Торайғыров университетінің
ҒЫЛЫМИ ЖУРНАЛЫ

НАУЧНЫЙ ЖУРНАЛ
Торайгыров университета

ТОРАЙҒЫРОВ УНИВЕРСИТЕТІНІҢ ХАБАРШЫСЫ

Филологиялық серия
1997 жылдан бастап шығады

ВЕСТНИК ТОРАЙГЫРОВ УНИВЕРСИТЕТА

Филологическая серия
Издается с 1997 года

ISSN 2710-3528

№ 2 (2023)

Павлодар

НАУЧНЫЙ ЖУРНАЛ
ТОРАЙГЫРОВ УНИВЕРСИТЕТА

Филологическая серия
выходит 4 раза в год

СВИДЕТЕЛЬСТВО
О постановке на переучет периодического печатного издания,
информационного агентства и сетевого издания
№ KZ30VPY00029268
выдано
Министерством информации и общественного развития
Республики Казахстан

Тематическая направленность
публикация материалов в области филологии

Подписной индекс – 76132

https://doi.org/10.48081/EIWC1999

SRSTI 16.41.21

https://doi.org/10.48081/MRWI7706

B. T. Kulbayeva[1], R. Zh. Saurbayev[2], *A. K. Zhetpisbay[3]
[1]A. Marghulan Pavlodar Pedagogical University,
Republic of Kazakhstan, Pavlodar;
[2]Toraighyrov University,
Republic of Kazakhstan, Pavlodar;
[3]A. Marghulan Pavlodar Pedagogical University,
Republic of Kazakhstan, Pavlodar
*e-mail: a_kanzhygaly@mail.ru

ON THE FEATURES OF STATISTICAL PROCESSING OF PHONETIC EXPERIMENTAL DATA IN LINGUISTIC RESEARCH

This article discusses the problem of using statistical methods for processing experimental data in the study of quantitative and qualitative properties of phonemes based on literary texts. The article studies the duration of vowels in the Russian language and its analogues in English in statistical terms. The duration of speech sounds is measured in thousandths of second - milliseconds (ms). In English, stressed vowels are longer than unstressed. The authors of the article note that it is not the statistical apparatus itself that presents the greatest difficulties. It turns out that it is most difficult for a phonetic researcher to see a statistical problem in his linguistic hypothesis, choose the correct statistical procedure, and then give an adequate interpretation of statistical calculations in terms of a linguistic problem. The authors regret that despite the rapid development of applied, statistical linguistics, there are no methodological developments yet, as in other applied fields. So far, the only way for a researcher is equal mastery of both linguistic problems and the apparatus of mathematical statistics.

In theoretical linguistics, linguistic research is usually divided into observation and experiment. The presented statistical methods apply mainly to observations. However, modern linguistics is becoming more and more experimental, especially in its field, which is called psycholinguistics.

Keywords: English, quantitative and qualitative linguistics, experimental phonetics, phonetic phenomenon, processing data.

Introduction

Experimental methods are also extremely developed in phonetic research. At the same time, there are no developments or manuals for processing linguistic experiments. The exception is a book published in a small edition by T. A. Brovchenko et al. [1], which certainly reflects the fact of the greatest «experimentality» of phonetics, at least in its Shcherbovsky version.

Practice, however, shows that the use of statistical methods is not a trivial procedure. The «collision» of the two sciences requires their adaptation to each other. So, on the one hand, the application of several criteria requires «fitting» the material to a certain framework. This is natural, since every statistical criterion model a well-defined «picture of the world». Therefore, if there is a desire to use this criterion, it is necessary that the experimental material exactly corresponds to the given «picture of the world», otherwise the obtained calculation results correspond to who knows what. However, the requirement of the correctness of the application of statistical procedures usually forces one to discard part of the material, thereby distorting the structure of the object.

On the other hand, linguistic material has the right to require the adaptation of the statistical apparatus and the development of specialized criteria and procedures. Some attempts are being made in this direction, but they are practically absent if we keep in mind the processing of data from linguistic experiments. Practice shows that the so-called standard software packages for processing data on computers are even less suitable for linguistic material. They take into account the specifics of the tasks and the material for which they are built and which is included in the «applied» experience of their creator. It is necessary to build a software package for solving linguistic problems properly, although one can doubt its universality in advance. Each time the task can be unique. And this means that it is necessary to adapt the statistical procedure to it, and not to adjust the task to the described canons.

