

<https://doi.org/10.52326/ic-ecco.2022/SEC.04>



An Interface for Phonosemantic Assessment of Russian Words

Tatiana Burlea¹, ORCID: 0000-0001-9454-4313

Victoria Bobicev¹, ORCID: 0000-0003-4450-3964

¹ Technical University of Moldova, bd. Ștefan cel Mare, 168, Chișinău, Republica Moldova, tatiana.burlea@iis.utm.md,
<http://lilu.fcim.utm.md/>, victoria.bobicev@ia.utm.md, <http://lilu.fcim.utm.md/>

Abstract— This paper describes an interface for the phonosemantic evaluation for Russian words. By phonosemantics we mean the subconscious emotional interpretation of the acoustic perception of words independent of their meaning. The given method is based on data about the emotional perception of sounds, obtained from a sample of respondents. The developed program evaluates the emotional characteristics of the words using data about the sounds. The interface allows users to evaluate phonosemantics for a given word as well as to introduce her/his personal perception of phonosemantics for individual sounds of Russian language.

Keywords—computational linguistics; natural language processing; phonosemantics; web application.

I. INTRODUCTION

Sound symbolism or phonosemantics is a branch of linguistics and refers to the idea that vocal sounds have meaning. In particular, sound symbolism is the idea that phonemes carry meaning in and of themselves.

Margaret Magnus, the author of a comprehensive book explaining phonosemantics to the lay reader [1] describes three types of sound symbol (first proposed by [2]):

Onomatopoeia is the imitation of a sound like roar or moo.

Clustering is an effect of the semantic association.

For example: *hacienda, hall, hangar, harem, haunt/s, haven, hearth, hive, hogan, hold, hole, hollow, home, host, hostel, hotel, house, hovel, hut, hutch* – all these words have similar meanings which refer to a concept “home” and begin with /h/. Hence, the first phoneme /h/ is in some way associated with the concept “home”.

Clustering is not entirely blind to reference, and hence has an element of arbitrariness.

True Iconism is the visceral effect of the sound on a person. True Iconism is completely predictable and completely blind to reference. It only directly affects our

understanding of what the word’s referent is like, the word’s connotation.

For example:

In the group of words:

flit, flitter, float, flutter, fly

the final /ur/ makes the movement repetitive, the short /i/ makes the movement quick and short.

In the words:

stamp, tramp, tump, tromp, step, stomp

a pre-final /m/ makes the contact with the ground heavy; a pre-vocalic /r/ makes the motion go forward, and so forth.

Webster’s dictionary defines phonosemantics as follows: “The study of the meaning and symbolism of vocal sounds.” This assumes that every sound can be perceived as pleasant or unpleasant, rounded or angular, warm or cold. Thus, for example, the English word “break” is seen as something sharp, brusque; the Romanian word “ou” (egg) as something rounded, “licurici” (firefly) is perceived as something small, pleasant and quick even for those who do not know the Romanian language.

[2] claims that “sound is not simply an imitation, but a sign, which reproduces a common quality of the sound and the object; to mark the object, language chooses the sounds which partly independently, partly in comparison with others produce the impression, which is similar to the listener as that produced by the object on the mind”.

In this paper phonosemantics is interpreted as the subconscious emotional perception of the sounds of the word independently of the actual meaning of the word. On the other hand, the meaning of the word is more important for its emotional perception, and the phonosemantic component emerges when the meaning of the word is unknown.

II. RELATED WORK

Phonosemantics is an intersection of phonetics, semantics, lexicology and psychology as theory of perception [3].

While the question of how things are named was studied since Ancient Times, some ancient philosophers thought that things are named by people's agreement and that there is no connection between the meaning and the "sounding" of the word. Plato, however, believed that the names for things were dependent on the features of the thing as well as on the features of the sounds.

The idea that the sounds of the language have their own separate semantics have been developed by Mikhail Lomonosov in its Rhetoric (1748). "According to Lomonosov, each sound has its own meaningful energy. But this energy in itself is not rational, but emotional. Moreover, it carries a clear sign of the irrational" [4]. However, these characteristics of sounds subconsciously affect the receiver emotionally.

