Methodological principles of Mišótika Cappadocian data collection

Nicole Vassalou[1, 2], Dimitris Papazachariou[1] & Mark Janse[2]
University of Patras[1] & Ghent University[2]
vasnicol@hotmail.com, papaz@upatras.gr & mark.janse@ugent.be

1. Introduction

The aim of this paper is to discuss and compare the methodological principles and tools that have been used in our study of the vowel system of Mišótika, a variety of Cappadocian Greek, as well as to analyse the outputs of the application of each method. In order to investigate the phonetics and phonology of the Mišótika vowel system, we have decided to combine ethnographic with experimental methods. Specifically, the present study examines the importance and the drawbacks of applying ethnographic with experimental procedures in order to collect data of conversational and elicited speech.

It is an accepted fact that methodology is an essential part of every linguistic research. The appropriate structure of methodology will determine the entire progress of the study. So this part should be studied and applied rigorously in order to be successful. While our research is still in progress, we think it is important to present the methods which constitute the basis for the analysis of a dialectal system and the implementation problems that have arisen so far.

This paper is structured in six parts. Section 2 contains some basic information about the historical and linguistic background of Cappadocian Greek and the variety of Misti. In section 3 we analyze the goals of the present study. In section 4 we present the methodology used to approach the dialectal system and the drawbacks that have occurred. In section 5 we present the methods used to overcome the problems and to examine the dialectal system as better as we could. Finally, in section 6 we detail the primary conclusions of the present research.

2. Historical and linguistic background

The Cappadocian dialect was spoken until 1924 in the Central Anatolian Region of present-day Turkey. Cappadocian is a linguistic variety of Greek origin which had been in contact with Turkish for almost nine centuries after the invasion of the Seljuks in the 11th century and the conquest of Byzantine Asia Minor by the Ottoman Turks in the 14th century. The result of this contact is apparent in the Cappadocian lexicon, phonology, morphology and syntax, although the exact impact varies among the different subdialects according to the nature and duration of the contact situation.

One of the Cappadocian villages was Misti, a homogeneous town according to Dawkins (1916: 19). In 1924, the Cappadocians were forced to leave their homeland as part of the population exchange between Greece and Turkey. The inhabitants of Misti, estimated at around 400 families, were scattered all over Greece and settled in over twenty different villages and towns (homogeneous and mixed).

More specifically, Cappadocians settled either in homogeneous villages with other Cappadocian speakers or in villages with other Greek dialect speakers or in big cities like Athens and Thessaloniki. As a consequence, the descendants of the Cappadocian refugees have been in contact with different varieties of Greek during the last hundred years, which complicates the study of their speech, as there are several different versions of contact for the same linguistic system.

According to Dawkins, who conducted fieldwork in Cappadocia in the years 1909-1911, the vowel system of Cappadocian, including of the dialect of Misti, consisted of eight vowels, aligning it with the vowel system of Turkish.
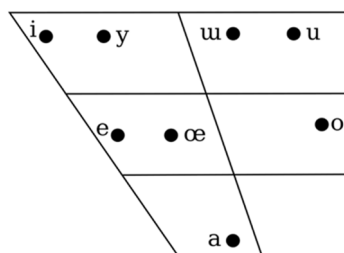

Figure 1: The older Cappadocian vowel system

However, in our recent research (Vassalou, Papazachariou & Janse in press) of one of the villages of Northern Greece with 2nd and 3rd generation native speakers, we found that the three Turkish vowels reported by Dawkins which do not exist in Standard Modern Greek (SMG), [y, œ, ɯ], are either lost entirely (the open-mid front rounded vowel [œ] and the high front rounded vowel [y]), or at the very last stage of elimination (the high back unrounded vowel [ɯ]). At the same time, the emergence of the near-open front unrounded vowel [æ], reported for Farašótika (Pharasa) and Pontic, but not for Cappadocian, by Dawkins (1916: 153), has prompted us to examine its incorporation into the linguistic system of present-day spoken Mišótika.

3. The present study

Our study aims to describe and analyse the vowel system of contemporary Mišótika, as it is used by different generations of native speakers in Greece – particularly in Cappadocian refugee communities in the prefectures of Kilkis, Thessaloniki, Larissa, Kavala, Ioannina and Alexandroupoli, and under varying contact conditions.