Materials and methods

The use of statistical methods in the processing of data obtained as a result of an experimental study is an important component of the success of the experimental study itself. It is extremely important not only to obtain the data themselves but also to conduct a comparative analysis, the results of which will allow an objective assessment of their linguistic significance.

In the linguistic literature, unfortunately, little attention is paid to the explanation, interpretation and comparison of statistical methods themselves [2]. This seems to be a big omission since the description of statistical methods allows researchers to expand the tools when conducting an experimental study. In the Russian linguistic literature, several authors can be noted who have studied this problem. Among them should be mentioned the well-known work of Golovin B. N. «Language and statistics» [3]. The problems of statistical processing of experimental data were also dealt with Grabe E. and Low E. [4], Fletcher J. and Grabe E. [5], Potapova R. K., [6], Kanter L. A. [7] and others. In foreign linguistics, methods of statistical data processing are actively used in their research by E. Grabe, Post B. and Nolan F. [8], Siok Wai Ting and Fletcher J. [9], P. Warren [10], as well as Ladd R. [11], et al.

In this paper, when describing methods of statistical data processing, it seems appropriate to rely on the results of an experimental phonetic study [2]. The sound material used in the study is part of a database created within the framework of the IViE project, the authors of which are E. Grabe, B. Post, F. Nolan [8]. The sound material is represented by the speech of speakers of nine regional varieties of the British version of modern English: Newcastle, London, Leeds, Liverpool, Cambridge, Cardiff, Dublin, Bradford, and Belfast.

Taking into account the above, we will focus in more detail on those moments where the specificity of the material is manifested, and where statistical methods are not unambiguous.

Statistics studies a varying trait, and a trait exists insofar as its gradations exist. For example, in Russian, sentences can be of different lengths. This means that length is a varying feature, and one, two, three, etc. words in a sentence are gradations of the «length» feature. Another example is the place of a stressed syllable in a word. In Russian, it can be a word's first, second, etc. syllable. This is also a variable feature that statistics can deal with. In French, the stress always falls on the last syllable. For phonetists, it makes sense to talk about the place of the French accent, at least in comparison with other languages, but for statistics, this is not a sign, since there is only one gradation, i.e. there is no variation.

Probability theory is based on the concept of a random event (A). It converges with the basis of statistics if we mean by a random event the appearance of gradations of a feature. But then it should be said that statistics consider such events for which the strict inequality $0<P(A)<1$ is valid, where P is the probability of an event A.

Linguistic features can be quantitative and qualitative. So, according to V. Yngve [12] quantitative is the length of a sentence measured in words, letters, or any other units, as the depth of the phrase. There is also can be included in the number of morphemes in a word, the pitch of the main tone, or the intensity of the sound. A quantitative feature can be compared with any measure or measured in «pieces». Note that we are often interested in measuring not only the feature itself but also the frequency of gradations.

Linguistic features can also have a qualitative nature, for example, «syntactic construction». «Stress» – you can set 2 gradations – a stressed and unstressed vowel, and you can set, for example, 3 gradations – stressed, semi-stressed, «letter», «vowel phoneme», etc. It is clear that these signs themselves are not measured, it is possible to investigate only one of their characteristics – the frequency of occurrence of gradations of these signs. This question becomes more complicated when the researcher is interested in the gradation of a trait. For example, we are interested in sentences with direct word order and want to know the frequency of such sentences in speech. The question arises on which array of sentences to look for this frequency – whether to include interrogative and exclamation sentences, or incomplete, complex sentences here. It is a question of which attribute and with which gradations we study.

Special difficulties for processing arise when we are dealing with a branching feature. The sign «rhythmic structure of the word» is branching, since first of all it is the length of the word, and then the place of the stressed syllable. One can imagine that in the conditions of some tasks, all gradations of such a feature are lined up: monosyllabic: choreic, iambic, dactylic, amphibrachic, etc. However, for several objectives, such a system is illogical – primarily due to the inclusion of monosyllabic words here, strictly speaking, having no rhythmic structure. Such linguistic incidents force a researcher who applies the statistical apparatus to very clearly determine which feature and which gradation of the feature is being studied since the results and their interpretation depend significantly on this.