[5] created an universal classification of signs that was an important step towards separating phonosemantics as an independent scientific discipline.

Phonosemantics became an independent branch of linguistics only in the middle of the 20th century by [6], the purpose of which is to study the relationship between sound and meaning in a word.

An important contribution to the development of phonosemantics have been made by Alexander Zhuravlev with his works of emotional perception of each sound of the language [7]. He developed a methodology of phonosemantic evaluation of words and entire texts based on this emotional perception of each sound in the word or text.

Online text evaluation system VAAL that used the above mentioned methodology has been developed and used for text content analysis.

[8] described an online interface for phonosemantic evaluation of Romanian words. The created application is mainly based on Zhuravlev methodology [7] and uses phonosemantic assessments of multiple Romanian speaking respondents collected through specific surveys.

Several works studied the similarities between phonosemantic perceptions in multiple languages.

In [9], the authors, classifying emotional words in Russian and English observed similarities of sounds in words used to convey similar meaning.

[10] by analyzing word lists covering nearly two-thirds of the world's languages, demonstrated that a considerable proportion of 100 basic vocabulary items carry strong associations with specific kinds of human speech sounds, occurring persistently across continents and linguistic lineages.

However, we considered that people speaking different languages associated sounds with slightly

different emotions and characteristics. Thus, our aim was to create a similar system for Russian language and to compare its assessments with Romanian application.

III. METHODOLOGY

The sound of the letters influences our impression of the sound of the whole word. Thus, we can create the dictionary of sounds, emotionally appreciate each sound and obtain so-called "phonosemantic aura" of the word as a sum of the emotional characteristics of the sounds that the word contains. The "phonosemantic aura" of the word represents the emotional perception of the word. It can be characterized by different methods.

We used the values proposed in (Журавлев & Павлюк, 1989), represented by 20 pairs of adjectives with opposite meanings describing sounds as "good" or "bad", "fast" or "slow", "big" or "small" etc. To check the answers we decided to use all the value variants. If the respondent chose the same values from similar pairs (such as "bad" and "evil") for a sound, then these answers probably reflect people's feelings about the given sound.

The respondents were asked to choose from the proposed values to mark the sounds of the Russian language. For the questionnaire, a web interface has been created with an HTML form and checkboxes (figure 1). Each respondent had to choose features for each sound. Five possibilities were presented for each pair of features: from the most right characteristic to the most left one. For example, for the first pair on Figure 1 (good-bad), the first blue dot indicates that the sound is very good, the second light blue indicates that the sound is somehow good, the middle gray dot indicates that the sound is neutral for this pair of characteristics; not bad, not good. Red dots indicate bad and very bad respectively. This selection is coded with a number from 1 to 5; 1 – very good, 5 – very bad.

Figure 1. Web interface for the questionnaire

IV. THE VOCABULARY OF RUSSIAN SOUNDS

Such interface repeated for each sound of Russian language. It should be mentioned that the number of sounds is much larger than the number of letters.

There are 33 letters in the Russian alphabet. Of these, 10 letters are vowels, 21 are consonants, a hard sign and a soft sign. There are 42 basic sounds in the Russian alphabet. Six of them are vowels, and thirty-six are consonants. The difference in the number of sounds and letters in Russian is due to the fact that some letters include two sounds. The list of letters and corresponding sounds is presented in the tables below.

TABLE I. RUSSIAN LETTERS AND CORRESPONDING SOUNDS

| letter | sound | letter | sound | letter | sound |
|--------|-------|--------|-------|--------|-------|
| а | а | б | б, б' | п | п, п' |
| е | йэ | в | в, в' | р | р, р' |
| ё | йо | г | г, г' | с | с, с' |
| и | и | д | д, д' | т | т, т' |
| о | о | ж | ж | ф | ф, ф' |
| у | у | з | з, з' | х | х, х' |
| ы | ы | к | к, к' | ц | ц |
| э | э | л | л, л' | ч | ч |
| ю | иу | м | м, м' | ш | ш |
| я | иа | н | н, н' | щ | щ |

The Russian alphabet has hard and soft consonants that is reflected in the table. A consonant becomes soft when specific vowel follows it, namely: *е, ё, и, ю, я*.

