particular words. This could be considered as a proxy for the individual differences in the word meaning: there are no words which result in the same pausing behavior.

Using the newly developed complex methodology which combines techniques and findings from distributional semantics, linear algebra, neurobiology of semantics, we have been able to show that the word meaning is highly variable among individuals, but these differences could be systematically described and classified. The proposed methodology could be easily applied for the analysis of not only the results of word association experiments but also of texts (including those from social media), contexts of the usage of the key words of any culture and so on to analyze the differences in the word meaning in people with various backgrounds which will facilitate the efficiency of human communication in different fields.

We highlight the usefulness of the proposed methodology for the field of foreign language teaching. L. Vygotsky claimed: "The child already possesses a system of meanings in the native language when he begins to learn a foreign language. This system of meanings is transferred to the foreign language"[9], and it is essential for the foreign language teachers to be able to explain the differences in the sociocultural and psychologically real meaning of the words in native and foreign languages.

The methodology proposed in this paper does not require any manual effort and is free from subjectivity which is typical for the body of research related to the problem of revealing individual differences in word meaning. It provides the results that could be easily replicated. Of course, the set of features currently implemented in our methodology, could be easily expanded using the existing psycholinguistic norms and could be updated as the new norms are created and new data from neurobiology about the word meaning structure is obtained.

Our nearest research plans are related to the close examination of the effects of part-of-speech, polysemy and concreteness/abstractness of the word on its semantic features using proposed methodology. We also plan to study the effect of different language models (including not only their type but also type of texts they are trained on) on proposed features as well as to expand our feature set. In addition, we aim to apply our methodology to analyze meanings of key words of Russian language and other world languages employing large language models replicating word association experiments.

REFERENCES

1. Lenci A. Distributional Models of Word Meaning. *Annual Review of Linguistics.* 2018;4:151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254

2. Boleda G. Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics*. 2020;6:213–234. https://doi.org/10.1146/annurev-linguistics-011619-030303

3. Stoltz D.S., Taylor M.A. Cultural Cartography with Word Embeddings. *Poetics*. 2021;88:101567. https://doi.org/10.1016/j.poetic.2021.101567

4. Lee M., Martin J.L. Coding, Counting and Cultural Cartography. *American Journal of Cultural Sociology.* 2015;3:1–33. https://doi.org/10.1057/ajcs.2014.13

[9] Vygotsky L.S. Vol. 1. Problems of General Psychology, Including the Volume Thinking and Speech (Cognition and Language: A Series in Psycholinguistics). P. 221.

5. Kozlowski A., Taddy M., Evans J.A. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review.* 2019;84(5):905–949. https://doi.org/10.1177/0003122419877135

6. Litvinova T., Panicheva P. Individual Differences in the Associative Meaning of a Word Through the Lens of the Language Model and Semantic Differential. *Research Result. Theoretical and Applied Linguistics.* 2024;10(1):61–93. (In Russ., abstract in Eng.) https://doi.org/10.18413/2313-8912-2024-10-1-0-5

7. Wengelin Å., Johansson V. Investigating Writing Processes with Keystroke Logging. In: Kruse O., Rapp C., Ansonet C.M., Benetos K., Cotos E., Devitt A., et al. (eds) Digital Writing Technologies in Higher Education. Cham: Springer; 2023. p. 405–420. https://doi.org/10.1007/978-3-031-36033-6_25

8. Torrance M., Rianne C. Methods for Studying the Writing Time-Course. *Reading and Writing.* 2024;37:239–251. https://doi.org/10.1007/s11145-023-10490-8

9. Vandermeulen N., Van Steendam E., De Maeyer S., Rijlaarsdam G. Writing Process Feedback Based on Keystroke Logging and Comparison with Exemplars: Effects on the Quality and Process of Synthesis Texts. *Written Communication.* 2023;40(1):90–144. https://doi.org/10.1177/07410883221127998