Intuitively, it is clear that if we want to make some kind of conclusion about the general population based on observations of the sample, then the sample from it should represent this general population quite well. In this case, statisticians speak about the quantitative and qualitative reliability of the sample. Strange as it may seem at first glance, the question of quantitative reliability is solved more simply. If there are preliminary ideas about the variation of the studied feature, then according to the appropriate formulas it is possible to calculate the necessary volume. It would ensure the specified accuracy of determining the studied parameter, or, in other words, it is possible to draw some kind of conclusion about the general parameters. Qualitative reliability is available when the sample structure repeats the structure of the general population. For example, we are interested in the length of a word in a newspaper. A novice researcher often does this: he takes 20 lines from editorials, political notes, sports news, etc., and on this combined sample gets the value of the average word length in the newspaper text. However, it is known that the headings in the newspaper do not occupy the same volume. Therefore, the result obtained in the described way is not reliable; the resulting parameter answers some other, but not the question posed. It would be necessary to make a sample so that the weight of each column in the newspaper was taken into account.

So, in the case when the structure of the general population is known, more precisely, the factors affecting the parameter under study are known, and the sample is built by these distributions. However, the structure of the general population is not always known. In this case, the sample is made according to a random law, i.e. with the help of any random number sensor. For example, if you need to find out the average length of a word in the novel «War and Peace», then you can use the random number table to select the pages that will be included in the selection.

Sometimes both ways are impossible. For example, you need to study the characteristics of vowels in normative pronunciation. These characteristics may vary from person to person. If it is not possible to increase the material so much that the group of speakers includes people of different ages, men and women with different speaking rates, different timbral features, etc., then they choose the principle of representativeness. This means that, based on linguistic considerations, a speaker is chosen – a typical representative of norms.

Results and discussions

The construction of the distribution is necessary to identify its shape. Statistical criteria do not apply to all types of distributions. The distribution must have a single vertex, i.e. be single-modal. If we have obtained two or more vertices, statisticians say that the observations belong to different general aggregates. Such a sample should be carefully analyzed and find a way to split it into two or more sets to obtain single-vertex distributions. Such distributions indicate the homogeneity of the studied feature.

So, the general totality in statistics is a set of objects organized not just by the studied attribute, but necessarily by a homogeneous attribute. And here there may be a contradiction between the linguistic and statistical understanding of the general population. Let us analyze the following example.

We study the duration of Russian vowels. Definitely, for a phonetist, this is a sign that is homogeneous at the phonemic level. The distribution of vowel durations turns out to be bimodal – with one vertex about 50–60 ms. On the other, it is – about 110–120 ms. The duration of speech sounds is measured in thousandths of second – milliseconds (ms). In English, stressed vowels are longer than unstressed ones. From the point of view of statistics, this feature is heterogeneous. It is not difficult to guess that unstressed vowels are grouped around the first mode and stressed vowels are grouped around the second. If we want to continue using statistical methods, we must comply with the requirements of mathematical statistics and consider the two general aggregates separately.

Let us note that the appearance of two or more modes in the construction of the distribution may also be a good device for establishing patterns. We might not have known anything about it before the experiment. It is not always a priori possible to assume heterogeneity of the material in the sense of the studied feature.

The question of the arithmetic mean, which is familiar to everyone from school, is not so simple. Let us note that the appearance of two or more modes in the construction of the distribution may also be a good device for establishing patterns. We might not have known anything about it before the experiment. Since it is not always a priori possible to assume heterogeneity of the material in the sense of the studied feature. The question of the arithmetic mean that is, the average duration of the vowels and consonants.:

Vowels	Consonants
$n1=40$, $\bar{x}1=90$ ms;	$n2=120$, $\bar{x}2=50$ ms.

where, ms – average value

The question is: what is the average duration of the sound? The average can be found in two ways. Method I:

$$\frac{\bar{x}1 + \bar{x}2}{2} = 70$$

This is the so-called unweighted average. Method II:

$$\frac{\bar{x}1n1 + \bar{x}2n2}{n1 + n2} = 60 \text{ мс.}$$

This is the weighted average. Which one is correct? Both averages are true, but each is for its case. The unweighted average is needed for the case when we want to find the average duration of a sound, regardless of how often these types of sounds occur in speech. It's like the average duration of a sound in a language system. The weighted average indicates the average duration of the sound in a given sounding text: it depends on how often long and short sounds occur in the text. This is the duration of the sound in speech. Thus, the types of arithmetic mean correlate with the opposition of speech and language.