For each of these sounds "phonosemantic aura" has been assessed through the interface presented on Figure 1.

V. METODOLOGY OF WORD PHONOSEMANTICS EVALUATION

After obtaining the characteristics of the sounds, we can calculate the characteristics of the words by summing all the values for all its sounds. The algorithm used was taken from [7]. Apart from the values given in the algorithm, the so-called information weight of the sounds is used. There are various factors that influence the informational weight of the sounds, among which the accent and the position of the sound in the word are the most influential. Sound frequency is also essential; rarer sounds contain more information. Calculation is made for each pair of the characters independently.

$$f_i = \frac{\sum_{j=1, m} n_i^j}{m}, \quad (1)$$

where f_i is the average value of the phonosemantic evaluation of the sound for a pair of values; n_i^j is the

phonosemantic evaluation of sound i by respondent j for a pair of values (can be 1, 2, 3, 4, or 5); m is the number of respondents.

F is the phonosemantic perception of a word for a pair of values.

$$F = \frac{\sum_{i=1, n} f_i k_i}{\sum_{i=1, n} k_i}$$

k_i is the information weight coefficient for each sound in the word:

$$k_i = \frac{1}{P_i}$$

where i is the position of the sound in the word ($i=1, n$);

P_i is the global frequency of the sound in speech.

The algorithm:

input data:

a word;

calculation:

- the word is divided into letters;

for each letter:

transformation letter \rightarrow sound i ;

for each sound:

for each pair of features:

- the phonosemantic evaluation of

the sound is calculated $k_i \cdot f_i$

- the summary F is calculated for the whole word.

F is displayed.

VI. INTERFACE FOR DATA COLLECTION AND WORD PHONOSEMANTIC EVALUATION

The created application¹ consist of two interfaces: (1) for data collection; (2) for word phonosemantic evaluation.

The data collection interface contains forms presented on Figure 1. There are one form for each sound; each form contains 20 rows with pairs of characteristics. The respondent selects the dot to check on the base of her/his feelings for this particular sound. For example, for the sound 'r' one may feel that it is more "frightening" than "safe" or it is more "rough" than "smooth" and so on. The checked forms then are submitted and the introduced results added to the database. In this way opinions were collected from multiple respondents and the average phonosemantic perception of each sound has been calculated.

The word evaluation interface contain a text input element where an user can introduce any word that is

¹ <https://phonosemantic-analysis-of-words.000webhostapp.com/>

analysed with the above described algorithm and the result is presented for every pair of the characteristics. Figure 2 presents a screenshot of the word evaluation interface with an introduced word. Figure 3 presents an answer of the application, namely, the phonosemantic evaluation of the given word.

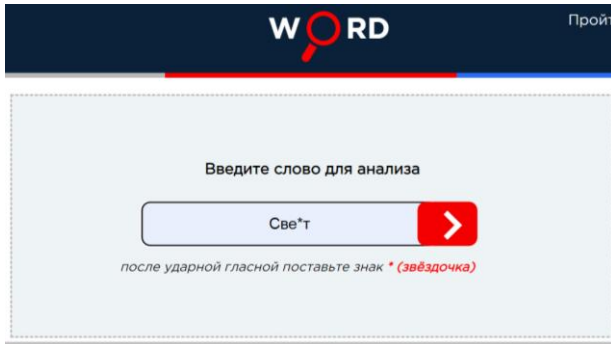


Figure 2. Web interface for the word phonosemantic evaluation

| | | |
|----------------------------|------|-------------|
| Горькое - Отталкивающее | 2.88 | Не выражено |
| Гладкое - Шероховатое | 2.32 | Гладкое |
| Легкое - Тяжелое | 2.43 | Легкое |
| Безопасное - Устрашающее | 2.5 | Безопасное |
| Величественное - Низменное | 3.18 | Не выражено |

Figure 3. Web interface with a fragment of the the word's phonosemantic evaluation result

If the calculated coefficient for a pair of characteristics is less than 2.5, the first characteristic of the pair is considered important for the evaluated word. If the coefficient is more than 3.5, the second characteristics is considered important. If the coefficient is between 2.5 and 3.5, no one of these two characteristics is important for the given word.

