DETERMINING THE SAMPLE IN THE LANGUAGE ORIENTED QUALITATIVE RESEARCHES

Abstract: Among other aspects, the quantitative researchers point out that defining the research sample is an extremely important question which contributes to the research validity and objectivity, thus raising dilemmas on specifying the sample in the qualitative researches related to language phenomena and questions. Therefore, this paper aims to answer the question: How can the sample be determined in the language oriented qualitative research, taking into consideration three different qualitative research scenarios developed by the authors, which, in fact, serve as a sample. The analysis of these scenarios exploits the general qualitative interpretative/inductive approach. The research results clearly show the complex process of choosing the sample, indicating that in the language qualitative researches the decision on the sample depends on the research question and the research goal, that the sample cannot be determined in advance due to the nature of the qualitative researches and that the sample directs the researches in many ways. Thus, determining the sample is an ongoing and constant process of feeding the research question with new information until the saturation is achieved and in the same time, while sampling, the researcher is, in fact, analyzing the data. This research has theoretical and practical implications as well, especially for those who are challenged to conduct the qualitative research on language issues. It provides them with knowledge that broadens their horizons and makes them aware of the qualitative researches’ relevance and, in particular, of the sample determination.

Key words: qualitative language research, sample.

1. Introduction

Worldwide a lot of researchers recognize the relevance not only of the quantitative researches but of the qualitative and mixed researches as well, pointing out the differences among them (Ackroyd & Hughes, 1992; Berg, 2001; Bernard, 1994; Brewer & Hunter, 1989; Burns, 2000; Bryman, 2012; Copland, Garton & Richards, 2010; Creswell, 2002; Creswell, 2009; Creswell & Clark, 2011; Creswell, 2014; Creswell, 2015; Creswell & Creswell, 2017; Croker, 2009; Denzin & Lincoln, 2000; Dörnyei, 2007; Freeman, 2009; Gilbert, 2008; Gorard 2001; Griffen, 2012; Herbst & Coldwell, 2004; Holliday, 2007; Hughes, 2015; Kumar, 2014; Long & Zhou, 2019; Mason, 2002;
Maxwell, 2005; Miles & Huberman, 1994; Onwuegbuzie & Leech, 2005; Patton, 2002; Raising, 2010; Rossman & Rallis, 2003; Sandelowski, 1998; Schoonenboom & Johnson, 2017; Silverman, 2000; Tashakkori & Teddlie, 1998; Yin, 2011 etc.). The main distinction between the quantitative and qualitative approaches, according to them, is the difference in the comprehension of the reality. They, also, highlight that the researcher’s choice of one of the mentioned paradigms depends, primarily on the research question and goal, i.e. whether he answers the question How many or the question How or Why.

From here, quantitative researchers, whose root is the positivism, believe that there is one fixed reality and that every truth find out for this reality can be measured and generalized and, therefore statistic methods for processing the data and reaching the conclusions are used, which contribute to the research objectivity and reliability. The quantitative researchers deal with hypothesis, i.e. with statements that predict the possible connection among the phenomena being researched, thus they are nonflexible, following one methodological framework which is defined in advance. For example, the success of a group of students is measured before the teacher uses a new teaching method and after the usage of this method in order to see if it has an impact on the students’ success. Regarding the statistical data processing, the descriptive statistics is not enough, i.e. these researchers look for statistical significance in the data, determining one working hypothesis (people who smoke have heart problems) which is accompanied by the null hypothesis, which, on the other hand, claims the opposite. The researchers always test the null hypothesis and follow the deductive thinking which goes from the general to the specific (top to bottom) and they exploit the well known Aristotle’s example on deductive thinking: All men are mortal, Socrates is human, and thus he is mortal too.