Here is another example. A conversation between several people, i.e. a polylogue, is recorded on tape. Let four announcers participate in the polylogue. You need to find out the average rate of speech. We note right away that the question asked in this way is incorrect. What tempo of speech are we interested in? The average tempo of the speakers who participated in the polylogue? Then this is the unweighted average. The average tempo of the sound of this polylogue, then it is a weighted average.

In linguistic works, there is a mixture of the arithmetic mean and the median, i.e. the use of the first instead of the second. Recall that if the studied feature only seems quantitative, the median should be an indicator of the central trend. In principle, in some cases, replacing it with an arithmetic mean does not change the conclusions, but you should make sure of this by making several parallel calculations. Our experience shows that when calculating factor analysis according to expert estimates, in some cases the results for averages and medians do not differ, in others the conclusions turn out to be different.

When processing experimental data, the question often arises about the significance of certain answers, which usually boils down to an assessment of the deviation of empirical probability (p). Auditors are invited to listen to a vowel sound cut from a magnetic recording and identify it. The subjects are informed that there are 6 possible answers. Let as a result of the experiment we obtain, f.e. such reaction frequencies:

V	a	o	u	e	y	i
l (V)	0,30	0,19	0,25	0,10	0,16	0,00

We draw attention to the fact that the responses |a| and |u| have maximum frequencies. The non-randomness of each answer is determined by the Student's t-criterion. However, this requires setting a theoretical probability, i.e. evaluating the subjects' prior knowledge of the possibilities of the appearance of various stimuli. There are several ways to do this.

1) Since all the answers are entered immediately before the experiment, we assume that they are all equally possible. And then p | a | = p | u | =...= 0.17. In the conditions of this problem, the answers |a| and |u| that interest us turn out to be random.

2) Although 6 responses were introduced, the |i| reaction is practically impossible since the vowel stood after a solid consonant. This is confirmed by = | i | = 0. It can be thought that the subjects, after presenting them with several main stimuli, refine this alphabet and work in a field of 5 answers. Then a priori p = 0, 20, and only the answer | a | turns out to be non-random.

3) Despite the introduction of the alphabet of answers, they are not equally possible, and the subjects use the frequencies that they acquired in speech experience. Then p | a | = 0, 42, p | g | = 0, 20, and only the answer | a| turns out to be non-random.

Let us complicate the task a little now. Let us be interested in the fact that the vowel is broken in the answers, then the frequency of the ruined is 0.44. If we calculate the probabilities |o| and |u| in the three proposed ways, we get that in the first and third cases the answers of the ruined are not random, and in the second – within the limits of random. However, one more way of calculation can be assumed: it can be assumed that the subjects initially react to the coarseness of the vowel, and only then choose a specific implementation by one of the methods. Then at the first stage p (or.) = 0.5 (or another p). And then the answers in the sense of coarseness turn out to be random.

Then it should be concluded that the results of calculations (and, accordingly, conclusions about the properties of stimuli) will depend on which procedure we take as the initial one, i.e., strictly speaking, on some «extra-statistical moment».

Examining only the «peaks» of distributions, and «modes», we lose most of the information, because, with this type of analysis, all the answers that turn out to be random from the point of view of the statistical criterion remain «overboard». At the same time, two stimuli may have the same «peaks» in responses and differ in distributions in general «phonemic images of stimuli», according to L. A.Chistovich [13]. Therefore, it is necessary to apply criteria for comparing distributions.

However, the situation is complicated in cases where the alphabet of answers is not limited. In this case, however, as in the previous one, the Shannon estimation of the entropy of experience (H) can be successfully applied. It has the advantage of avoiding questions about the a priori distribution of reaction frequencies and the alphabet of answers. However, at the same time, we are moving away from the problem of the "quality" of reactions. As a result, the uncertainty of the experience, or the uncertainty of the stimulus in the perceptual experiment, is evaluated.