VII. ANALYSIS OF THE RESULTS

First, we analysed the average characteristics for the individual sounds. Figure 4 presents the results for the sound 'a'. It is seen in the diagram that the most silent features of this sound are "good", "big" and "strong".

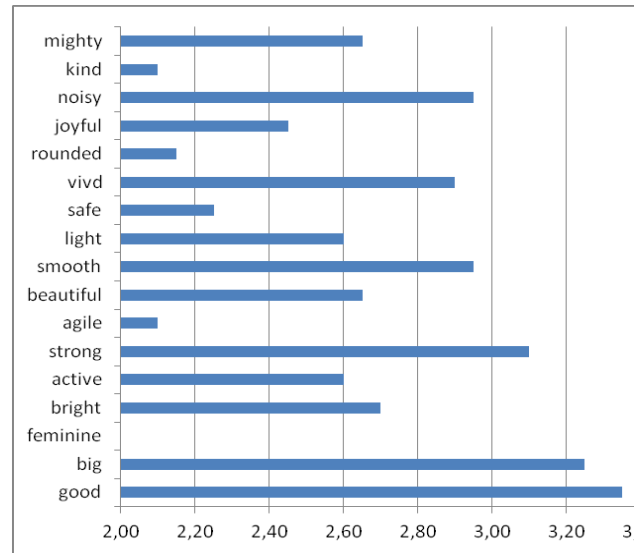


Figure 4. Phonosemantic evaluation result for the sound "a"

Figure 5 presents the results for the sound 'o'. The most silent feature of this sound is "noisy". Then it is "smooth", "vivid" and "strong". Less silent is the feature "active". There are no any other characteristics that are important for this sound.

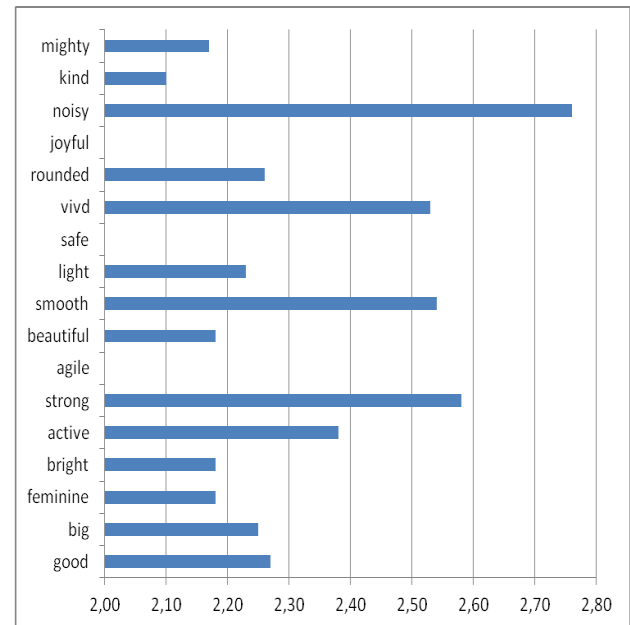


Figure 5. Phonosemantic evaluation result for the sound "o"

Little less silent are features "smooth", "vivid" and "noisy". Even less silent but still important are features "mighty", "beautiful" and "bright". Finally, somehow important are features "light" and "active". The other features were not selected by respondents as specific for this sound.

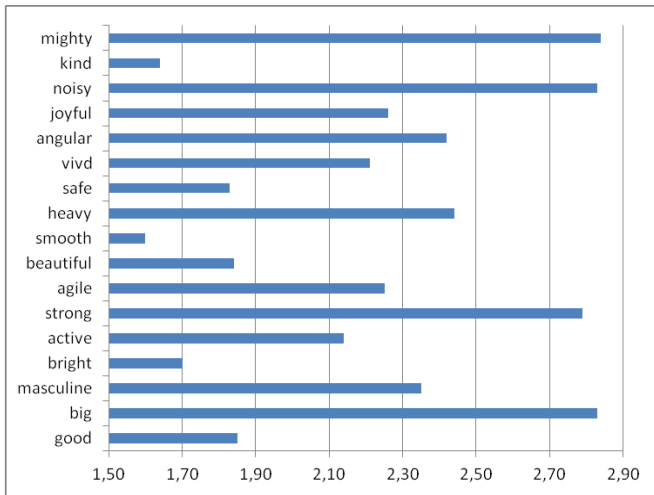


Figure 6. Phonosemantic evaluation result for the sound “r”

The examples on figures 4 and 5 are vowels. Consonants evoke other associations and features.

