DOI 10.5281/zenodo.1212360 UDC 159.9.072

# The Problem of Sample Representativeness for Conducting Experimental and Broad Psychological Research

# Проблема репрезентативності вибірки при проведенні експериментальних та широких психолінгвістичних досліджень

**Iia Gordiienko-Mytrofanova** [1]
Dr. in Psychology,
Professor

**Ія Гордієнко-Митрофанова** [1]
доктор психологічних наук,
професор

E-mail: ikavgm@gmail.com
orcid.org/0000-0001-5332-0663

**Yevhen Pidchasov** [1]
Ph.D. in Psychology,
Associate Professor

**Євген Підчасов** [1]
кандидат психологічних наук,
доцент

E-mail: psyevgeniy@gmail.com
orcid.org/0000-0001-5057-4828

**Serhii Sauta** [1]
MA in Psychology

**Сергій Саута** [1]
магістр психології

E-mail: sergeisauta@gmail.com
orcid.org/0000-0002-1213-2655

**Iuliia Kobzieva** [1]
Postgraduate student

**Юлія Кобзєва** [1]
аспірантка

E-mail: yu.kobzeva88@gmail.com
orcid.org/0000-0001-8483-9695

[1] *H.S. Skovoroda Kharkiv National Pedagogical University, Ukraine*
✉ 2, Valentinovskaya Str., Kharkiv, 61000

[1] *Харківський національний педагогічний університет ім. Г.С. Сковороди, Україна*
✉ вул. Валентинівська, 2, Харків, 61000

*Original manuscript received November 10, 2017*
*Revised manuscript accepted January 20, 2018*

***ABSTRACT***

*The article is devoted to the problem of creating a representative sample of respondents in the course of experimental and broad psycholinguistic research, first of all, its quantitative composition and structure.*

*The primary method of research was the psycholinguistic experiment, the main stage of which is a free association experiment with «playfulness» as a stimulus word. The use of mathematical and statistical procedures confirmed the hypothesis and helped to achieve the goal.*

*The hypothesis of the experimental psycholinguistic research into «playfulness» stimulus as a stable personality trait was to prove the effectiveness of applying specific strategies to determine the representative quantitative composition of samples by means of comparing the frequency of the studied characteristics.*

*In terms of theoretical substantiation, the approaches to the determination of the quantitative composition of the experimental sample and the statistical calculations based on the results of the practical research into associative reactions to the stimulus «playfulness», it is proved that as the size of the sample decreases, the distinction grows, meaning that groups with fewer respondents do not reflect all the characteristics of the general population. The assumption was proved that the number of 100 people or close to that could not meet this requirement in the case of extensive research, the minimum size of the sample should be about 400–500 people, though this number is not always sufficient either. The sufficient quantitative composition of the experimental sample, in large general populations, varies from 400 to 1500 persons and depends on the quantitative and qualitative structure of the general population and the organizational peculiarities of the research. So, the sample should be enough to meet the requirement, which is to reflect the main tendencies and characteristics of both the general population and the purpose of the study.*

***Key words:*** *experimental research, psycholinguistic experiment, respondents, general population, representative sample, strategies and criteria of sample formation, the quantitative and qualitative composition of the sample.*

## Introduction

The composition and the nature of reactions inside associative fields bring forward an idea of the context of word usage, finding a meaning which is mentally relevant for the native speakers. Probably,

this is the reason why the question of the validity of associative experiments is so relevant and remains open until now.

How many respondents should be interviewed in the course of an associative experiment, so that we can talk about the validity of the results obtained?

As it has been shown in our review of psychological, psycholinguistic and sociological works, which all resort to free-association experiment (FAE) as a research tool, the quantitative characteristics of a sample, i.e. the sample size, which is used to explore the «associativity» of the stimulus, is still one of the most challenging and difficult-to-solve problems, both in methodological and in practical aspects. This problem, in its turn, is associated with a number of some other problems. In particular, it is the extremely insufficient use of broad statistical methods of varifying the hypotheses and the results of associative research, which would increase the reliability of the data obtained, and help to better outline the general population, as well as the sample, and provide the possibility of further transfer of the results over the entire population and making forecasts (Titova, 1975; Goroshko, 1997–2000; Karaulov, 2000).

G. Cherkasova wonders if it would be sufficient to rely on one hundred reactions (a sample which is most often used by researchers) in order to make conclusions on the associativity of the stimulus words, and she comes to the following conclusions summing up the quantitative studies of associative dictionaries: an increase in the number of respondents by 100 people that respond to a stimulus, increases on average by 36,41% the number of different reactions in the vocabulary entries of the direct vocabulary; «constant» reactions (no more than ten) have a relative frequency of occurrence of more than 5% in any sample (surves) of at least 100 respondents; whereas the increase in the sample size (survey) from 100 to 200 respondents causes relative frequencies of occurrence (ranks) of «constant» reactions to stabilize (Cherkasova, 2005).

O. Palkin, answering the question, «What is the bottom threshold, at which the researcher who works on the basis of associative technique has the right to speak about the reliability and completeness of the data obtained?» claims that «the bottom threshold is at the level of 100 respondents». A sample of 100 reactions per stimulus can be considered reliable and sufficiently representative provided that the sample is

correctly distributed among the general population. The competent distribution of the sample implies that the number of men and women acting as respondents should be approximately equal and the survey should be conducted in different places and embrace different segments of the population to ensure a variety of opinions and minimize the element of randomness (Palkin, 2008: 82). At the same time, the author concludes that the compilation of an «ideal» sample is a hardly feasible task and it is even theoretically difficult to model such a survey that would cover all segments of the population evenly. Moreover, as he observes, this task seems to be unrealistic for practical purposes (Palkin, 2008: 83).

At the same time, O. Palkin makes a curious remark that one of the leaders of the Moscow Psycholinguistic School, E. Tarasov, «when communicating with the doubters, had a habit of having along a pen and a piece of paper, and afterward, a rough graph appeared on paper. In that graph, up to 50, the increase of new reactions was significant, then up to the value of 100 the increase continued but it was not very significant, and after the value of 100, the growth, although it was present, it was at the negligible level. He further explained that the described tendency is repeated each time when processing reactions to the next stimulus. So, he concluded that 100 responses to each stimulus would be sufficient, but it would be better to obtain a little bit more» (Palkin, 2008: 82).

In our opinion, based on the sample of 100 people, it is impossible to obtain complete and non-distorted information about the features of the general population, since such a sample does not fully reflect the characteristics of the population under study unless the latter is considered as a rather narrow, homogeneous pool of respondents with clear, specific characteristics. But in this case, the general population already requires special selection, and there is a high probability of spoiling the naturalness of the population. Moreover, in the case described above, one can observe an attempt to approximate the general population to the characteristics of the sample rather than on the contrary, although it should be the «opposite».

However, there are some scholars who hold a different point of view, which we share. Thus, O. Goroshko, a recognized Ukrainian expert in the field of associative research, in one of her overviews on the application of quantitative methods to FAE data processing, notes

that the number of respondents sufficient to provide the given reliability is about 1,000 people (Goroshko, 2001).

O. Goroshko also notes that certain statistical laws have been established in the structure of an associative field with the help of quantitative analysis: ***the idea of the «golden section»*** and the existence of the connection between the number of reactions in an associative field and its ranks. The concept of the golden section is embodied in the proportions of numerical characteristics of the associative field (its richness and diversity), which reflect the stability of its structure as the values of language proficiency. The optimal quantitative composition of the field as the value of language proficiency one should consider the number of reactions within 400–500 units. In this case, the mentioned indicators (lexical richness and variety) approach the golden section value (Goroshko, 2001).

Yu. Karaulov believes that another statistical regularity in the structure of the associative field bears **the rank index of reaction frequencies**. As it was calculated, the number of ranks is directly related to the number of reactions. The larger the number of reactions is, the greater is the number of ranks in it. In this case, the described dependence is characterized by certain constancy: if the field is composed of 100 reactions, the number of ranks fluctuates within 5–8, the field of 500 reactions fits within 16–18 ranks, and the field which creates the associative norm (1,000 reactions), distributes its frequencies across 21–23 ranks (Karaulov, 2000).

I. Berezin mentions that the representative sample, which represents the whole population of Russia should consist of 3,600–9,000 people and 180 groups (two genders, three age groups, two educational levels, three income groups, and five types of settlements). The size of the sample depends on the number of parameters we want to achieve representativeness for. If we are satisfied with representativeness only by gender and age, then a sample of 400 people in one settlement will be more than enough. If we choose three parameters, the number of respondents should be increased to 600. Achieving the representativeness of the sample by five parameters simultaneously (gender, age, income, education, sphere of professional activity) is possible only with a sample of 1,000–1,200 people in one settlement (Berezin, 2012).