Researchers should be warned against possible errors in calculating the average entropy estimates. It is important to distinguish between two ways. In one case, you can get the entropy for each stimulus, and then average it for the entire group of stimuli – H. In another case, you can get the average phonemic image of a group of stimuli, and then calculate the entropy for it – $H\{\bar{p}\}$. These estimates do not match: $H<H\{\bar{p}\}$ The degree of discrepancy between the two H is greater, the more the phonemic representations vary within a group of stimuli; equality is possible only if the phonemic images of all the stimuli of the group coincide. Let us show this with examples:

Example 1

	A	o	u	I
S1	50	50		
S2	50	50		
S3	50	50		

Example 2

	a	O	U	e	i	Y
S1	50	50				
S2			50	50		
S3					50	50

In Example 1, the uncertainty of each stimulus S is 1 bit and $H\{\bar{p}\}$ =1. Average distribution of responses (R) – {50, 50, 0, 0, 0, 0} and $H\{\bar{p}\}$ = 1. In Example 2, the entropy of each stimulus is also equal to 1 and H=1. But the average distribution of incentives is {17, 17, 17, 17, 17, 17}. And then $H\{\bar{p}\}$ =2.58 bits. The qualitative interpretation of each of them should be taken into account H indicates the average uncertainty of a group of stimuli, and a $H\{\bar{p}\}$ – o indicates the uncertainty of the average stimulus from a certain group, taking into account the variation of phonemic images.

Note also that H, if there are many different stimuli, can be considered as a quantitative trait and, thus, investigated using distributions and other methods for quantitative traits. The solution to some problems of perceptual phonetics using this method is described in the work of L. V. Ignatkina, and Stern A. S. [14], and the study of the associative power of the Russian word is presented in the work of I. G. Ovchinnikova, and A. S. Stern [15].

When selecting verbal material, it turns out that there are few long high-frequency words in English or, for example, there are no monosyllabic verbs in German, etc. This, on the one hand, seems to distort the objective picture of the action of factors, but, on the other hand, reflects the specifics of the language as a system.

The spread of variance analysis in linguistic research unexpectedly encountered psychological «rejection» of the results on the part of some researchers, since the weights of the factors (η 2/x) are small. However, a small value of the correlation ratio, provided that it is significant, may correspond to the significance of the difference according to the Student's t –criterion at a 5 % level with a fairly small difference between the averages. However, the fact that the Student's criterion has been used in linguistics for a long time does not confuse researchers, but the corresponding small correlation relation is unusual. From the accumulated experience, it only follows that language is a very multifactorial and poorly formalized system.

The correlation coefficient can become a tool for finding decision-making units in which a person works when perceiving speech. In some cases, it is known that there must be a correlation between the sign and the results of perception. Thus, it has been widely confirmed by experiments of different modalities that probabilistic forecasting occurs during perception, i.e. the more frequent the stimulus, the better it is perceived. However, the «frequency» attribute itself is initially continuous. You can specify several ways to divide the frequency range into gradations. For example, in a linear scale, in a logarithmic scale, with more complex methods. The number of gradations may also be different. For each method, you can get your correlation coefficient with the frequency of correct identification. The maximum coefficient in this case will correspond to the optimal assignment of class intervals; optimal – in the sense of units of this attribute.

When working with correlation coefficients, the greatest number of questions arise on the interpretation of these coefficients: what is a large and what is a small correlation, etc. There is, say, the Guilford scale, where it is proposed to consider the correlation from 0.8 to 1.0 very high, and from 0.4 to 0.6 average. But these interpretations are proposed for psychological and pedagogical research. We should not forget about the specifics of linguistic features and language as a system. It may turn out that with a very large number of experiments, the same type of correlation does not exceed 0.7. Apparently, in this case, based on the nature of the trait, it is necessary to change the scale and assume that 0.6 – 0.7 is a very high correlation, and 0.4 – 0.6 is high, etc.

Conclusion

In conclusion, we would like to note that it is not the statistical apparatus itself or even the mutual adaptation of methods, material, and apparatus to each other that presents the greatest difficulties. It turns out that it is most difficult for a researcher to see a statistical problem in his linguistic hypothesis, choose the right statistical procedure, and then give an adequate interpretation of statistical calculations in terms of a linguistic problem. Unfortunately, there are no such methodological developments yet. So far, the only way for a researcher is equal mastery of both linguistic problems and the apparatus of mathematical statistics.