For example, Fig.6 presents characteristics for the sound ‘r’. In overall, it is perceived as ‘mighty’, ‘noisy’, ‘strong’ and ‘big’ that seems adequate.

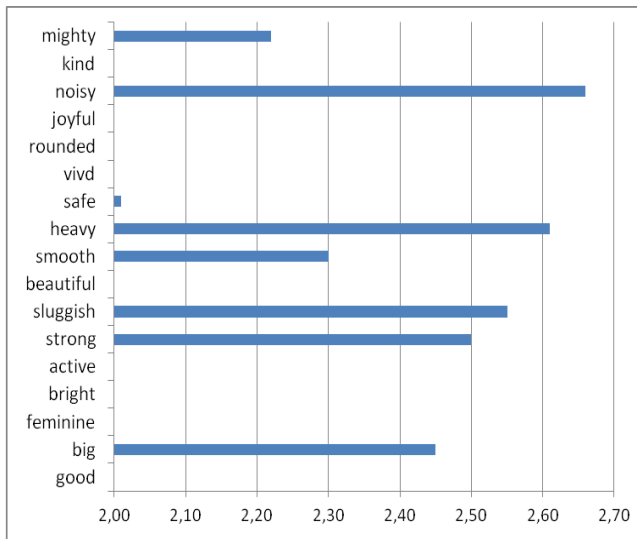


Figure 7. Phonosemantic evaluation result for the word “гром” (thunder)

Overall phonosemantic description of a word is formed as a sum of the characteristics of the sounds of the given word. Figures 7, 8 and 9 presents the examples of phonosemantic evaluation of words “гром” (thunder), “булка” (bun) and “цифра” (number).

Fig. 7 demonstrates the phonosemantic characteristics of the word “гром” (thunder). The most salient are the characteristics ‘noisy’, ‘heavy’, ‘sluggish’ and ‘strong’. In this case these characteristics correspond to the real word sense; however it is not always the case.

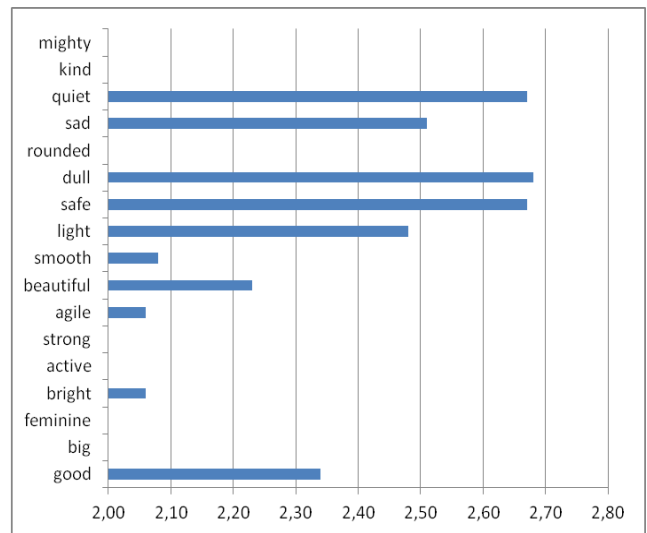


Figure 8. Phonosemantic evaluation result for the word “цифра” (number)

For example, Fig. 8 presents the phonosemantic characteristics of the word “цифра” (number). The sense in this case is abstract and does not evoke much emotional association. The sound of this word, however, evokes the emotional impression that it is ‘quiet’, ‘dull’, ‘safe’ and ‘sad’.