On the contrary, the qualitative researches, having a root in the constructivism, are mainly a reaction on the overforcing of the quantitative researches. The qualitative researchers assume that there are various interpretations of the reality that emerge from the data, bearing in mind that every individual perceive the world in its own way, thus stressing out the role of the context and the interpretation of the data relations in order to better describe the phenomenon being examined and to generate knowledge. These interpretations are subjected to change, depending on the time and the circumstances hence, the reality is not universal but connected with the person and the context. The qualitative researchers use a theory which is the frame of the research as well as different approaches, and try to answer the research question, examining different social settings, groups and individuals who live in them. In fact, these researchers consider the participants’ thoughts on the phenomenon being explored as extremely important, thus interpreting the meaning that others have for the world that surrounds them. The qualitative researches do not separate the people from their natural setting, which is complex, multidimensional and dynamic, meaning that the information are gathered by talking with the people and by observing how they behave and act within the context in which they are. Surely, the researcher’s interpretation cannot be detached from his own history, context and previous understanding of phenomena, thus the qualitative interpretation is, in fact, a combination of the researcher’s and participants’ interpretation.

The mixed researches, i.e. quantitative-qualitative are often named as pragmatic, and the researchers use the methods, procedures, and techniques which are usually related with the quantitative or the qualitative researches, claiming that these combinations lead to best results. For instance, the data may be obtained through directive interviews from a focus group, which is a qualitative approach, and then the results may be used for designing a survey to examine the attitudes of a larger group of participants, which is a quantitative approach, and at the end the data may be statistically processed. Obviously, various combinations are possible and they
depend on the research question and goal as well as on researchers’ creativity, need, and knowledge to mix these two methodologies.

The qualitative researches on language problems and phenomena and the way that language is used in the real world is quite interesting topic for the researchers (Benson, Chic, Gao, Huang & Wang, 2009; Burns, 1999; Brown, 2001; Brown & Yule, 1983; Duff, 2002; Fairclough, 2004; Freeman & Long, 1994; Heigham & Crocker, 2009; Lazaraton, 2004; Long & Zhou, 2019; McCarthy, 1991; Mengmeng, 2018; Merriam, 2002; Nunan, 1992; Richards, 2009; Richards & Morse, 2007; Van Dijk, 1997; Wodak, 2002; Wolcott, 2008; etc.). A lot of language qualitative researches exploit approaches that are typical for the social sciences and the health care and use, especially, the Grounded theory (Glaser & Strauss, 1967) and its modification (Strauss & Corbin, 1990) as well as other approaches. It can be said that these approaches used in the social sciences are very appropriate and suitable for examining language related researches, in the first place, because the language is a social phenomenon.

Albeit, the language qualitative researches are recognized as a relevant field of research and their number is permanently growing, they are still being criticized, in particular, regarding the question of the decision on the sample. Therefore, this study aims to answer the question: How can the sample be determined in the language oriented qualitative research, taking into considerations the views of some of the most influential researchers worldwide and offering three research scenarios which indicate that sampling in qualitative researches differs from sampling in quantitative researches and that while sampling, researchers are, in fact analyzing the data. Answering the question contributes to the researchers previous assumption that the sample in the qualitative language researches cannot be determined in advanced, that specifying the sample in these researches is an ongoing process which lasts till the very end of the research and that it directs the research.

1.1. Broader research context

The question of defining the sample in the social sciences and humanities is long ago introduced by Mead (1952), stating that the field note in anthropology should be done counting at least two participants, which enables to justify the notion community and further comparative experiences. Since then, a lot of researchers bring out their insights and on the techniques for its determination, stating the difference between the sample in the quantitative and the qualitative researches (Clarke, 2005; Corbin & Strauss, 2015; Curtis, Gesler, Smith & Washburn, 2000; Gentles, Charles, Ploeg & McKibbon, 2015; Kaufman, 1989; Kindsiko & Poltimäe, 2019; Kopper, 2008; Luborsky & Rubinstein, 1995; Marshall, 1996; McLeod, 2019; McCombes, 2019; Morse, 1994; Omana, 2013; Onwuegbuzie & Leech, 2005; Rapley, 2014; Richards, 2005; Ritchie et al., 2003; Spradley, 1979; Taherdoost, 2016; Ulas, 2006; Van Rijnsoever, 2017; Weller et al., 2018; Wilmot etc.) and stressing out its significant role for the validity of the research. Marshall (1996) even considers that there are methodological differences between the quantitative and the qualitative researches and that the techniques for determination of the sample in the quantitative and the qualitative researches cannot be mutually exchanged. This Marshall’s claim is supported by Patton (2002) who assumes that the decision on the sample in the qualitative research is connected with specific people, situations or places because they offer a specific perspective.