At the same time, O. Shmeliov states that the size of a representative sample cannot be determined a priori and is not reduced

to a predetermined constant (or a derivative from the size of the general population). All these attempts (to calculate something without carrying out a real survey) are characteristic of abstract mathematicians, who tend to ignore the unpredictability of natural processes (Shmeliov, 2016).

We fully share the opinion of scientists that the compilation of an «ideal» sample is hardly achievable. And it cannot be feasible if the general population is represented by an experimental sample, which is formed without any strategy, since it is impossible to reflect all its characteristics, especially through its careful analysis. Of course, if the general population is not considered as a separate category of respondents (a stratum), of a determined age and gender, and rather narrow in characteristics – by gender, age, educational background, involvement in the certain activity, etc.

In this respect, a question arises if there is a minimum number of respondents which would satisfy all associative experiments? The procedure of creating a sample in every research has its internal logic and strategy, which is defined by the goals, objectives, and hypothesis of the study, and, of course, by the features of the general population (population), including its volume.

**The aim of the present paper**: on the basis of applied psycholinguistic research to determine and to provide rationalization for the peculiarities of creating an experimental sample of respondents – its structure, quantitative characteristics, as well as certain qualitative aspects in the course of experimental and broad psycholinguistic research.

**The main objectives:** to identify key concepts and criteria that reflect the characteristics (and components) of creating experimental groups; to experimentally substantiate the peculiarities of forming the quantitative and qualitative composition of samples in the conducted psycholinguistic research; to present the main strategies and methods of determining and estimating the composition when creating representative experimental groups of respondents.

Therefore, in the course of the preparation of the study and selection of subjects, carrying out the experiment and presenting the results there are some methodological complications and stereotypes regarding the sample of the respondents, which usually indicate two most common cases. The first case is ignorance of methodological requirements regarding the characteristics of creating groups of respondents, and secondly, insufficient attention to the methodological

requirements for that matter. Both can be considered as significant errors, which respectively reduce the objectivity of the result, up to its inconsistency with the trends in the population.

As for the first case, it is advisable to pay attention to the methodology of scientific research within the framework of the modern paradigm, which provides sufficiently detailed requirements regarding the formation of the sample, including the most common artifacts and research errors. In the second case, it is necessary to emphasize the crucial importance of adherence to the basic requirements for the creation of an experimental sample for the possibility of further correct transfer of the results to a general population to be followed by the possibility to make predictions, which is usually the ultimate goal of the vast majority of studies, unless there is a goal to determine an essentially individual result.

Summarizing the main methodological mistakes in the formation of an experimental sample, one can distinguish the following:

– most frequently, the authors of the research do not use or describe the strategy of selecting respondents, which does not allow others to understand either the technique of research of the author or what was taken as the characteristics of the general population. In the best case scenario, there is a more or less detailed description of the group of respondents that slightly clarifies the situation, but there is no complete picture;

– it often happens that the number of respondents is accepted arbitrarily, and due to lack of information it is impossible to understand, why the author considers this number to be sufficient;

– it is often accepted arbitrarily that the sufficient number is 100 to 200 respondents, even in the study of large general populations (for example, when transferring the result to the national level or to all people of the given age);

– only a few authors analyze the quantitative and qualitative composition and usually do not explain the sufficiency of such quantity. Firstly, it violates the requirements to the description of the particular research technique and, in fact, it does not allow others to conduct a repeated research according to the author's criteria in order to check the assumptions and assertions that have been made on the basis of the obtained results;

– the authors do not introduce the concept of the general population, which makes it impossible to determine the population to

which the results have been transferred by the author, but it can be extremely important, especially in the case of narrowly focused and culturally colored researches;

– in the description of respondents who participated in the study one way or another, there are no distinctions between such concepts as the sample of representation, representative sample, the final number of research results; in fact, researchers seldom differentiate between a sample and the respondents studied in the analysis of the final results;

– besides, in the presence of control and experimental groups in the study, the way they have been formed is not always explained, but here the situation is somewhat better due to the fact that the selection of the experimental group is always explained by the logic of the particular research.

Unfortunately, in modern studies, presented by the authors in their scientific works, we observe a lot of methodological violations that undermine the objectivity of the results and do not allow us to achieve a complete understanding of the author's strategy. In this way they violate one of the basic conditions of scientific approach, i.e. the possibility to verify the results of any other researcher by means of following the author's methodology, which generally prevents any further scientific discussion.

In this paper we will not consider the peculiarities of key strategies of sample formation as they have been discussed in greater details in the methodological sources and they are not too complicated. But as far as the definition of the quantitative and qualitative composition is concerned, everything is a little more complicated, so let us try to illustrate the primary indicators and techniques we must rely upon while forming a group of respondents.

In the very beginning let us define *the key concepts* of our work, such as general population, sample representation, representative sample and the final number of results (respondents). The scope and purpose of the article do not allow us to carry out a detailed review of the basic concepts and approaches to these terms; therefore, we present their generalized interpretation by the authors of the study along with the main characteristics of these psychological phenomena. This material is based on a broad analysis of scientific papers and authors' experience of being practical scholars, including the authors of this work (Gordienko-Mytrofanova & Sauta, 2016).

*General population* – its definition is more or less clear – refers to all representatives of the population that can potentially participate in the study and which will subsequently be transferred to the results of a specific study. In this case, general population may be synonymous to the term population (Druzhinin, 2004; Atramentova & Utevska, 2007; Bochelyuk V.Y. & Bochelyuk, V.V., 2008; Maksimenko & Nosenko, 2008).

Problems begin on the next level, i.e., *a sample of representation*. The selection of representations is that part of the general population, which utterly reflects its characteristics, it is achievable to the researcher, and from which the selection of respondents to participate in the experiment will be carried out. In this case, it should be emphasized that the sample of representation is not identical to the general population, but reflects its main characteristics, equivalent to it for key characteristics (Druzhinin, 2004; Bochelyuk, V.Y. & Bochelyuk, V.V., 2008). But it is necessary to take into account the fact that the transfer of the result is carried out on the general population, and not on the sample of representation. Therefore, when calculating the quantitative and, if necessary, qualitative composition of the sample, it is necessary to take into account the characteristics and the volume of the general population, on which the result will be transferred.

*A representative sample* is a group of respondents that reflect the main characteristics of the general population (that is why it is so important to determine these characteristics), and these respondents will take part in the research in this or that way. It is the group that is commonly referred to as a sample (Yadov & Semenova, 1998; Druzhinin, 2004; Atramentova & Utevska, 2007; Bochelyuk, V.Y. & Bochelyuk, V.V., 2008; Maksimenko & Nosenko, 2008). It is important here to note two aspects: firstly, the quantitative composition must be sufficient to reflect the characteristics of the general population; secondly, it goes on about the selected subjects at the beginning of the work, and not about the final result. After all, any research process is dynamic, and its course and result may differ significantly from what was intended, although substantial deviations should be avoided as far as methodical procedures are concerned.

The next step is *to create a control and experimental group* within the representative sample, either at the beginning of the study or as a result of one of its stages. At the moment, it is not our purpose to

analyze the peculiarities of creating of these groups; we only note some of the points important for understanding the peculiarities of determining the quantitative and qualitative composition of the representative sample. These features include an understanding of the fact that under conditions of sufficient equivalence of the control and experimental groups, they represent a homogeneous general population. In the case of opposing them on the basis of a key feature (gender, education, profession, etc.), we end up with two separate general populations, and this should be taken into account when creating representative samples.

Turning back to the concept of a representative sample and its involvement into direct research, we note that, as it is mentioned above, the study of the phenomenon is dynamic and even for attempts to maximally control its course, it is impossible to avoid certain methodological deviations (artifacts). One of the sources of artifacts can be the dynamics of groups. Many researchers who developed the methodological foundations of the experiment focused on these facts, which they considered it necessary to take into account when analyzing the final results. For example, D. Campbell and D. Martin distinguish the following factors that influence the composition of the sample, such as experimental dropouts (mortality), history (background), selection (selection), statistical regression, (Campbell, 1963; Martin, 2004). In our case, the most important is the experimental dropouts as a detachment subjects from the group for one or another reason. It can only be a partial participation in the study (illness, refusal to continue work), non-compliance with the instructions, which leads to distorted results (and the result has to be discarded), prohibition to use its results, etc. (Druzhinin, 2004; Martin, 2004; Bochelyuk, V.Y. & Bochelyuk, V.V., 2008; Maksimenko & Nosenko, 2008).

All this leads to possible differences between the number of subjects included in the representative sample (sample) and the number of the subjects, and their results are taken into account in the final analysis, which becomes the basis for regularities and facts. And it is in this case we are talking about *the final number of results (respondents)*, and this index can both coincide with the number of respondents in the experimental sample, and significantly differ from the predetermined quantitative composition of the sample. Therefore, when forming a sample, it is essential to consider not only the features and the volume of the general population but also *the desired final number of results*.