Our research focuses on the study of the linguistic changes that the dialect has undergone since the 1924 population exchange. Through a detailed study of contact situations of the Cappadocian communities in Greece, the main purpose is to find how the Mišótika vowel system is formed, to determine the phonological status of the vowels and to analyse their distribution in the vowel spectrum. Furthermore, we want to examine the effects of language contact in the vowel system and the role of linguistic and non-linguistic parameters which are displayed, activated, and interact in contact situations, affecting changes in the linguistic system.

The basic research question is whether language contact with SMG or other varieties of Modern Greek is gradually leading to the language death of Mišótika or the creation of a new koine[1], i.e. a new variety which includes mixed features from Mišótika, SMG, and/or other dialectal varieties. We can perfectly well appreciate the pressure that the Cappadocians felt from the locals with whom they were in contact after the population exchange, and the severe stigma that any Turkish characteristics carried for many decades, not just in the language but in other aspects of social behaviour as well. One

---

[1] The koineization process results in the creation of a new variety, which includes mixed features from the mutually comprehensive linguistic systems that have been in contact (cf., e.g., Trudgill 1986: 107; Hinskens 1992: 15).

of the results of this stigmatization is to be found in the linguistic system of the variety, as the dialect speakers were trying to accommodate to the new linguistic environments.

The social stigmatization and the attrition of the dialect complicate the study of Mišótika very considerably. Focusing on these conditions, an ethnographic study seems to be the only valid method to collect dialectal data. Despite this, we decided to approach the dialectal system by using a combination of ethnographic and experimental methods to collect data of conversational and elicited speech. In fact, the ethnographic and experimental principles are in a sense opposite methods. Both have equally positive and negative effects. However, our goal is to incorporate both in our research, aiming to complement each other and trying to highlight only the positive aspects of each method. We maintain that the combination of these methodological principles will provide a larger and more in-depth study of the Mišótika vowel system.

Post & Nolan (2012) stress that the combination of the two types of speech (i.e. conversational and elicited) must accomplish 'ecological validity', that is, speech recordings should be a sample of natural speech communication in register with informal-dialectal authenticity. There is the need for balance between the 'ecological validity' of the elicitation, that is the closeness of recordings to spontaneous speech, and how tightly controlled the elicitation needs to be to induce the speaker to produce the required dialectal speech in experimentally comparable ways.

According to Kainada & Baltazani (2013), linguists doing dialectal fieldwork should use a combination of conversational speech and tightly controlled speech tasks. They claim that both can elicit representative dialectal speech, provided they avoid informants' accommodation to the dominant standard variety. Without loss of authenticity of speech, we can control critical aspects of both linguistic and extralinguistic context, which cannot be controlled in spontaneous speech corpora.

This combination is also the method suggested by Himmelmann and Ladd (2008), who propose that researchers start with conversational speech (something we have already done to some extent in our recent study), identify the linguistic issues of interest and then move on to experimental approaches for data elicitation. Nevertheless, it is not obvious that these two methods can always be combined and there are several problems which we will address in the present paper.

4. Methodology

To approach the dialectal system, our recordings of native speech have taken place in two phases. Each phase was aimed at producing different results. In this section we will analyse the methodological principles that have been used in our study of the vowel system of Mišótika, and the reasons of the application of each method. At the same time we will present the problems that we encountered in attempting to implement experimental tasks.

4.1 Ethnographic method

The first phase of data collection focuses on fieldwork with recordings of friendly and casual conversations, which has been accomplished according to the ethnographic methods. The aim has been to conduct spontaneous dialectal speech, since conversational speech provides the most naturally uttered and representative dialectal data.

The ethnographic method suggests that the researcher should be in a direct and continuous relationship with the native speakers of the dialect in order to understand

their way of living and to be able to elicit natural dialectal speech. A main feature of ethnographic research is the effort to develop personal relationships with the informants. Hence, in the present research the ethnographic study precedes the experimental study. It is an accepted fact that only if the informant trusts the fieldworker during the ethnographic study, will he or she accept to participate in later experiments.