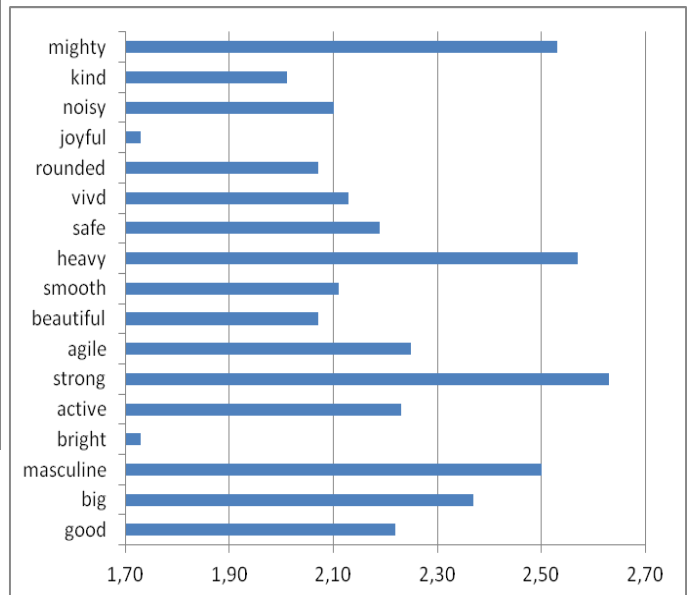


Figure 9. Phonosemantic evaluation result for the word “булка” (bun)

The last example on Fig. 9 is phonosemantic perception of the word “булка” (bun). It is perceived as “heavy”, “strong”, “mighty” and even “masculine”. We can easily notice that in this case, Russian and English words share the combination of sounds ‘b’ and ‘u’ in this

word and this combination ‘bu’ makes the word sound “heavy” and “strong”.

VIII. CONCLUSIONS

The paper presents an application with a web interface that evaluated phonosemantic perception of Russian words. The application has two interfaces: one is designed to collect users’ perceptions of the individual sounds of Russian language and in the other any user can evaluate phonosemantic perception of a whole word that is calculated as the combination of the phonosemantics of its sounds. In some cases sound of the word has something in common with its meaning but in some cases word sounds are perceived differently and are in contrast with its meaning. Phonosemantic characteristics of words can be used for new names evaluation in advertisement [11].

REFERENCES

- [1] Margaret Magnus, *The Gods of the Word: Archetypes in the Consonants*, Thomas Jefferson University Press, 2010.
- [2] Alexander von Humboldt, *Über die Verschiedenheit des Menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*, Dummler, Bonn, 1960.
- [3] Воронин С. В. *Основы фоносемантики*. М.: ЛЕНАНД, 2006.
- [4] Гукровский Г. А. Ломоносов-критик // *Литературное творчество М. В. Ломоносова: Исследования и материалы* / Под ред. П. Н. Беркова, И. З. Сермана. — М.; Л.: Изд-во АН СССР, 1962.
- [5] Пирс Ч. С. *Избранные философские произведения*. — М.: Логос, 2000.
- [6] С.В.Воронин, *Основы фоносемантики*. Изд. 3, репринтное, URSS, 2009.
- [7] Журавлев А.П., Павлюк Н.А. (1989). *Язык и компьютер*, Москва, Просвещение.
- [8] Bobicev V., Zidraşco T., *Estimating Word Phonosemantics*. In Proceedings of LREC 2008, Marrakech, Morocco, 2008.
- [9] Sokolova, M., Bobicev V. Classification of Emotion Words in Russian and Romanian Languages. RANLP 2009 conference, Borovets, Bulgaria, 2009.
- [10] Blasi, Damián E.; Wichmann, Søren; Hammarström, Harald; Stadler, Peter F.; Christiansen, Morten H. (27 September 2016). "Sound-meaning association biases evidenced across thousands of languages". Proceedings of the National Academy of Sciences. 113 (39): 10818–10823.
- [11] Klink, Richard R. (1 February 2000). "Creating Brand Names With Meaning: The Use of Sound Symbolism". *Marketing Letters*. 11 (1): 5–20.