Following the quantitative logic, let us illustrate defining the sample with a simple example. If the researcher wants to explore the connection between drinking coffee and the high blood pressure, the whole number of people who have a high blood pressure on the planet represents the population. Yet, it is understandable that the researcher cannot have access to all of the
units (to all of the people who have a high blood pressure). Therefore, from the population (all the people who have a high blood pressure), the researcher chooses certain (definite) number of people (who have a high blood pressure) and those chosen people should be a representative sample, i.e. they represent that population. Then, on the base on the results obtained through researching of the sample, the researcher reaches conclusion and generalized them to the whole population. Clearly, all the relevant research features should be determined in advance (what counts as a high blood pressure, how many cups of coffee should be considered as harmful, according to the participants’ age, which age is appropriate for the study etc.). A lot of researchers exploit statistic operation to determine the sample size because they assume that the bigger the sample the more valid and objective the results are. The population refers to the overall number of units (people, events, phenomena, texts, document etc.) in the universe. In most cases, the researcher may not have access to all those units, thus from the overall number, he chooses definitive number of units which represents the sample. As shown in the example, specifying the sample in this way cannot be implemented in a large number of qualitative researches (surely there are qualitative researches which use a sample determined before the research itself), primarily, due to their complex nature. The qualitative researchers explain, interpret, describe etc., and they do not measure the collected data. Therefore, it appears that in the qualitative researches as suggested by Gentles et al. (2015), the sample, in fact, refers to the sources from which the information are obtained in order to realize the settled goals. As Lubrosky and Rubinstein (1995) state, the sample is about the units of one system (universe) which have to be taken into account in the research to provide valid representation of that system, meaning that the chosen sample should represent the population being examined.

According to the quantitative researchers the bigger sample is a confirmation for the representativeness of the population in order to obtain statistically significant results. Hence, the quantitative researchers cannot accept that the small sample in the qualitative researches may lead to meaningful results and contest their scientific value. Other researchers as Fugard and Potts (2015) even propose statistical operations for determining the size of the sample in the qualitative researches.

Although different techniques to specify the sample and its size are mentioned from the researchers, especially in the quantitative researches, there is an agreement among the qualitative researchers that the best and appropriate sample is the sample which gives answer to the research question and that fulfils the research goal (Marshal, 1996), bearing in mind that unlike the quantitative, the sample in the qualitative researches is usually small (smaller), though this is not a general rule.

However, it is not simple to determine how many units the sample should consist of. According to Marshall (1996) the appropriate number of the units representing the sample becomes obvious when the researcher is in progress, as new concepts and categories emerged. Further, he mentions the convenience sample, the sample based on researchers’ judgment on the most productive sample which answers the research question (key informants are a subtype) and the theoretical sample, i.e. the sample that is guided by the theory which asks for interpretative theories that emerge from the data. Morse (1994) argues that the question of the sample definition in the qualitative researches depends on the research scope and nature and on the quality of the data. Regarding the size of the sample, Green and Thorogood (2009) propose 20 interviews, Ritchie, Lewis and Elam (2003) 50, and Britten (1995) also 50. Lincoln and Guba (1995) introduce the notion information redundancy and Malterud, Siersma and Guassora (2015) the notion information power as parameters to specify the sample size, meaning that if the sample has bigger information power, then the less units are needed and that the sampling
stops when the saturation is achieved, i.e. the units are not providing new information. Glaser and Straus (1967) use the term theoretical sampling referring to the constant process of collecting and analyzing the data and to development of a theory, thus stressing out that the sampling is guided by the theory which emerges from the data and not from the features of one in advance defined population. There are researches that examine the academic discourse in regard to the sample (Gentles et al., 2015) and conclude that in typical qualitative researches the notion sample and techniques for determining sample are missing. They confirm that the determination of the sample in qualitative research, according to the research goal proposed by Patton (2002), is widely used, although there are researchers as Yin (2011) who assumes that the notion sample should be avoided in the qualitative research. It seems that the most logical is the Lincoln and Guba’s (1995) statement which indicates that all strategies for defining the sample are in a way purposeful, bearing in mind that the sample should provide answer to the research question and to fulfill the research goal.