Accordingly, when estimating the volume of a representative sample, in addition to taking into account the qualitative composition of the general population, the quantity can be calculated according to the following approaches:

1. Selection on the basis of the estimated quantitative composition of the general population, upon which the result will be transferred.
2. Selection on the basis of the well-grounded desirable number of final results that can be transferred upon the general population.
3. In both previous cases, the estimation should take into account statistical errors.

Regarding the qualitative composition of a group or a population, within the scope of this work, we are interested only in the definition of the population that is the conceptual basis of the study. Besides, we consider the possibility of dividing one general population into several smaller ones by the purpose and hypothesis of the research. The latter is observed in the case described at the beginning of this work (which pushed us to this analysis). Namely, inside the general population, there are several key characteristics specified, which form a homogeneous sample. But first, this method does not allow us to objectively represent the general population, since it is not homogeneous and is already a violation of methodological requirements (if we want to get a generalized result). So, it refers to the balancing of a limited sample of criteria, which requires an equivalent transfer of the result. Secondly, there is certain subjectivity in the selection of criteria. To avoid this, it is usually recommended to apply randomization as a way to level external variables by way of mixing. In addition, by means of defining some clear categories within a very limited sample, we invariably divide the general population into separate smaller sets. However, in this case, each of them will be represented by too few respondents.

## Methods and techniques of research

The main method of the conducted research is experimental, in particular, a psycholinguistic experiment, having been aimed at description the psycholinguistic meaning of the word «playfulness» as the most adequate and reliable model of systemic significance that reflects the reality of linguistic consciousness (Sternin, 2011: 188). The main stage of the research was the free association experiment (with the word-stimulus «playfulness») as the most elaborated technique of

semantic analysis. It is worth mentioning that the choice of this stimulus is not random. Over the last fifteen years it has been possible to observe a radical reorientation from the fragmentary mention of playfulness in psychological texts to the foundation of the latter in psychology as a stable personality trait (Guitard et al., 2005; Yarnal & Qian, 2011; Proyer, 2012–2014b; Shen et al., 2014; Waldman-Levi et al., 2015; Bar-Haim Erez et al., 2016; Yue et al., 2016).

As additional methods the surveys have been applied, with the aim of refinement of a free associative experiment results; questioning for the specification of the characteristics of the sample; the psychodiagnostic method, in particular, the methodology of the «Differential-Diagnostic Questionnaire (DDQ)» by E. Klimov for dividing the experimental sample into stratum. As a mathematical-statistical method, the analysis of the results of the study used frequency and cluster analysis, the φ* criterion (Fischer z-transformation), which allowed to identify tendencies in the distribution of associations of experimental groups and strata to achieve the intended goals.

The hypothesis of experimental psycholinguistic research of the «playfulness» stimulus being a stable personality trait is to substantiate the effectiveness of special strategies of determination of the representative quantitative composition of the samples by comparing the frequency of the studied characteristics.

The actual psycholinguistic tasks are aimed at showing that «playfulness» is a relevant lexeme in the linguistic consciousness of the subjects; gender, age and professional factors influence the stimulus of «playfulness» only on the far periphery.

In our study, we define general population as the inhabitants of Ukraine (aged 18 to 75), whose linguistic consciousness is characterized by knowledge (including understanding) of Russian. This is primarily because the association tasks have been presented in Russian, and it was also assumed that in the territory of modern post-Soviet Ukraine, the majority of the population either fluent in Russian or understands it sufficiently freely to produce associations. The answer was not limited and was not emphasized on the language of the answer, but the vast majority of respondents provided the answer in Russian regardless of the region, there was no refusal or an excuse on the misunderstanding of the task.

According to statistics as of 01.01.2015 (the State Statistics Service of Ukraine website): the main general population of adult

people, in total (from 18 to 75) is 35,145,000 people, 82% of the total population (42,759,700 people). This includes Russian-speaking, Ukrainian-speaking, bilingual, and those who understand Russian and able to express themselves in this language.

There are 460 cities, 885 urban-type settlements, as well as 28,385 other settlements (including settlements) in Ukraine. The country is divided into 24 oblasts. The urban population accounts for 62,13%, and the rural population is 37,87%. Ukrainians make up 77,8% of the population of Ukraine, and Russians – 17,3%. Belarusians, Moldovans, Crimean Tatars, Bulgarians, Hungarians, Romanians, Poles, Jews, Armenians, Greeks, Tatars and other ethnic groups, who were also involved in the research, live on the territory of the country.

Besides, according to the hypothesis of the study, further analysis of the results was foreseen for such indicators as age, gender, profession. This condition led to the division of the main general population into smaller ones according to the following key criteria.

Gender criterion: male – 49,09%, female – 50,91% of Ukraine's population.

Age criterion – youth (from 18 to 35 yo) is estimated to be 10,448,900 people; maturity (from 36 to 60 yo) – 15,366,100 people. The elderly (from 61 to 75 yo) – approximately 9,330,400 people – are not presented in this research as our work with this as a group is still in progress.

Here we consider it expedient to explain why at this stage of the study we have combined youth (17/18–21) and young age (21/22–35) into one age group. Systematic studies of the intellectual functions of the adults into the age range 18–35 years, conducted under the guidance of B. Ananiev, showed that the most considerable changes occur in short-term verbal memory, dealing with visual and auditory modality. The highest development rates are from the age of 18 to 30 years; then there is a slight decline. In turn, figurative memory is subjected to less change with age, and verbal impressions of long-term memory are characterized by a high continuity of indicators at the age of 18–35 years (Ananiev, 2001).

Among respondents, the key social categories (strata) were widely represented: in marital status (married, not married, divorced, cohabitation with a partner, engaged, widowed, having relationships, etc.). By level of education (with higher education, with incomplete higher education,

secondary education, etc.). By the level of professional activity (students, employed, unemployed, retired, etc.).

Taking into account this data and the fact that smaller general aggregates (strata) will be singled out in the course of the analysis, the selection of the subjects in the sample has been carried out both quantitatively and qualitatively. The main strategy of selection is a random selection (randomization), stratification with quantitative balancing.

*Sample 1* reflects the characteristics of the main general population. The purpose is to identify general trends in associations. At the initial stage of the study, all 24 oblasts of Ukraine were covered, as well as all types of settlements, together with 40 cities. It is this available part of the main general population that constituted *the sample of representation*. Investigators presented the subjects with word-stimulus «playfulness» and suggested to express the first (five) words that came to mind. The task was offered in the visual modality. According to the results of the activity, there were 2,902 respondents in Sample 1. General parameters of the sample: males – 1,187 persons (40,9%), age from 18 to 35 – 755, from 36 to 60 – 412, over 61 – 20; female – 1,715 people (59.1%), age from 18 to 35 – 1,310, from 36 to 60 – 381, over 61 – 24, according to characteristics close to the general population.

When the object of research or the general population is sufficiently large (for example, the population of Ukraine, districts, cities, families, etc.) and is multileveled, the sample is also of a multi-stage nature, the process of the sample construction proceeds in several stages. At the initial and every further intermediate stage, the objects of representation (selection units) are first selected, and only in the final stage the units of observation are added (for example, the oblast – the district – the city – the quarter – the house – the family – the person – the object). The main disadvantage of a multi-stage sample is the fact that the larger the number of stages is, the greater is the magnitude of the error. And for the preservation of objectivity, it is necessary to divide it into such groups of units that will fully reproduce its heterogeneity and, accordingly, the general population.

As a rule, at the first stage of building a multi-stage sampling, the *stratification* of the general population is applied. The *strata* are inherently homogeneous parts of the totality that are different from each other and which completely cover the totality (Maksimenko & Nosenko: 103). For example, the territory is divided into political and

administrative regions, each of them – into settlements, of urban and rural types; the cities are divided by size, can be separated by age groups, social status, profession, etc. Stratification is appropriate to continue until the certain uniform distribution of the feature within each stratum is achieved. In fact, stratification is the procedure for the formation of several separate random subsamples from a certain population, which then are combined into one.

The number of units that are selected from each stratum can be determined in several ways:

1. From each stratum, some units are selected which is proportional to the volume of the stratum:

$$ni = n\ (Ni / N);\ (1)$$

where $n$ is the sample size, $N$ is the volume of the general population, $Ni$ is the volume of an i-stratum.

For example, in our case, the general population – adult age, in total (from 18 to 75) is 35,145,000 people, when stratified by the age criterion, first stratum – youth (from 18 to 35) is approximately – 10,448,900 people, the number of respondents in Sample 1 was 2,902 persons, the respondents selected from the first stratum should be 863 persons (in our study this stratum is 400 people).