One of our main considerations was to avoid the 'Observer's Paradox' (Labov 1972), something that would have devastating results to our study due to the severe stigmatization of the Cappadocian variety. However, we strongly believe that the emergence of the Observer's Paradox is not simply due to the presence of the microphone and the recorder, but the result of the communicative settings of the recording, in particular, the absence of real social bonds between the fieldworker and the informants (Papazachariou 2006). Therefore, we decided to use a native speaker who is a central member of the speech community as the 'intermediary', aiming to reduce the impact of the alien presence and trying to create a friendly, and everyday casual communicative situation to produce natural dialectal speech.

Up to now, we have succeeded in collecting a sufficient number of recordings with the implementation of the ethnographic method, while the presence of the 'intermediary' has yielded the intended dialectal speech.

4.2 Experimental methods

After the completion of ethnographic research, we moved on to the second phase of data collection. This phase has been conducted utilising experimental methods. At this point, we should note that the appropriate stimuli of the tasks are based on our previous findings about the vowel system of Mišótika (Vassalou, Papazachariou & Janse in press). In a pilot study we collected a corpus of dialectal words including the target vowels in different metrical environments (stressed or unstressed, ultimate or other syllables, di- or polysyllabic words, etc.). The corpus was created on the basis of analyses of older recordings of dialectal speech, as well as on the basis of existing glossaries of Mišótika.

To investigate the production of all the words collected, we have implemented experimental methods. In the ethnographic data collection, it is almost certain that the cases we want to examine either will not be produced at all or will be realized only by some informants. In other words, significant linguistic phenomena may be missed during the ethnographic data collection which will make it difficult if not impossible to make relevant comparisons.

For these reasons, we decided to conduct additional naming and production experiments. The tasks are structured so as to guide the informants to be accommodated to the variety under investigation. The participants' responses in both tasks have been recorded in order to be analysed with PRAAT (Boersma & Weenink 2013). Praat will be used for the transcription, annotation and formant analysis of the data, so as to showcase the distribution of the vowels in the vowel spectrum and also the linguistic environments and position of stress that affect the realization of each vowel.

Moreover, according to the method proposed by Kainada & Baltazani (2013), the instructions and aural prompts directed at the participants had been pre-recorded in Mišótika by a native speaker of the dialect and incorporated in the PowerPoint presentation for reasons of homogeneity of the experimental process across participants and as a stimulus for eliciting dialectal speech. Each controlled experiment does not exceed 20 minutes so as to avoid any discomfort to the informant. Finally, we have to stress that the experiments were conducted on a pilot basis in a target group of native

speakers in order to determine their proper functionality and whether the informants could cope with the requirements of each experiment.

4.2.1 Picture – Naming Task

The first task created was a picture, i.e. a naming task, where participants were shown pictures on a computer screen using a PowerPoint presentation and asked to name what the pictures present – in their dialect, and not in SMG. All the pictures in the task attempted to elicit words with the target vowels in different metrical structures. The purpose is to identify the realization of each vowel in each individual word. For example, they were shown the following picture and were asked to identify the pronunciation of the word in Mišótika.

Picture 1: the expected answer → πέντζερε [ˈpʰænʤæræ] 'window'

Such tasks have the advantage of target phenomena being elicited in a controlled fashion, while the resulting data is still spontaneous in that it is unscripted. Through this experimental pilot attempt on a small group of native speakers of the dialect it was concluded that while the experiment was developed with particular care in order to incorporate all the vowels of the dialectal system and their realisation in all possible environments, in practice essential weaknesses arose.

More specifically, a fundamental problem which most researchers face while attempting such experimental methods relates to the negative attitude of the informants, mainly the elderly, towards any kind of experiment. On most occasions, experimental procedures cause the exasperation of the informants especially in the presence of the computer. This results in the participants refusing to perform the task at hand. It is a fact that the elderly are not accustomed to the use of the computer and generally panic at the idea that they must sit in front of one to perform an experiment. Consequently, an experimental procedure such as a picture-task may not seem easy to implement. During the fieldwork, elderly people had difficulty engaging even in an open conversation in their own dialect, let alone using a computer.

Besides the general negativity which exists in such cases, another basic problem arose through the pilot picture task method, which was the way in which native speakers comprehend most of the images included in the experiment. In particular, many of these images seemed to confound the informants who, in turn, failed to produce a response. This was due to the fact that even though the depicted images were familiar, the form which the informant possessed in his/her mental lexicon differed from the given image and so could not answer properly. In many cases, the native speakers of Mišótika lacked the form as we perceive them in contemporary terms.