References

1 **Brovchenko, T. A., Varbanets, P. D., Taranets, V. G.** Method of statistical analysis in phonetic research. – Odessa, 1976. – 100 p.

2 **Nenasheva, Y. A.** Codification of the phonetic norm in the intonation contour of neutral utterances in modern English : experimental phonetic research: abstract of the dissertation of the Candidate of Philological Sciences. 2007. – 167 p.

3 **Golovin, B. N.** Language and statistics. – Moscow : 1971. – 160 p.

4 **Grabe, E., Low, E. L.** Durational. Variability in Speech and Rhythm Class Hypothesis. – 2002. – 7. – Mouton.

5 **Fletcher, J., Grabe, E.** Intonational variation in four dialects of English : the high rise tune. Prosodic typology. The Phonology of Intonation and Phrasing. Oxford, 2004. OUP. – P. 98–110.

6 **Potapova, R., Potapov, V.** Sentiment analysis of digital communication. Bulletin of the Moscow State Linguistic University. – Humanities. – 2023. – № 1.

7 **Kanter, L. A.** System analysis of speech intonation. – Moscow, 1988. – 128 p.

8 **Grabe, E., Post, B., Nolan, F.** Modelling intonational Variation in English. The IViE system. Proceedings of Prosody. Adam Mickeiwitz University, Poznan. 2001. – 207 p.

9 **Siok, Wai Ting, Fletcher, J.** The role of phonological awareness and visiual-orthographic skills in Chinese reading acquisition. Developmental Phycology, 2001. 37(6), p. 886–899.

10 **Warren, P.** Introducing Psycholinguistics. Cambridge University Press. 2012. – 265 p.

11 **Ladd, R.** Intonational phonology. – Cambridge : 1996. – 350 p.

12 **Ingve, A.** Model and hypothesis for language structure. Economics. // Proceedings of the Amer. Philosophical Society. – 1960. – № 104. – No.5.

13 **Chistovich, L. A.** Speech : Articulation and Perception. Department of Commerce, Clearinghouse for Federal Scientific and Technical Information // Joint Publications Research Service. – 1969. – 272 p.

14 **Ignatkina, L. V., Stern, A. S.** Uncertainty in the perception of vowels. Hearing and speech in norm and pathology. Issue 5. – Leningrad, 1984. – P. 3–10.

15 **Ovchinnikova, I. G., Stern, A. S.** Associative power of the Russian word. Psycholinguistic problems of phonetics and vocabulary. – Kalinin, 1989. P. 110 – 117.

Material received on 09.06.23.

*Б. Т. Кульбаева[1], Р. Ж. Саурбаев[2], *Ә. Қ. Жетпісбай[3]*
[1]Ә. Марғұлан ат. Павлодар педагогикалық университеті,
Қазақстан Республикасы, Павлодар қ.;
[2]Торайғыров университет,
Қазақстан Республикасы, Павлодар қ.;
[3]Ә. Марғұлан Павлодар педагогикалық университеті,
Қазақстан Республикасы, Павлодар қ.
Материал 09.06.23 баспаға түсті.

ЛИНГВИСТИКАЛЫҚ ЗЕРТТЕУЛЕРДЕГІ ФОНЕТИКАЛЫҚ ЭКСПЕРИМЕНТТІК ДЕРЕКТЕРДІ СТАТИСТИКАЛЫҚ ӨҢДЕУДІҢ ЕРЕКШЕЛІКТЕРІ ТУРАЛЫ

Бұл мақалада авторлар көркем мәтіндер негізінде фонемалардың сандық және сапалық қасиеттерін зерттеу кезінде эксперименттік деректерді өңдеудің статистикалық әдістерін қолдану мәселесін қарастырады. Мақалада орыс тіліндегі дауысты дыбыстардың ұзақтығы және оның ағылшын тіліндегі аналогтары статистикалық *түрде зерттеледі. Сөйлеу дыбыстарының ұзақтығы секундтың мыңнан бір бөлігімен – миллисекундпен (мс) өлшенеді. Ағылшын тілінде екпінді дауысты дыбыстар екпінсіз дауыстыларға қараганда ұзағырақ. Мақала авторлары ең үлкен қиындықтар статистикалық аппараттың өзі емес екенін атап өтті. Фонетика зерттеушісіне лингвистикалық гипотезадағы статистикалық мәселені көру, дұрыс статистикалық процедураны таңдау, содан кейін лингвистикалық мәселе тұрғысынан статистикалық есептеулерге барабар түсінік беру қиынға соғады екен. Қолданбалы, статистикалық лингвистиканың қарқынды дамуына қарамастан, басқа қолданбалы салалардағыдай әдістемелік әзірлемелер әлі жоқ. Әзірге зерттеушінің жалғыз жолы – лингвистикалық мәселелерді де, математикалық статистика аппаратын да тең меңгеру.*