2. The number of units in each sub-sample is proportional to the standard deviation $si$ of the i-stratum (this indicator is being calculated on the basis of the results of the study sample):

$$ni = n\ (Ni\ si / S\ Njsj);\ J = 1;\ (2)$$

where $n$ is the sample size, $Ni$ is the volume of the i-stratum, l is the number of groups.

3. The same number of units is selected from each stratum:

$$N1 = N2 = N3 = .... = Nk;\ (3)$$

In this case, the divergence in the number of different strata of the general population is ignored, but at the same time, there is an opportunity to sufficiently represent smaller strata in the sample. This very method, taking into account the width of the general population, was chosen to be the major one.

If at a certain stage of stratification uniformity is achieved in the distribution of a characteristic, which is interesting for the researcher. As a rule, the procedure of allocating the nests or series (serial sample)

is usually carried out in the future. Unlike the strata, the nests are similar to each other; they are relatively homogeneous by certain attributes (Cochran, 1974). As a rule, the series is the smallest strata (if they are similar to each other on a certain significant sign): family, profession, level of education, etc. Selected samples of the series (nests) are subject to continuous or selective observation. It should be noted that, under the condition of a constant examination of the selected nests, the homogeneity of objects within each of them too understate the dispersion of a feature in comparison with the general population (for example, pupils of a particular school class or workers of a certain collective and, moreover, members of one family, would have many standard features).

It is possible to make a selection of a certain nest (series) within a stratum in the following ways:

1. on a random basis.
2. on the principle of the typical nest.

On the last stage of the actual implementation of random or systematic selection, as a rule, there appears a task to select a particular respondent from a circle of potentially possible.

At the end of these calculations, it is necessary to focus attention on the fact that the use of multistage samples results in one of the paradoxes of experimental psychology – on the one hand, with the help of stratification we distinguish groups from the general population, but due to the homogeneity of the strata, we are moving away from the heterogeneity of the population. This leads to the fact that in the case of finding the distinctions in executions, we will no longer be able to transfer the result to the main population, and the transfer should be carried out by the characteristics of the strata. If we want to analyze the peculiarities of the general population, we will have to return to Sample 1 (2,902 persons) that underwent stratified sampling.

Thus, using stratification, we obtained a general multilevel population and a representative sample out of that and were already formed from its number of samples of respondents depending on the aim.

*Sample 2.* The purpose is to determine the influence of gender and age on the nature and composition of associations.

Stratification of the general population has been carried out, which caused the sample to become multistage.

The sample according to the criteria «gender» (males and females – included both age groups) and «age» (18–35 and 36–60) included 1,600 respondents: 400 people in each sub-group of respondents. Within the framework of our research, we reduced the sample size from the point of view that the response rates in the samples of 1,600 and 3,000 respondents do not statistically differ (experimental data is given below). 400 people have been selected for each stratum in accordance with the above described approach of stratification (formula 3) together with the actual absence of significant statistical distinctions between samples of 1,600 and 400 people on the verge of non-significance/significance (see Table 3). This fact of approximation to the appearance of distinctions between the totality (a sample of representation) is objective, because, as it was mentioned above, in shows greater homogeneity.

*Sample 3*. The purpose is to determine the influence of the profession on the nature and structure of associations.

The next stage of the research was to identify common and specific features of respondents belonging to different professional backgrounds (types of occupations): «person – nature», «person – person», «person – sign systems», «person – technology», «person – artistic image», which are defined on the basis of the methodology developed by I. Klimov and his Differential-Diagnostic Questionnaire (DDQ), within the strata of youth (group 18–35), which is a period when a person tries to find his or her self. The sample comprised 500 people: 100 people in each «type of occupation», males and females being equally represented. In this case, we used the next level of the multilevel population – the allocation of nests. Regarding the quantitative composition, the allocation of nests implies a high degree of homogeneity in comparison with the general population, so that the conformity is not emphasized, which is what was proved by the statistical authenticity of the distinctions (see Table 3).

## Results and discussions

Now, after having analyzed the quantitative and qualitative composition of the general population and the samples of our research, in order to prove the validity of our number and show those failures, that often remain unnoticed, we present several variants.

Unfortunately, in modern psychological studies, the topic of methods and strategies of quantitative (and to a certain extent qualitative) formation of experimental samples is not sufficiently revealed and filling

the gap is the purpose of this work. The orientation and affinity of this issue with similar in other scientific fields allowed us to rely not only on the work of psychologists, but also on the peculiarities of the disclosure of this issue in linguistics, sociology, political science, medical research, physiology, economics, and rely on mathematical and statistical developments. And it is this breadth of coverage of the problem that allows you to most effectively outline the main ways to solve it.

First, let us return to the concepts of large and small populations. *Sometimes the quantitative characteristics of the sample are being taken arbitrarily*, about 5–10% of its volume. Based on these notions, the research identifies the following indicative quantitative characteristics:

– the small general population of up to 2,000 people – a sample of minimum 50 people;

– the large general population of 5,000 people – a sample of minimum 200 people.

To demonstrate a certain validity of these data, give an example based on the most accessible and widespread group of the respondents: the pupils and the students. So often the researches carried out within the limits of one settlement with an average (city of regional significance) or large (city of regional significance, sometimes some small settlements) with a small population. In this case, we deal with a group of 3–10 schools within the settlement (Merefa, Kharkiv region, district center, with a population of about 25 thousand people – 5 schools and 2 lyceums; the city of Krasnograd, district center, population of about 23 thousand people, 5 secondary schools; one art school and one sports school) and more in big cities, but in the latter case a separate district (for example, not the largest district of the city of Kharkiv – Novobavarsky, with the population of about 110 thousand people, represented by 19 schools).

Coming back to quantitative indicators, regarding the composition of one school, if one speaks a single degree, for example, a junior school, it is usually represented by two parallels of 25 to 30 people from each form from 1 to 4, in total it is about 240 junior pupils. In the case of a separate school, there are 25 to 30 people (10%) in the sample. If we talk about the whole city (a territorial unit, a conglomeration) – this is in total about 1,000–2,500 thousand junior students – the sample can range from 50 to 200 people, which is already determined additionally, based on the purpose and hypothesis of the research.

It is more complicated than with large populations – large cities, oblast, country because even at lower percentages, quantitative indicators of the sample shows significant growth. In this case, we observe a tendency, which is directly opposite to small communities – in which, for a small number of respondents, the result is closer to the individual. This is a tendency to an excessive number of respondents – even at the lower limit, a minimum sample size of 250 people, or even more. Therefore, it should be noted that this is a very rough, indicative calculation, for precise studies, which require a special calculation or statistical survey, is not desirable.

For accurate, methodologically grounded research it is better to use special formulas to estimate the quantitative composition of the sample. The calculation, as it was noted, can be carried out in two ways – from the general population and from the number of desired final results. And these failures require the introduction of additional concepts and indicators, such as the number of objects depends on the determined conditions. Among these conditions is the accuracy of the study, the required probability of forecast, the power of the criterion, and so on. There should be large enough objects within the sample to form a correct idea of the general population. The more objects are there, the less statistical error is, but after reaching a certain amount of further increase in the sample does not significantly affect the result (Atramentova & Utevska, 2007).

Generalized methodological and terminological tools for calculating a sample and sample errors have not been fully formed yet in the modern methodology of experimental research. And against the background of the well-established mathematical and statistical apparatus, one can find various terminological and symbolic designations of the same criteria.

In order to provide rationalization for our research, we used the most widely-spread parameters:

– *Percentage of responses and losses during registration.* Usually the acceptable volume of losses ranges between 5% and 20%, meaning that the volume of the experimental sample increases by 1,05–1,2 times;

– *Statistical power of research*. It is arbitrarily considered as 100% of the result minus error probability. The power is usually taken as 80–99,9%. It can't be lower than 80%, but if it is absolutely important to make sure that the research does not miss any probable effect, the power of research must be no less than 90%;

– *Statistical error*. As a rule, the level of significance is accepted as 5% and lower (the equivalent p=0,05);

– *Confidence interval* takes into consideration the statistical error. For example, if a certain feature was detected with 70% of respondents, the real position of this feature is within the range of 65–75%, meaning that the error is +/– 5%.

These are the key criteria, but there are a number of other criteria that enhance the reliability of data:

*Error of sample survey (accuracy)* – difference (deviation) between the values of the parameters in general population and its sample value (Yadov & Semenova, 1998);

*Share of feature (expected frequency of the result)* – expected share of feature, for which the error is estimated. If no information on the share of feature is available, it is necessary to use a value that equals 50% when the maximum error occurs.

*Standard deviation (dispersion)* – variability of observations;

*Minimal clinically important effect* – minimal changes and differences that we do not want to ignore.