For example, when asked to realize the word [tʃœˈʃmæ] which in Mišótika means 'tap' and also providing an image of a simple contemporary 'tap', the participants could not retrieve the correct meaning on the basis of the image in Mišótika, because as it seems the word [tʃœˈʃmæ] has a different depiction in their mental lexicon (which corresponds to central communal fountain taps found in villages many years ago). Therefore, many of the words could not be examined through such an experiment as many images were inappropriate to lead the informant to retrieve the anticipated structure.

Finally, the focal point was that the stage at which the informants were given an image and were unable to produce an immediate response, when they were assisted with the word in SMG orally, then they could produce a response with ease in Mišótika. This fact proves that while the speaker is aware of the target word, during the implementation of such an experiment he/she is often unable to realise the task and as a result provides erroneous data.

4.2.2 Translation Task

The second task that had been created was a translation task, where participants read from the computer screen short utterances in SMG and were asked to translate/produce these utterances in the Mišótika dialect, as in the following examples:

(1)
αυτοί μαθαίνουν τη δική μας γλώσσα
ιτούρα μαχαίν'νι **τεμέαρ** γλώσσα (the expected answer)
[iˈtura maˈhenːi **tæˈmær** ˈɣlosa]
'They learn our own language'

(2)
τι σκέφτεσαι τώρα;
τι **ντϋσϋνντίεις ντερέ**; (the expected answer)
[ˈtʰi **dyʃynˈdiis dæˈræ**]
'What are you thinking now?'

The marked words are the target words of each utterance. The translations are designed to give the linguist information about the vocabulary, metrical structures and accent. All the utterances to be used are short and simple. The goal here was to make the participants produce the structures that were not included in the previous experiment, as they were not displayed, such as verbs, adverbs, adjectives and pronouns.

Unfortunately, during the pilot implementation of this experiment, it turned out that the informants did not always produce the desired translation. More specifically, when they read the given sentence from the computer in SMG, they had difficulty translating all the words correctly and forming the sentences in Mišótika properly. As a result, in many cases, the informant did not concentrate enough and he/she translated some of the words in the sentence while omitting or distorting the translation of the target word, which was the main purpose of this experiment.

The drawbacks of this translation task are likely to reproduce the gradual linguistic change that this dialect has undergone and mainly due to language contact with SMG following the settlement of the Cappadocian refugees in Greece. It seems that because the dominant language is SMG, it is not easy for the informant to make a swift and fluent transition from the linguistic system of SMG to the Mišótika version. As a result,

participants have difficulty in delivering the appropriate translation of the given sentences and also the structure of their answers basically depict in their speech the mixed features of both linguistic systems.

Also, it is worth noting that the experiment was conducted in a second stage without the use of a computer because of the negative response of the informant towards this. For this reason, the sentences were produce orally in SMG through a native speaker of the dialect in the role of the 'intermediary', who in turn asked the informants to translate into Mišótika. The results of this experiment were virtually the same. The informant had difficulty translating the whole sentence and in most cases translated selectively, producing the rest of the sentence in SMG. This resulted in the omission of the realisation of the target word and hence the experimental procedure was considered unsuccessful.

In conclusion, although the target words were incorporated into a small number of sentences so that the experiment would be completed successfully, the participants not only displayed a negative reaction towards the computer but also seemed to be discouraged by the demands of a translation task.

5. The adjustment of the tasks

In order to investigate the whole corpus of the dialectal words including the target vowels in different metrical environments (stressed or unstressed, ultimate or other syllables, di- or polysyllabic words, etc.) we had collected, it was decided to adjust the experiments based on the capabilities of the informants. As a result, taking into consideration the problems mentioned in previous sections, we attempted to elicit the realization of these words in a different method so as to gather the desired data for our study.

In particular, we gathered all the target-words in both experimental procedures, and incorporated them into a traditional oral questionnaire. In order to avoid our participation during the collection of data as in the previous gathering of spontaneous dialectal speech, we asked our 'intermediary' to perform the questionnaire and record the data.