On the other hand, researchers argue that determining the sample in advance is not in accordance with the conceptual and methodological bases of the qualitative research. They argue that it is extremely hard to define the correct sample size in advance because in these researches the principle of saturation is accepted, which according to Sandelowski (1998) directs the collection of data and analysis in the sense of information redundancy or as O’Reilly and Parker (2013) say the sample is in relation with the theoretical insights which emerge from the data.

Although in the language qualitative researches the sample may consist of people, various researchers mention different units as sample: a) individuals, artifacts, events, organizations or groups of research participants (Parahoo, 1997); b) number of people being interviewed, cases and focus groups (Wilmot); c) sources of information – informants and every other source which informs the researcher as observational data, existing documents, archival data, transcribed and various types of documents (Gentles et al., 2015; Van Rijnsoever, 2017; Yin, 2011); d) the number of the articles in journals (Kindsiko & Politmäe, 2019); e) documents (letters, pamphlets, agendas, reports, other researches and journals’ articles (Yin, 2003); f) persons and events (school events, ceremonies, cultural rituals) (Schensul, 2008); interview transcripts, journalistic articles, observation, documents and visual images (Cheek, 2008); g) groups, events or occurrences, but not people (Strauss & Corbin, 1998); h) people, behaviours, events or processes (Marshall & Rossman, 2006); i) people, events or information that clears the relevance of a category or categories (Charmaz, 2014); j) everything that is potential source of information (Richards, 2005) etc. These researchers’ insights contribute to the authors’ claims that people and what they say on particular language phenomenon are not the only units that should be count as sample.

2. Methods

In order to answer the research question, i.e. to show the various ways in which the sample can be determined in a qualitative language research, three different qualitative research scenarios are developed by the authors, and they are the sample in this study. The scenarios are analyzed by using the generic interpretative/inductive qualitative approach (Backett & Davison, 1995; Bryman & Burgess, 1994; Caelli, Ray & Mill, 2003; Creswell, 2009; Dey, 1993; Elliott & Gillie, 1998; Harding & Whitehead, 2016; Hunt, 2009; Kahlke, 2014; Marshall, 1999; Meriam, 2009; Morse, 2011; Ranse, Yates & Coyer, 2012; Sandelowski, 2010; Thomas, 2006; Thorne, Kirkham & McDonald-Emes, 1997), which does not follow the frame of one or more qualitative methodologies asking for specific methodological direction as the Grounded theory, phenomenology, ethnography etc., but rather the practical demands of the research question.
dictate the method of data collection and analysis. The analysis of the data includes constant use of comparison method and iterative methods which lead to broader understanding of the data and it could be said that while sampling, the authors do the analysis, which, on the other hand, is in accordance with Hamersley and Atkinson’s claim (1983) that in the qualitative research the data collection, analysis and interpretation are interwoven.

3. Findings and discussion

What follows are three different research scenarios that rely on the theoretical insights of the most influential researchers in this field, showing that sampling in qualitative researches is a complex and ongoing process and that it is very hard to specify the sample in advance.