Therefore, we intend to give ***examples of how a sample size can be calculated taking into account the main and additional criteria***. It should be noted separately that, in addition to universal approaches, the emphasis has been put on techniques that take into account the form of representation of data. In our case, this is the frequency (percentage) of qualitative data.

An approach similar to the one discussed above, when the researcher can rely on predecessors instead of going deep into calculations, is present in more precise methods of determining the quantitative composition of a sample.

1. For example, *Altman's Nomogram* is often used to determine the quantitative composition of the sample, which gives an approximate sample size depending on the power of the criterion, the level of significance and the value of the effect (Altman, 1991). In our case, according to its data, we get to the number of 100 dates (values, persons), as it was suggested in some studies, only in case of a significant reduction in the power of the criterion (80% and lower).

2. Rather approximate, though mathematically grounded data presented by sociologist V. Paniotto, *who proposed the formula for calculating the sample from the volume of the general population and the acceptable error (level of reliability)* (Paniotto, 1984).

Paniotto's formula has the following form:

$$\Delta = (\sqrt{\frac{1}{n} - \frac{1}{N}}) * 100\% \quad (4)$$

where Δ – sample error;
n – sample size;
N – the volume of the general population.

Using this formula, V. Paniotto proposed the following approximate ratio of the volume of the general population and that of the sample (Table 1) with an acceptable statistical error Δ at 5%.

**Table 1.** The Ratio of General Population and Sample

| The volume of general population | 500 | 1000 | 2000 | 3000 |
|---|---|---|---|---|
| The volume of sample | 222 | 286 | 333 | 350 |
| | | | | |
| The volume of general population | 4000 | 5000 | 10000 | 100000 |
| The volume of sample | 360 | 370 | 385 | 400 |

As we can see from the table, Paniotto's calculations confirm the insufficiency of 100 people for large general populations, and this data is most closely related to the number of respondents we have chosen (according to Paniotto, a large population is the one that has more than 5,000 people). This pattern persists and if we pay attention to Sample 2, where individual sample groups (strata) are formed from the main general population and are smaller than the main general population, but they are still related to large general populations due to their size – in our study, they include 400 people.

The calculation by the formula based on our general population in these groups of subjects also gives the required 400 people, so this number of respondents is minimal and sufficient, and a further increase in the quantitative parameters of the sample is not essential. And the level of reliability of 5% is a minimum threshold for humanitarian areas. First of all, it indicates statistically significant reliability, and secondly, increasing the level of reliability even up to 1% increases the sample exponentially, complicating the research.

*The essential decrease of the quantitative indicators of the sample can lead to the following consequences:*

– it can affect the objectivity of the results due to narrowing dispersion;

– it can violate the representativeness of the sample in terms of general population and thus distort the result;

– it reduces the possibility to transfer the results, which changes the population represented by the sample;

– it calls for maximum justification for the selection of the subjects with maximum consideration of the most important and additional characteristics of the general population, which, in turn, requires a thorough analysis of the latter and complicates the selection even further.

It is also necessary to note that Paniotto's formula is not the only formula available for such calculations. There are other formulas as well, but they are usually adapted for quantitative data and take into account the scope of data, standard deviations, etc. In our case, we deal with the analysis of associations, which means that we deal mainly with qualitative analysis, so quantitative data are used only indirectly, as the frequency (quantity of observations) of qualitative entities.

Let us cite several examples of methods of calculation of the quantitative composition of the sample that meet our requirements to the form of data presentation and the hypothesis of our research.

3. *The formula for estimating sample size at one single frequency* (Bland, 2000):

$$n = 15.4 * p * (1-p) / W^2; \quad (5)$$

where n – the required sample size,
p – the expected frequency of the result,
W – the width of the confidence interval.

We select our data at the minimum levels: the controlled variable is the frequency of deviations in associative responses, p = 50% (0.5); W - +/- 5%, i.e. 10% (or 0.1). We calculate the sample size according to our data:

$$n = 15.4 * 0,5 * (1-0.5) / 0.12 = 385$$

Thus, to get the result within the statistical error 0,05 (5%), with the estimated frequency of observing productive associations being 50%, the size of the sample is 385 people. And given the possible 10% withdrawal from the group, we obtain the quantitative composition of the sample is 385 x 1.1 = 424 persons.

4. *In case if it is necessary to compare frequencies of two groups (nests) within a single sample*, the following method of calculation can be applied. But before that, *it is necessary to calculate the number of respondents for each group represented in the research*.

The following data is selected for the calculation: controlled variable – the frequency of compliance of the associations; the value of significant differences – 20% (0.2); the level of significance – 5%; power – 80%; for a two-way test (accepting or denying an alternative hypothesis).

*A formula for calculating the sample size when comparing two frequencies* (Bland M., 2000):

$$n = [Z\acute{a}+Z\beta]^2 \times [(p_1 \times (1-p_1) + (p_2 \times (1-p_2))] / [p_1-p_2]^2; \quad (6)$$

where n – the sample size for each group (the total sample size is twice as large),

$p_1$ – the first frequency, we select its index as 60% (0.60);

$p_2$ – the second frequency, we select it at the level of 40% (0.40);

$p_1$–$p_2$ = clinically important differences, chosen as 20% (0.2);

Zά – depends on the level of significance, determined by special tables (the table of critical values of Student's coefficient, t) – is 1.96 at the level of significance is 0.05;

Zβ – depends on the selected power (determined by the tables of critical values) – in our case, it is 0.84 with power being 80%.

As we process our data with the help of this formula, we obtain the following:

$$n = [1{,}96 + 0{,}84]\ 2 \times [(0{,}6 \times 0{,}4) + (0{,}4 \times 0.6)] / [0{,}2]\ 2 = 95$$

Thus, we receive the number of observations needed to be included in each group. Consider possible withdrawals of 10% – 95 x 1,1 = 105 people in the group. The total sample size will be twice as large, i.e., 210 people. It means that in this case it can be argued that a sample consisting of 210–230 persons will be sufficient to detect differences in the frequency of associations, with 80% power, 5% confidence, and 20% level of minimally important differences.

5. Let us show another formula for the case of a partial (share) representation of the results (percentage). *For that, let us determine the sample size on the basis of the estimated confidence interval* (Koichubekov, Sorokina, Mkhitaryan, 2014). The initial information necessary for the implementation of this approach is the magnitude of variation, which is believed to be inherent to the population; desired

accuracy; the level of reliability, which should correspond to the results of the conducted research.

The size of the sample is determined by the following formula:

$$n = \frac{z^2(pq)}{e^2}; \quad (7)$$

where n – sample size;

z – normalized deviation, which is determined according to the chosen level of significance (the table of critical values of Student's t-coefficient),

p – estimated frequency of variation for the sample, q = (100-p),

e – permissible error.

For the assumed minimal variation of 50%, the permissible error is 5%, z – for the power of 80% – 1,29, and for the power of 95% – 1.96, we get the sample size of 166 respondents at the power of 80%, and 384 respondents if the power is 95%.

6. For the breadth of our analysis of the approaches to determining the quantitative composition of the sample, let us give some examples of such *calculations using modern Network Calculators* that use similar approaches and formulas, but also have a possibility to automate certain processes:

6.1. «Calculator» website (http://allcalc.ru/node/100) with minimum possible criteria (trustworthiness, confidence, reliability (power) – 85%; confidence interval, error – 5% and higher) for the chosen general population (young people aged 18–35, approximately 10,448,900 people), yields the result of 384 respondents for the necessary sample.

6.2. «Sociopolis» website (http://sociopolis.ua/ru/servisy/kalkulator-vybirky/) with minimum possible criteria (trustworthiness, confidence, reliability (power) – 95%, error – 0,05 and higher) for the chosen general population (young people aged 18–35, approximately 10,448,900 people) yields the result of 384 respondents for the necessary sample.

6.3. «Medical Statistics» website (http://medstatistic.ru/calculators/calcsize.ht) provides wider opportunities for considering various criteria and methods of their calculation (with due references to sources). Here we once again applied our selection of minimum possible limitations and rigidity of the criteria:

– level of precision – medium;
– significance rate – 0.05;
– variable A – 1.96;
– research power – 80%;
– variable B – 0.84;
– confidence interval – 2;
– permissible error – 5;
– the width of confidence interval – 10%;
– standardized variability $\frac{p_1 - p_2}{\sqrt{\bar{p}(100 - \bar{p})}} = 0.4$;

– minimum important difference between values (shares or means) – 20 %;
– type of sample formation – random;
– units for data presentation – percentage;
– expected frequency of the phenomenon in the experimental group – 60%;
– withdrawal rate in the experimental group – 40%;
– expected frequency of the phenomenon in the control group – 40%;
– withdrawal rate in the control group – 60%;
– general population – young people (aged 18–35), approximately 10,448,900 people.

Table 2 demonstrates the results of sample size calculation.