Under these new conditions, the 'intermediary' produced the version of every target word in SMG orally and requested the informant to deliver the structure in Mišótika directly. Furthermore, in order to avoid any inconvenience on the part of the informant during the procedure which we subjected him/her to, we combined the gathering of conversational and elicited speech during the same recording of each native speaker. More specifically, we asked the 'intermediary' to engage casual conversation with the informant for 45 minutes so as to gather the desired spontaneous dialectal speech and then asked him/her to realize every word in the questionnaire. This had as a result the whole recording of every informant to last approximately one and a half hour.

When we utilized the above procedure, the results of the study proved to be quite encouraging. On the one hand, the placement of an 'intermediary' as the facilitator of the interviews led the informants to feel more comfortable and deliver native dialectal speech while at the same time the informants accepted to realize the corpus of the questionnaire. On the other hand, the elicitation of the structures under investigation in the form of interactive oral questionnaire seemed to function better, as the informants, in the majority of cases, realized every target word directly and with relative ease.

It is worth noting that throughout the recording, the informants showed no signs of fatigue and were quite satisfied with the fact that their job had been completed after the ending of the recordings. Conversely, the implementation of the experiments based on

the initial version, which was presented above, would require the speaker to complete the procedure in at least two phases (i.e. 1st face: conversational speech, 2nd phase: the experiments). This fact could have caused further problems as it is deemed difficult, when the number of the informants is significant, to manage to persuade all the informants to devote the required time for the realization of the study.

6. Closing remarks

Wrapping up the previous discussion, we hope to have shown that our goal is to present the methodological principles that determine our study, so as to identify the vowel system of Mišótika. We have tried to showcase how spontaneous and controlled elicited speech can be combined in terms of extracting dialectal data and how both the ethnographic and experimental principles can be implemented in our research. Furthermore, we wanted to present the problems which could arise during the implementation of the experiments and also to show the ways we chose to overcome these problems.

The basic conclusion we wish to highlight is how important it is to scrutinize all the methodological parameters of the fieldwork and more specifically the methods involved in the implementation of the experiments so as to realise the required collection of dialectal data. We strongly believe that the combination of both methodological principles offers the best insights into the vowel system of Mišótika, and it constitutes a follow-up on our previous research.

References

Boersma, P. and D. Weenink. (2013) *Praat: Doing Phonetics by Computer*. Version 5.3.51, retrieved June 2, 2013. www.praat.org.

Dawkins, R. M. (1916) *Modern Greek in Asia Minor: A Study of the Dialects of Silli, Cappadocia and Pharasa, with Grammar, Texts, Translations and Glossary*. Cambridge: Cambridge University Press.

Himmelmann, P.N. and R.D. Ladd. (2008) Prosodic description: An introduction for fieldworkers. *Language Documentation and Conversation 2 (2)*: 244-274.

Hinskens F. (1992) *Dialect Levelling in Limburg: Structural and Sociolinguistic Aspects*. Ph.D. thesis. University of Nijmegen.

Kainada, E. and M. Baltazani. (2013) Evaluating methods for eliciting dialectal speech. In: M. Janse, B. D. Joseph, A. Ralli and M. Bagriacik (Eds.), *Proceedings of the 5th International Conference on Modern Greek dialects and Linguistic Theory*, 101-123.

Labov, W. (1972) *Sociolinguistic patterns*. Philadelphia: Pennsylvania University Press.

Papazachariou, D. (2006) The Quantitative Study of the Lateral Variable (l) in the Dialect of Patras. In: M. Janse, B. Joseph & A. Ralli (Eds.), *Proceedings of the Second International Conference on Modern Greek Dialects and Linguistic Theory*. Mytiline: Patras University Press, 298-313.

Post, B. and F. Nolan. (2012) Data analysis for prosodic analysis of continuous speech and dialectal variation. In: A.C. Cohn, C. Fougeron, & M. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology*. Oxford: OUP, 538-547.

Trudgill, P. (1986) *Dialects in Contact*. New York: Blackwell.

Vassalou, N., Papazachariou D. and M. Janse. (In press) The Vowel System of Mišótika Cappadocian. In: A. Alexiadou, J. Androutsopoulos, Th. Georgakopoulos, A. Kalokairinos & Th. S. Pavlidou (Eds.), *Proceedings of the 12th International Conference on Greek Linguistics*. Edition Romiosini - CeMoG, Freie Universität, Berlin.