3.1. Research scenario 1

Let us assume that the researcher notices that one orthographic rule $x$ in the written practice is used incorrectly and that he wants to examine this phenomenon. How can the sample be determined in this case? Starting from the researchers’ insights that the sample represents the sources from which the information is obtained, there is a space to think in several different aspects. From one hand, the sample may be this one rule (there is no other rule about that phenomenon) and this is in accordance with Patton (1990) who proposes that in the case study one individual or one institution represents the sample and with Myers and Burns (2005) who state that the small number of units, representing the sample, as well the whole population may be used when one problem or phenomenon is examined. In this case, the sample, in fact, represents the whole population, i.e. the researcher has the possibility to analyze the whole population which is reduced to the sample. Further, the researcher can take into account the existence of this rule in the various editions of the Orthography (if there are such), still, the sample represents this one rule $x$. The researcher gets information for this rule in the only or in the various editions of the orthography of one language. On the other hand, as analysis goes on, he may feel that he needs additional information, thus including two specialists in the given area, and this is in accordance with Spradley (1979) who states that the sample can be represented by key informants or as Patton (1990) names it by the sample according to intensity. In this case, the two experts, in fact, provide expert information for this specialized subject. Furthermore, the sample may be supplemented with the views of a group of teachers and students, although according to Morse (1994) there is no rule which will determine how many teachers or students are enough to be involved in the sample. It is obvious that the teachers and the students are people with different beliefs, thoughts, experiences and attitudes, and this is confirmed by Kaufman (1989) who argues that the group of people should be as homogenous as possible. In addition, as analysis goes on, if it is a rule from the Macedonian language orthography, the researcher can esteem that it is a good idea to look for this rule in various Slavic languages orthography as well in other world languages and to find similarities and differences etc. Therefore, it is on the researcher and on his understanding about the relation of the sample with the research question and research goals to determine whether the sample will consists only of this one rule or he will take into consideration all the mentioned aspects, which is in positive correlation with Luborsky and Rubinstein (1995) who highlight the access to the sources as well the economic resources as relevant aspect in determining the sample. In addition, while sampling, the researcher may realize whether his research question is worth to be answered and whether it needs to be changed or modified. Finally, although the researcher works with only one rule, there is a place for generalization which is confirmed by Berg (2001) who argues that, in certain degree, the research results can be generalized, by Gobo (2004) who states that working with a small sample does not mean that the generalization is not possible and by Gubrium and Holstein (2014) who claim that in the
qualitative researches the researchers are conceptualizing and hypothesizing, but they do all this constantly, while collecting data and analyzing. Therefore, it is clear that the sample cannot be determined in advance and that the research can take many directions, depending on the research question and goal, the nature of the research itself as well as on the sources of information the researcher collects. What is most significant is the fact that while looking for sources of information, i.e. sampling, the researcher is, in fact, doing the analysis and he constantly move forward and backward, until he reaches saturation, i.e. until he perceives that the sources he further takes into consideration do not feed his research with new information.

3.2. Research scenario 2

Let us assume that the researcher perceives that it is hard for his student to understand the term jargon and they cannot illustrate it with proper examples. What can be the sample in this case? With the researches’ first notice his curiosity is awakening, which is in accordance with Kumar (2014) who states that in the beginning of the qualitative researches it is enough for the researcher to identify what sparks his interest and what he wants to explore. Like in the previous scenario, the researcher may think of the sample in many ways. At first, he looks at the definitions in the textbooks and he sees that one and only one definition in several textbooks is present. From here, this one definition in the textbooks is the one source of information. Reading the definition more carefully and deeper, he understands that his students are right, i.e. the definition does not explain the essence of the term and that it is, fact, contradictory. Further, he tries to find answers in the existing literature in his mother tongue and reveals that all of the definitions he finds out in various sources are in, fact, the same, i.e. they have the same information given with two or three lines. Furthermore, he realizes that all the sources he explores give the same examples and that two to three jargon units are given. After that, conceiving that so far he has not reach the answer of his research question, he has to find more sources of information, thus he turns on to the existing dictionaries in his mother tongue and sees if there are any clearer explanations. While reading the explanations in the dictionaries, he comprehends that they do not give any additional answers. Hence, in addition, he has to search deeper, he has to address the foreign literature and tries to find the answers there. Then, he addresses the researches on the term jargon in the other Slavic languages as well in the other world languages, after that, he examines as more dictionaries of various Slavic and other world languages as possible, and during this process he constantly makes comparison between the content of that one definition and what he finds in other sources he uses, analysis and interpretation. Further, in the literature, he finds out that the term jargon is connected with a particular social group, and therefore he has to search within the sociology and perceives the way that the social groups are defined. Furthermore, he reveals that there are different social groups and that each group has its own jargon. Additionally, he has to relate these different jargon units within the Macedonian language system and see that some of them are terminology units as well, thus he has to address the terminology literature, and this will surely expand the research scope etc. Therefore, when starting to sample, the researcher cannot specify the sample in his research in advance and while sampling, he determines what sources he additionally needs to answer his research question.