As we can see from the table 2, four out of seven available variants suggest using samples consisting of 400 and more respondents. Two variants deal with the size of a separate group within the sample as frequencies or strata are compared (see formula (6)). This leaves only one remaining variant that suggests the size of a sample consisting of 100 respondents, according to the values given the author in the table. So, as far as the problem of defining the size of the sample of the general population is considered, the analysis described above proves our assumptions about the minimal size of the experimental sample consisting of 400 respondents.

**Statistical analysis of the obtained results** is another empirical and objective proof that testifies not only to the above-mentioned assumptions but also to the feasibility of using this approach in the empirical psycholinguistic association experiment.

**Table 2.** Results of Sample Size Calculation Using Different Methods

| Method | Formula used for calculation | Result |
|---|---|---|
| K. Otdelnova's method | According to the table | 100 |
| Formula for repeated selection (one sample) | $n = \frac{t^2 \times p \times q}{\Delta^2}$ | 384 |
| Formula without repeated selection (one sample) | $n = \frac{N \times t^2 \times p \times q}{N \times \Delta^2 + t^2 \times p \times q}$ | 385 |
| Plokhinskiy's formula for comparing two groups | $n = \frac{t^2}{\Delta^2} \times (p_1 q_1 + p_2 q_2)$ | 768 |
| Lera formula for relative values (determines the size of each group which is compared) | $\frac{16}{(p_1 - p_2)/\sqrt{\bar{p}(100 - \bar{p})}}$ | 101 |
| Sample calculation formula where one single frequency is estimated | $n = \frac{15.4 \times p_1 \times (100 - p_1)}{W^2}$ | 371 |
| Sample calculation formula where two frequencies are estimated | $n = \frac{(A + B)^2 \times (p_1 \times (100 - p_1) + p_2 \times (100 - p_2))}{(p_1 - p_2)^2}$ | 95 |

In the first place, it is necessary to mention that the statistical analysis of empirical data obtained via association experiments is aggravated by the fact that this data is usually presented in a quantitative form, which means that they have to be converted into numbers in order to apply statistical coefficients. In this case, we can only use those coefficients that are compatible with statistical analysis based on frequency parameters. The number of these coefficients is rather limited, the most common out of which are: Pearson coefficient $\chi^2$, binomial criterion *m*, Kolmogorov-Smirnov criterion λ, and φ* criterion (Fisher z-transformation). As long as we evaluate the feasibility of applying these coefficients in our empirical research, let us analyze the peculiarities of representation and distribution of this data and the goal of our study considering the conditions of applying this or that coefficient (this analysis can be useful for other experiments as well). It should not be forgotten that the core idea of our empirical research was to show

that «playfulness» is a relevant lexeme in the linguistic consciousness of respondents, as well as to determine the influence of such factors as gender, age, and profession on the stimulus «playfulness». All this being considered, the general population and the sample appear to have a multi-level structure (randomized experimental sample (Sample 1) – stratified sample (Sample 2) – cluster sample (Sample 3)). The goal of this scientific paper was to provide evidence for the importance of determining the sufficient minimum of the sample size in experimental research and in this way to confirm the relevance of the chosen quantity of respondents in the sample groups used in our psycholinguistic association experiment.

The analysis of the peculiarities and structure of research, as well as the peculiarities of presenting the data, make it possible to choose a relevant statistical procedure (i.e., the coefficient). Thus, the binominal criterion (m) is used if research uses only one sample and the number of respondents does not exceed 300 people, which does not meet our requirements. Kolmogorov-Smirnov criterion λ assumes that the categories according to which the analysis is supposed to be made should be sorted out in ascending or descending order, and this codification cannot be random. This condition does not satisfy the peculiarities of presenting data in our research. In the first place, it does not have a precise distinction between different categories (levels), if one considers frequencies of reactions as categories. Secondly, the distribution of categories is a random value, so it would be wrong to speak about the accumulation of frequencies. As far as Pearson coefficient $\chi^2$ is concerned, it applies to our research, but it has certain limitations. Notably, in our study, it is necessary to consider the number of reactions rather than the number of respondents by way of the number of surveys. It is explained by the fact that a respondent gives several associative responses, and this very fact does not meet the main goal of the present paper, which is to provide rationalization for the selecting a particular number of respondents rather than reactions.

That leaves us only one statistical criterion for this case, which is *φ* criterion (Fisher z-transformation)*, which has minimal restrictions and is aimed at comparing samples according to the frequency of the observed effect, which makes this approach useful for our conditions. The only significant requirement here is to make sure that all reactions are reduced to the alternative scale «effect present – effect absent». In

our case, this problem is solved using dividing all reactions into high frequency and low-frequency reactions so that groups can be further compared according to these very criteria.

As long as we deal with social groups in our research, we decided to turn to the fundamentals of social psychology for the definition of the notion of a «group» to explain why certain reactions are considered to be high-frequency reactions. The minimal threshold of his notion is claimed to be 3–5 people. In this way, reactions, whose frequencies allowed us to define them as a group within certain parameters (starting from 3 people), can be considered as high-frequency reactions.

It would be worthwhile to mention that groups that are being compared, as far as the aim of the research is concerned (general population – a sample of representation – representative sample – stratified sample – cluster sample), can be perceived as nested one inside the other, as long as they were selected consecutively. It should also be noted that to prevent cluster homogeneity from influencing the results, randomization was used for all statistical calculations when creating strata and clusters (nests). This explains identical associative reactions and possible frequency distributions (high and low-frequency reactions) and allows us to compare these groups according to the ways this parameter is distributed. It means that the ratio of high frequency and low-frequency reactions were calculated within each of these key groups, and later these groups were statistically compared with each other with the help of φ* criterion. This helped to obtain data about statistical relevance (equivalence) of the groups with the possibility of transferring results from one group into another to confirm the relevance of reactions and their distribution, which also proves the representativeness of these groups. The results of statistical analysis are displayed in Table 3 and Figure 1.

When analyzing and comparing the data of table 3 about the declared issue, we see the sample of 100 respondents showing sufficient differences at the level of $p \leq 0{,}01$ with all other experimental groups. Therefore, this quantity of the respondents is absolutely insufficient for ensuring transfer of the results obtained to the general population. In other words, a nested sample, even on condition of monitoring the homogeneity factor does not reflect the characteristics of the general population, which entirely confirms the calculations above.

**Table 3.** Comparison of the Samples by the ratio of the Reactions Frequency (φ* criterion)

| Sample, Quantity | 100 Nested sample | 400 Stratified sample | 1600 Randomized sample | 3000 Sample of randomization |
|---|---|---|---|---|
| 100 | | 4.441** | 6.194** | 7.198** |
| 400 | 4.441** | | 1.754* | 2.758** |
| 1600 | 9.194** | 1.754* | | 1,004 |
| 3000 | 7.198** | 2.758** | 1,004 | |

Note: * – ambiguity area (p≤0,05), ** – relevance area (p≤0,01)

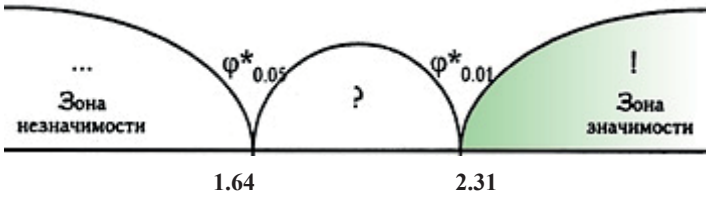


**Fig. 1.** *Relevance areas upon criterion φ**

Comparing the group of 400 respondents (stratified sample), we observe some distinctions from the randomized sample (1600 people) within ambiguity area (p≤0,05) at the edge of the non-relevance area. That is to say, showing specific differences from general randomized sample, that limits representativeness even on this level of correlation. It is also confirmed by above substantiations in differences of the strata formation and their influence on the result. As for correlation with the standard generalized sample, we have got insufficient distinctions, albeit at the edge of the non-relevance area. In this way, it is possible to state firmly, that even the sample level substantiated above of 400–500 people might be considered as a minimum, but it is not always sufficient to ensure representativeness of the sample, as well as a result in comparison with the general population.

Statistical analysis of the randomized experimental sample compliance (1,600 people) with the randomization sample (3,000 people) and the general population has not revealed substantial distinctions in the distribution of the results. We can safely say that the given quantity of people is sufficient to confirm representativeness of the respondents,

and the data as well; therefore the sample keeps the characteristics of the general sample.

It is possible to observe certain regularity: while reducing the sample level, distinctions increase, and groups of less quantity of respondents do not reflect all characteristics of the general population, even if the influence of homogeneity factor experience targeted prevention. A substantial level of quantity combination of the experimental sample in case of large general populations lies within an interval from 400 to 1500 people and depends on quantity and quality combination of general population, the goal, and features of the research arrangement.