10. Ismail M.G., Salem M.A.-M., Abd El Ghany M.A., Aldakheel E.A., Abbas S. Outlier Detection for Keystroke Biometric User Authentication. *PeerJ Computer Science.* 2024;10:e2086. https://doi.org/10.7717/peerj-cs.2086

11. Acien A., Calcagno N., Burke K.M., Mondesire-Crump I., Holmes A.A., Mruthik S., et al. A Novel Digital Tool for Detection and Monitoring of Amyotrophic Lateral Sclerosis Motor Impairment and Progression via Keystroke Dynamics. *Scientific Reports.* 2024;14:16851. https://doi.org/10.1038/s41598-024-67940-8

12. Borj P.R., Bours P. Detecting Liars in Chats Using Keystroke Dynamics. In: Proceedings of the 2019 3rd International Conference on Biometric Engineering and Applications (ICBEA 2019). New York: Association for Computing Machinery; 2019. p. 1–6. https://doi.org/10.1145/3345336.3345337

13. MacNiven S., Tench R. Keystrokes: A Practical Exploration of Semantic Drift in Timed Word Association Tasks. *PLoS ONE.* 2024;19(7):e0305568. https://doi.org/10.1371/journal.pone.0305568

14. Utsumi A. Exploring What Is Encoded in Distributional Word Vectors: A Neurobiologically Motivated Analysis. *Cognitive Science.* 2020;44(6):e12844. https://doi.org/10.1111/cogs.12844

15. Johns B.T. Determining the Relativity of Word Meanings through the Construction of Individualized Models of Semantic Memory. *Cognitive Science.* 2024;48(2):e13413. https://doi.org/10.1111/cogs.13413

16. Thompson B., Roberts S.G., Lupyan G. Cultural Influences on Word Meanings Revealed through Large-Scale Semantic Alignment. *Nature Human Behaviour.* 2020;4:1029–1038. https://doi.org/10.1038/s41562-020-0924-8

17. Wang X., Bi Y. Idiosyncratic Tower of Babel: Individual Differences in Word-Meaning Representation Increase as Word Abstractness Increases. *Psychological Science.* 2021;32(10):1617–1635. https://doi.org/10.1177/09567976211003877

18. Johns B. Computing Word Meanings by Aggregating Individualized Distributional Models: Wisdom of the Crowds in Lexical Semantic Memory. *Cognitive Systems Research.* 2023;80:90–102. https://doi.org/10.1016/j.cogsys.2023.02.009

19. Li P., Schloss B., Follmer D.J. Speaking Two "Languages" in America: A Semantic Space Analysis of How Presidential Candidates and Their Supporters Represent Abstract Political Concepts Differently. *Behavior Research Methods.* 2017;49:1668–1685. https://doi.org/10.3758/s13428-017-0931-5

20. Diallo A., Fürnkranz J. Unsupervised Alignment of Distributional Word Embeddings. In: Bergmann R., Malburg L., Rodermund S.C., Timm I.J. (eds) KI 2022: Advances in Artificial Intelligence. Cham.: Springer; 2022. p. 60–74. https://doi.org/10.1007/978-3-031-15791-2_7

21. Stoltz D.S., Taylor M.A., Dudley J.S.K. A Tool Kit for Relation Induction in Text Analysis. *Sociological Methods and Research.* 2024. https://doi.org/10.1177/00491241241233242

22. Grand G., Blank I.A., Pereira F., Fedorenko E. Semantic Projection Recovers Rich Human Knowledge of Multiple Object Features from Word Embeddings. *Nature Human Behaviour.* 2022;6:975–987. https://doi.org/10.1038/s41562-022-01316-8

23. Binder J.R., Conant L.L., Humphries C.J., Fernandino L., Simons S.B., Aguilar M., et al. Toward a Brain-Based Componential Semantic Representation. *Cognitive Neuropsychology.* 2016;33(3–4):130–174. https://doi.org/10.1080/02643294.2016.1147426