Теориялық тіл білімінде лингвистикалық зерттеулер әдетте бақылау және эксперимент болып бөлінеді. Ұсынылған статистикалық әдістер негізінен бақылауларга қолданылады. Дегенмен, қазіргі лингвистика, әсіресе психолингвистика деп аталатын сала да дами келе эксперименталды болып келеді.

Кілтті сөздер: ағылшын тілі, сандық және сапалық лингвистика, эксперименттік фонетика, фонетикалық құбылыс, деректерді өңдеу.

*Б. Т. Кульбаева[1], Р. Ж. Саурбаев[2], *А. К. Жетписбай[3]*
[1]Павлодарский педагогический
университет имени А. Маргулана,
Республика Казахстан, г. Павлодар;
[2]Торайгыров университет,
Республика Казахстан, г. Павлодар;
[3]Павлодарский педагогический
университет имени А. Маргулана,
Республика Казахстан, г. Павлодар.
Материал поступил в редакцию 09.06.23.

ОБ ОСОБЕННОСТЯХ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ФОНЕТИЧЕСКИХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

В данной статье рассматривается проблема использования статистических методов обработки экспериментальных данных при изучении количественных и качественных свойств фонем на основе художественных текстов. В статье исследуется *длительность гласных в русском языке и ее аналогах в английском языке в статистическом выражении. Длительность речевых звуков измеряется в тысячных долях секунды – миллисекундах (мс). В английском языке ударные гласные длиннее безударных. Авторы статьи отмечают, что наибольшие трудности представляет не сам статистический аппарат. Исследователю фонетики труднее всего увидеть статистическую проблему в своей лингвистической гипотезе, выбрать правильную статистическую процедуру, а затем дать адекватную интерпретацию статистических вычислений в терминах лингвистической проблемы. Несмотря на стремительное развитие прикладной, статистической лингвистики, пока нет методологических разработок, как в других прикладных областях. Пока что единственный путь для исследователя - это равное владение как лингвистическими проблемами, так и аппаратом математической статистики.*

В теоретической лингвистике лингвистические исследования обычно делятся на наблюдение и эксперимент. Представленные статистические методы применимы в основном к наблюдениям. Однако современная лингвистика становится все более экспериментальной, особенно в своей области, которая называется психолингвистикой.

Ключевые слова: английский язык, количественная и качественная лингвистика, экспериментальная фонетика, фонетическое явление, обработка данных.

Теруге 09.06.2023 ж. жіберілді. Басуға 30.06.2023 ж. қол қойылды.
Электронды баспа
3,18 МБ RAM
Шартты баспа табағы 22,79. Таралымы 300 дана. Бағасы келісім бойынша.
Компьютерде беттеген: А. К. Темиргалинова
Корректорлар: А. Р. Омарова, Д. А. Кожас
Тапсырыс № 4100

Сдано в набор 09.06.2023 г. Подписано в печать 30.06.2023 г.
Электронное издание
3,18 МБ RAM
Усл. печ. л. 22,79. Тираж 300 экз. Цена договорная.
Компьютерная верстка: А. К. Темиргалинова
Корректоры: А. Р. Омарова, Д. А. Кожас
Заказ № 4100

«Toraighyrov University» баспасынан басылып шығарылған
Торайғыров университеті
140008, Павлодар қ., Ломов к., 64, 137 каб.

«Toraighyrov University» баспасы
Торайғыров университеті
140008, Павлодар қ., Ломов к., 64, 137 каб.
67-36-69
e-mail: kereku@tou.edu.kz
www.vestnik.tou.edu.kz