Therefore, we have confirmed our assumption, that to obtain a representative result, the sample should be sufficient to meet the requirements on reflection of main characteristics of the general population. The quantity of 100 people or close to it, cannot meet the given demand, minimal quantity composition of the sample (but not always sufficient!), should not be less than 400–500 people. These data are confirmed by above theoretical substantiation, approaches to the calculation of quantity combination of the experimental sample, and first of all, by statistical estimates based on the results of practical research of the associative reactions.

## Conclusions

There is a rich variability of methods in the calculation of the sample quantitative combination, and one should choose the method depending on the list of the following research indicators: general population, the quantity of required final results, the hypothesis, the features of data representation, preciseness, level of representativeness, level of relevance, etc. But, in any case, despite of the calculation methods, the basic issue is that the sample should reflect the characteristics of the general population, and this requires the substantiation of the latter, and the presence of minimal limit, which imply that if the sample is less, then it is impossible to create a substantiated representative sample.

The procedure is complicated because of the absence of stable and agreed methodological criteria and requirements to the determination of the quantitative combination of the sample. Therefore, even attempts to get as much as close to the existing methodological approaches and purity of research, we couldn't entirely avoid specific deviations

in the formation of the representative samples. Since approaching to maximal representativeness violates the peculiarities of strata, formation of clusters and keeps us from the goal of our psycholinguistic study (definition of distinctions). Therefore, to form maximal representative sample one needs an accurate definition of the general population with all its characteristics, and this is exceedingly difficult to execute. Analysis of these possibilities is the perspective of our further experimental and methodological work in this direction. Also, there is a certain terminological confusion brought by a variety of terms and names applied to the same indicators, which is also should be agreed.

So, the issue about the volume of the general population is uncertain, unsolved and challenging to be solved. Some principles of the sample volume definition have been considered above, but a researcher, while choosing the sample volume, is influenced by some other factors, including the resources – time, finances, and from another side, one's wish to engage as many people in the survey, as it will be necessary for obtaining the maximally reliable information.

As it is mentioned above, the sample volume directly depends on width and homogeneity of the aggregate (population) under study. The less homogeneous the population in research is, the more peculiarities it has. In most cases, the members who constitute it, either group of people, or particular individuals always differ (by gender, age, education, profession or other distinguishing marks). Of course, there is no need to reflect within the sample all the qualities of the researched object, but it is necessary for the most relevant ones. The more informative and detailed the analysis of general population will be, the more qualities of the given object we take into consideration, the more massive should be the sample volume. It is because the respondents divide into subgroups, which can be compared on statistical basis only under the condition of their adequate quantity. Thus, the number of subgroups within the sample influences their volumes directly. But, at the same time, excessive detail of the marks can prevent the study of the large groups; transfer the vector of research to the category of branch-oriented, or even individual. This problem is solved directly by the author of the particular study about its goal, hypothesis, and tasks, which allows to ignore part of the characteristics of general population under the condition of substantiation and to observe the critical methodological requirements. And one of the essential stages, which mostly define the

qualities of the results obtained, is the formation of a representative sample of the respondents.

In many cases, it is reasonable to draw on the experience, by the researchers while carrying out sample surveys, and focus on it, depending on the scope and character of research. Thus, «typical» samples for national surveys vary within 1000 – 2500 respondents (depending on quantity of subgroups (strata), which is under analysis), the strata in frame of these researches and regional surveys – from 200 to 500 (while analyzing of numerous subgroups the volume of regional sample can increase to 1000 person and more), and nests within the strata defined in quantity from 100 to 200 person (Osipov, 2009).

Empirical research confirmed the given calculations with the application of statistical procedures. There is a regularity revealed that under the condition of decreasing volume of the sample the distinctions increase and the groups with fewer numbers of the respondent don't reflect all the characteristics of the general population. The assumption that quantity of 100 people or close to it, cannot meet this requirement, the minimal quantitative composition of the sample (not always sufficient!), should be not less than 400–500 people. The adequate level of the quantitative composition of the experimental sample is an interval from 400 to 1500 people and depends on the quantitative and qualitative composition of the general sample and the features of research. Therefore it should be sufficient to meet requirements in the reflection of the general population features, as well as the goal of research.

## References

Altman, D.G. (1991). *Practical Statistics for Medical Research*. London : Chapman and Hall.

Ananiev, B.G. (2001). *O problemakh sovremennogo chelovekoznaniia [On the problems of modern human knowledge]*. St. Petersburg : Peter [in Russian].

Atramentova, L.O., & Utevska, O.M. (2007). *Biometry. Part I. Kharakterystyka rozpodiliv [Characteristics of dispersions]*. Kharkiv : Ranok [in Ukrainian].

Bar-Haim Erez, A., Katz, N., & Waldman-Levi, A. (2016). Protective personality variables and their effect on well-being and participation in the elderly: a pilot study. *Healthy Aging Research*, *5*(12), 1–9. doi: 10.1097/01.HXR.0000511876.87334.cb

Berezin, I.S. (2012). *Marketingovyie issledovaniia. Instruktsiia po primeneniiu [Marketing research. Instructions for use]*. Moscow : Yurait-Izdat [in Russian].

Bland, M. (2000). *An Introduction to Medical Statistics* (3rd. ed.). Oxford University Press.

Bochelyuk, V.Y., & Bochelyuk, V.V. (2008). *Metodyka ta orhanizatsiia naukovykh doslidzhen z psykholohii [Methodology and organization of scientific research in psychology]*. Kyiv : Tsentr uchbovoi literatury [in Ukrainian].

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Boston : Houghton Mifflin Company.

Cherkasova, G.A. (2005). Kvantitativnyie issledovaniia assotsiativnykh slovarei [Quantitative studies of associative dictionaries], *Obshcheniie. Yazykovoie soznaniie. Mezhkulturnaia kommunikatsiia – Communication. Language consciousness. Intercultural communication* (pp. 227–244). RAS. Institute of Linguistics. Retrieved from http://it-claim.ru/Library/Articles/publications_Cherkasova_Galina/Cherkasova40/Cherkasova40.html [in Russian].

Cochran, W.G. (1976). *Metody vyborochnogo issledovaniia [Sampling Techniques]*. Moscow : Statistics [in Russian].

Druzhinin, V.N. (2004). *Eksperimentalnaia psikhologiya [Experimental psychology]*. St. Petersburg : Piter [in Russian].

Gordienko-Mytrofanova, I.V., & Sauta, S.L. (2016). Playfulness as a peculiar expression of sexual relationships (semantic interpretation of the results of the psycholinguistic experiment). *European Humanities Studies: State and Society*, *1,* 46–62.

Goroshko, Ye.I. (1997). Osobennosti muzhskoi i zhenskoi assotsiativnoi kartiny mira v russkom yazyke [Features of the male and female associative picture of the world in Russian]. *Wiener Slawistischer Almanach*, *40,* 203–246 [in Russian].

Goroshko, Ye.I. (1998). Spetsifika assotsiativnogo soznaniia nekotorykh grupp russkoyazychnogo naseleniia Ukrainy [Peculiarity of the associative consciousness of some groups of the Russian-speaking population of Ukraine], *Yazykovoie soznaniie: formirovaniie i funktsionirovaniie – Language consciousness: formation and functioning* (pp. 186–200). Moscow : Russian Academy of Sciences [in Russian].

Goroshko, Ye.I. (2000). Izucheniie verbalnykh assotsiatsyi na tsveta [The study of verbal associations on colors]. *Yazykovoie soznaniie i obraz mira – Language consciousness and the image of the world* (pp. 291–313). Moscow : Russian Academy of Sciences [in Russian].

Goroshko, Ye.I. (2001). *Integrativnaia model svobodnogo assotsiativnogo eksperimenta [Integrative model of the free associative experiment]*. Russian Academy of Sciences. Institute of Linguistics. Retrieved from http://www.textology.ru/article.aspx?aId=95 [in Russian].

Guitard, P., Ferland, F., & Dutil, É. (2005). Toward a better understanding of playfulness in adults. *OTJR: Occupation, Participation and Health, 25*, 9–22. doi: 10.1177/153944920502500103

Karaulov, Yu.N. (2000). Semanticheskii geshtalt assotsiativnogo polia i obrazy soznaniia [The semantic Gestalt of the associative field and the images of consciousness]. Proceedings from *XIII Mezhdunarodnyi simpozium po psikholingvistike i teorii kommunikatsii – The 13th International Symposium on Psycholinguistics and Communication Theory,* (pp. 107–108). Moscow: Editor Tarasov, Ye.F. [in Russian].