24. Chersoni E., Santus E., Huang C., Lenci A. Decoding Word Embeddings with Brain-Based Semantic Features. *Computational Linguistics.* 2021;47(3):663–698. https://doi.org/10.1162/coli_a_00412

25. Aldridge M., Fontaine L., Bowen N., Smith T. A New Perspective on Word Association: How Keystroke Logging Informs Strength of Word Association. *WORD.* 2018;64(4):218–234. https://doi.org/10.1080/00437956.2018.1535365

26. Ivanouw J. Stimulus Affectivity of the Danish Word Association Test as Measured by Response Heterogeneity and Rasch Scaled Number of Prolonged Reaction Times. *Scandinavian Journal of Psychology*. 2006;47(1):51–59. https://doi.org/10.1111/j.1467-9450.2006.00492.x

27. Zagorovskaya O.V. Key Words of the Russian Culture in the Aspects of the New Realities of School Language Education. *Sovremennye problemy lingvistiki i metodiki prepodavaniya russkogo yazyka v vuze i shkole.* 2018;(28):73–78. (In Russ., abstract in Eng.) Available at: https://new-journal.ru/nomer/28-nomer/ (accessed 09.08.2024).

28. Hills T.T., Jones M.N., Todd P.M. Optimal Foraging in Semantic Memory. *Psychological Review.* 2012;119(2):431–440. https://doi.org/10.1037/a0027373

29. Litvinova T., Zavarzina V., Panicheva P., Lyubova S., Mamaev I. RuPersWordAssociation: A New Dataset to Study Individual Association Behavior. In: Proceedings of the International Conference "Internet and Modern Society" (IMS-2024) (in press).

30. Stoltz D.S., Taylor M.A. text2map: R Tools for Text Matrices. *Journal of Open Source Software*. 2022;7(72):3741. https://doi.org/10.21105/joss.03741

31. Taylor M.A., Stoltz D.S. Integrating Semantic Directions with Concept Mover's Distance to Measure Binary Concept Engagement. *Journal of Computational Social Science*. 2021;4:231–242. https://doi.org/10.1007/s42001-020-00075-8

32. Stoltz D.S., Taylor M.A. Concept Mover's Distance: Measuring Concept Engagement via Word Embeddings in Texts. *Journal of Computational Social Science*. 2019;2:293–313. https://doi.org/10.1007/s42001-019-00048-6

33. Miklashevsky A. Perceptual Experience Norms for 506 Russian Nouns: Modality Rating. Spatial Localization, Manipulability, Imageability and Other Variables. *Journal of Psycholinguistic Research*. 2018;47:641–661. https://doi.org/10.1007/s10936-017-9548-1

34. Panicheva P., Litvinova T. Matching LIWC with Russian Thesauri: An Exploratory Study. In: Filchenkov A., Kauttonen J., Pivovarova L. (eds) Artificial Intelligence and Natural Language. AINL 2020. Communications in Computer and Information Science. Cham.: Springer; 2020. p. 181–195. https://doi.org/10.1007/978-3-030-59082-6_14

35. Wengelin Å. Examining Pauses in Writing: Theory, Methods and Empirical Data. In: Sullivan K.P., Lindgren E. (eds) Computer Key-Stroke Logging and Writing. Leiden: Brill; 2006. Vol. 18. p. 107–130. https://doi.org/10.1163/9780080460932_008

36. Bates D., Mächler M., Bolker B., Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015;67(1):1–48. https://doi.org/10.18637/jss.v067.i01

37. Lê S., Josse J., Husson F. FactoMineR: A Package for Multivariate Analysis. *Journal of Statistical Software*. 2008;25(1):1–18. https://doi.org/10.18637/jss.v025.i01

38. Tomczak M., Tomczak E. The Need to Report Effect Size Estimates Revisited. An Overview of Some Recommended Measures of Effect Size. *Trends in Sport Sciences*. 2014;21(1):19–25. Available at: https://tss.awf.poznan.pl/The-need-to-report-effect-size-estimates-revisited-An-overview-of-some-recommended,188960,0,2.html (accessed 09.08.2024).