3.3. Research scenario 3

Let us assume that the researcher wants to explore the semantic formulas for expressing compliments in the Macedonian language among the students of Macedonian language and literature. First, the researcher has to be familiar with the Austin’s and Searle’s theory of the speech acts, which is in agreement with the researchers’ claims that the researcher should have some previous knowledge of the phenomenon he chooses to examine. The researcher
considers all the 20 students in the group and asks them to write down all the compliments they use when complimenting each other. What is the sample in this case? Given that there is an agreement among the researchers that people are the units of sampling, it may be said that the sample consists of these 20 students, i.e. they are the source of information. Yet, these 20 students are listing compliments that they are using in their daily interactions with friends. Thus, the researcher collects, in fact, let us say, 100 compliments and these compliments are the corpus that the researcher analyzes. From here, it can be said that the sample represents these 20 students, but in the same time, the sample is the 100 compliments. These compliments are, in fact, a type of corpus and this is in correlation with Kida (2013) who lists several advantages for researchers when making their own corpus, primarily, because this corpus is adjusted to the research needs and reflects the attitudes of the researcher who composes it. In corpus-based qualitative language researches, the size and the possibility for generalization are also stated as a problems (Baker, 2010; Biber, Conrad & Reppen, 1998; Kohnen, 2007; Leech, 2007). However, although it is a general rule that the bigger corpus means bigger research validity, Stubbs (1996) reveals that each text or collection of texts can be comprehended as corpus, and although there are certain criteria, for now the question of the corpus size remains open. This is confirmed by Baker (2010) who says that particular journalistic articles can be seen as a corpus, and by Hardie (2016) who notices that even several sentences can be named as a corpus. Thus, the researchers encourage the DIY corpuses (Do it yourself).

4. Conclusions

From these three different scenarios several inferences emerge:

In the qualitative language researches it is almost impossible to specify the sample and its size in advance (of course, depending on the research question and the goal, there are qualitative language oriented researches which determine the sample and its size in advance), due to the complexity of the type of researches.

The researcher cannot know in advance what sources of information he will need.

Further, it is clear that the sample and its size determine the research course in many directions and only the researchers should decide which course he follows to answer his research question the best.

Furthermore, while sampling, the researcher decides whether the research question is worth to be answered, and he has the opportunity to change or to modify it, and this indicates the flexibility of these researches.

In addition, during this ongoing process, the researcher constantly moves backwards and forwards in the data, making constant comparison and many iterative steps until he gets the answer of the research question and fulfils the research goal.

Finally, the research suggests that the sampling process and the analysis are done simultaneously, because while sampling, i.e. deciding what sources of information he includes in the research, the researcher is inevitably analyzing the data.

This research is extremely meaningful for those interested to realize a qualitative research connected with language phenomena, primarily, because it shows them the complex nature of determining the sample and encourages them to think profoundly on what should represent the sample in their own research. At the same time, those interested in qualitative research
benefit from this research comprehending that the sample cannot and should not be always determined in advance and that the research question and aim, in fact, direct the way the sample is defined.

References


Biographical notes:

Violeta Janusheva is a full time professor at the Faculty of Education within the University “Sv. Kliment Ohridski” in Bitola, Republic of North Macedonia, teaching subjects related to the Macedonian language and to the assessment of the students’ achievement. She is a doctor in philological sciences and has published 129 research articles.

Jove Dimitrija Talevski is a full professor and vice dean at the Faculty of Education within the University “Sv. Kliment Ohridski” in Bitola, Republic of North Macedonia, teaching subjects related to geography, demo-geography, military field, educational sciences and assessment of the students’ achievements. Having achieved the professional specialization in the field of cartography in Tuluz, France, he acquired the title expert in field cartography. He was a representative from the Republic of Macedonia at the NATO and Partnership for Peace Geographical Conference in Brussels. As an expert he is a member of the Board for demarcation and marking of the border line between the Republic of Macedonia and Kosovo. He is a doctor in geographical sciences and has published over 70 research articles and dozen textbooks, guidebooks and teaching aids.

Milena Pejchinovska-Stojkovikj – PhD, works at the Faculty of Education – University “Sv. Kliment Ohridski“ in Bitola, as an associate professor for natural and social sciences and teaching methods courses. She has published more than 40 articles and scientific research papers in the field of Teaching and Learning Science in primary school and Assessment in education. Her research interest is in the field of Teaching Methods Pedagogy, with emphases to methods and strategies of learning and teaching in elementary classroom and the assessment and evaluation of students’ achievements.