Koichubekov, B.K., Sorokina, M.A., & Mkhitaryan, K.E. (2014). Opredeleniye razmera vyborki pri planirovanii nauchnogo issledovaniia [Determination of the sample size in the planning of scientific research]. *Mezhdunarodnyi zhurnal prikladnykh i fundamentalnykh issledovanii. Meditsinskiie nauki – International Journal of Applied and Fundamental Research. Medical sciences*, *4*, 71–74 [in Russian].

Maksimenko, S.D., & Nosenko, E.L. (2008). *Eksperimentalnaia psikhologiya [Experimental psychology]*. Kyiv: Tsentr uchbovoi literatury [in Ukrainian].

Martin, D.W. (2004). *Psikhologicheskiie eksperimenty. Sekrety mekhanizmov psikhiki [Psychological experiments. Secrets of the mechanisms of the psyche]*. St. Petersburg: Prime-Evroznak [in Russian].

Osipov, G.V. (2009). *Rabochaia kniga sotsiologa [Handbook of the sociologist]*. Moscow: Librocom [in Russian].

Palkin, A.D. (2008). Assotsiativnyi eksperiment kak sposob krosskulturnogo issledovaniia obrazov soznaniia [Associative experiment as a way of cross-cultural study of images of consciousness]. *Voprosy psikhologii – Questions of psychology*, *4*, 81–89 [in Russian].

Paniotto, V.I. (1984). *Programmnoie obespecheniie prikladnykh sotsiologicheskikh i sotsialno-psikhologicheskikh issledovanii [Software of applied sociological and socio-psychological research]*. Kyiv: Znaniye [in Russian].

Proyer, R.T. (2014a). Perceived functions of playfulness in adults: Does it mobilize you at work, rest, and when being with others? *Revue Européenne de Psychologie Appliquée – European Review of Applied Psychology*, *64*(5), 241–250.

Proyer, R.T. (2012). Examining playfulness in adults: Testing its correlates with personality, positive psychological functioning, goal aspirations, and multimethodically assessed ingenuity. *Psychological Test and Assessment Modeling*, *54*(2), 103–127.

Proyer, R.T. (2013). The well-being of playful adults: Adult playfulness, subjective well-being, physical well-being, and the pursuit of enjoyable activities. *European Journal of Humour Research*, *1*(1) 84–98.

Proyer, R.T. (2014). Playfulness over the life-course and its relation to happiness: Results from an online survey. *Zeitschrift für Gerontologie und Geriatrie*, *6,* 508–512.

Shen, X., Chick, G., & Zinn, H. (2014). Playfulness in Adulthood as a Personality Trait. A Reconceptualization and a New Measurement. *Journal of Leisure Research*, *46*(1), 58–83.

Shmeliov, A.G. (2016). *Optimalnaia chislennost vyborki i psikhofiziologicheskiie metodiki [Optimal sample size and psychophysiological methods]*. Retrieved from http://forum.ht-line.ru/threads/optimalnaja-chislennost-vyborki-i-psixofiziologicheskie-metodiki.1568 [in Russian].

Sternin, I.A., & Rudakova, A.V. (2011). *Psikholingvisticheskoie znacheniie slova i yego opisaniie [Psycholinguistic meaning of the word and its description]*. Voronezh: Lambert [in Russian].

Titova, A.I. (1975). Ustoichivyie assotsiatsii leksem v nekotorykh slavianskikh yazykakh [Stable associations of lexemes in some Slavic languages]. *Candidate's thesis*. Minsk: BSU [in Russian].

Waldman-Levi, A., Bar-Haim Erez, A., & Katz, N. (2015). Healthy aging is reflected in well-being, participation, playfulness, and cognitive-emotional functioning. *Healthy Aging Research, 4*(8), 1–7.

*Website «Calculator».* Retrieved from http://allcalc.ru/node/100

*Website «Medical statistics».* Retrieved from http://medstatistic.ru/calculators/calcsize.html

*Website «Sotsiopolis».* Retrieved from http://sociopolis.ua/ru/servisy/kalkulator-vybirky

*Website of State Statistics Service of Ukraine.* Retrieved from http://www.ukrstat.gov.ua/

Yadov, V.A., & Semenova, V.V. (1998). *Strategiia sotsiologicheskogo issledovaniia. Opisaniie, obiasneniie, ponimaniie sotsialnoi realnosti [Strategy of sociological*

*research. Description, explanation, understanding of social reality]*. Moscow: Dobrosvet: Knizhnyi dom «Universitet» [in Russian].
Yarnal, C., & Qian, X. (2011). Older-adult playfulness: An innovative construct and measurement for healthy aging research. *American Journal of Play*, *4*(1), 52–79.
Yue, X.D., Leung, C.L., & Hiranandani, N.A. (2016). Adult Playfulness, Humor Styles, and Subjective Happiness. *Sage Journals: Psychological Reports, 119*(3), 630–640.

***АНОТАЦІЯ***

*Статтю присвячено проблематиці формування репрезентативної вибірки досліджуваних при проведенні експериментальних та широких психолінгвістичних досліджень, у першу чергу її кількісного складу та структури.*

*Основним методом дослідження виступив психолінгвістичний експеримент, головним етапом якого є вільний асоціативний експеримент зі словом-стимулом «грайливість». Підтвердження гіпотези й досягнення мети здійснено за допомогою математико-статистичних процедур.*

*Гіпотеза експериментального психолінгвістичного дослідження стимулу «грайливість» як стійкої особистісної властивості полягала в обґрунтуванні ефективності застосування спеціальних стратегій визначення репрезентативного кількісного складу вибірок через порівняння частоти досліджуваних характеристик.*

*За теоретичним обґрунтуванням, підходами щодо визначення кількісного складу експериментальної вибірки й статистичними розрахунками за результатами практичного дослідження асоціативних реакцій на стимул «грайливість», доведено, що при зменшенні обсягу вибірки розрізнення зростають і групи з меншою кількістю досліджуваних не відображують усіх характеристик генеральної сукупності. Підтвердилося припущення про те, що кількість у 100 осіб чи близько до того не може задовольнити цю вимогу при широких дослідженнях, мінімальний чисельний склад вибірки (але не завжди достатній) має бути близько 400–500 осіб. Достатній рівень кількісного складу експериментальної вибірки, за великих генеральних сукупностей, знаходиться в інтервалі від 400 до 1500 осіб і залежить від кількісного та якісного складу генеральної сукупності й особливостей організації дослідження. Тобто вибірка повинна бути достатньою для задоволення вимог у відображенні основних тенденцій та характеристик як генеральної сукупності, так і мети дослідження.*

***Ключові слова:*** *експериментальне дослідження, психолінгвістичний експеримент, досліджувані, генеральна сукупність, репрезентативна вибірка, стратегії та критерії формування вибірки, кількісний та якісний склад вибірки.*

***Гордиенко-Митрофанова Ия, Подчасов Евгений, Саута Сергей, Кобзева Юлия. Проблема репрезентативности выборки при проведении экспериментальных и широких психолингвистических исследований***

***АННОТАЦИЯ***

*Статья посвящена проблематике формирования репрезентативной выборки испытуемых при проведении экспериментальных и широких психолингвистических исследований, в первую очередь, ее количественного состава и структуры.*

*Основным методом исследования выступил психолингвистический эксперимент, главным этапом которого является свободный ассоциативный эксперимент со словом-стимулом «игривость». Подтверждение гипотезы и достижение цели осуществлено с помощью математико-статистических процедур.*

*Гипотеза экспериментального психолингвистического исследования стимула «игривость», как устойчивого личностного свойства, заключалась в обосновании эффективности применения специальных стратегий определения репрезентативного количественного состава выборок путем сравнения частоты исследуемых характеристик.*

*Теоретически обосновывая подходы к определению количественного состава экспериментальной выборки и применяя статистические расчеты по результатам практического исследования ассоциативных реакций на стимул «игривость», доказано, что при уменьшении объема выборки различия растут, и группы с меньшим количеством исследуемых не отображают всех характеристик генеральной совокупности. Подтвердилось предположение о том, что количество в 100 человек или около того не может удовлетворить данное требование при широких исследованиях, минимальный численный состав выборки (но не всегда достаточный), должен быть около 400-500 человек. Достаточный уровень количественного состава экспериментальной выборки, при больших генеральных совокупностях, находится в интервале от 400 до 1500 человек и зависит от количественного и качественного состава генеральной совокупности и особенностей организации исследования. Таким образом, выборка должна быть достаточной для удовлетворения требований в отражении основных тенденций и характеристик, как генеральной совокупности, так и цели исследования.*

***Ключевые слова:*** *экспериментальное исследование, психолингвистический эксперимент, испытуемые, генеральная совокупность, репрезентативная выборка, стратегии и критерии формирования выборки, количественный и качественный состав выборки.*

 © *Гордієнко-Митрофанова Ія, Підчасов Є., Саута С., Кобзєва Ю.*