39. Wang X., Men W., Gao J., Caramazza A., Bi Y. Two Forms of Knowledge Representations in the Human Brain. *Neuron.* 2020;107(2):383–393. https://doi.org/10.1016/j.neuron.2020.04.010

40. Warriner A.B., Kuperman V., Brysbaert M. Norms of Valence, Arousal, and Dominance for 13,915 English Lemmas. *Behavior Research Methods*. 2013;45:1191–1207. https://doi.org/10.3758/s13428-012-0314-x

41. Galac Á., Zayniev D. Paths of Linguistic Synesthesia across Cultures: A Lexical Analysis of Conventionalized Cross-Sensory Meaning Extensions in Europe and Central Asia. *Cognitive Linguistic Studies*. 2023;10(2):450–479. https://doi.org/10.1075/cogls.00108.gal

42. Mroczko-Wąsowicz A., Nikolić D. Semantic Mechanisms May be Responsible for Developing Synesthesia. *Frontiers in Human Neuroscience*. 2014;8:509. https://doi.org/10.3389/fnhum.2014.00509

43. Litvinova T.A., Zavarzina V.A., Kotlyarova E.S., Lyubova S.G. Mapping the Field of Word Association Research Using Text Mining Approach. In: ITCC '23: Proceedings of the 2023 5th International Conference on Information Technology and Computer Communications. New York: Association for Computing Machinery; 2023. p. 90–98. https://doi.org/10.1145/3606843.3606858

About the authors:

Tatiana A. Litvinova, Dr.Sci. (Philol.), Head of Research Laboratory of Psycholinguistic Textual Modelling, Professor of the Chair of Russian Language, Modern Russian and Foreign Literature, Voronezh State Pedagogical University (86 Lenin St., Voronezh 394043, Russian Federation), **ORCID: https://orcid.org/0000-0002-**

6019-3700, **Scopus ID:** **56638057700**, **Researcher ID:** **P-3809-2016**, **SPIN-code:** **3050-5653**, centr_rus_yaz@mail.ru

Olga V. Dekhnich, Cand.Sci. (Philol.), Associate Professor, Deputy Director for Science and Research of the Institute of Intercultural Communication and International Relations, Associate Professor of the Chair of English Philology and Cross-Cultural Communication, Belgorod State National Research University (85 Pobedy St., Belgorod 308015, Russian Federation), **ORCID:** **https://orcid.org/0000-0001-6088-2656**, **Scopus ID:** **56436702200**, **Researcher ID:** **AAM-9877-2020**, **SPIN-code:** **3426-6630**, dekhnich@bsuedu.ru

Об авторах:

Литвинова Татьяна Александровна, доктор филологических наук, заведующий научно-исследовательской лабораторией психолингвистического текстового моделирования, профессор кафедры русского языка, современной русской и зарубежной литературы Воронежского государственного педагогического университета (394043, Российская Федерация, г. Воронеж, ул. Ленина, д. 86), **ORCID:** **https://orcid.org/0000-0002-6019-3700**, **Scopus ID:** **56638057700**, **Researcher ID:** **P-3809-2016**, **SPIN-код:** **3050-5653**, centr_rus_yaz@mail.ru

Дехнич Ольга Витальевна, кандидат филологических наук, доцент, заместитель директора Института межкультурной коммуникации и международных отношений по научной деятельности, доцент кафедры английской филологии и межкультурной коммуникации Белгородского государственного национального исследовательского университета (308015, Российская Федерация, г. Белгород, ул. Победы, д. 85), **ORCID:** **https://orcid.org/0000-0001-6088-2656**, **Scopus ID:** **56436702200**, **Researcher ID:** **AAM-9877-2020**, **SPIN-код:** **3426-6630**, dekhnich@bsuedu.ru