

A Folk Songs Corpus for Constructing and Updating Dialect Dictionaries

Katerina T. Frantzi

Dept. of Mediterranean Studies

University of the Aegean

Dimokratias 1, 85100

Rhodes, Greece

frantzi@rhodes.aegean.gr

Abstract

We present the construction and processing of a folk songs corpus for the construction and updating of dialect dictionaries. The Greek folk songs are an important form of art to reveal dialect information. However, the information used for the existing dialect dictionaries has been until now acquired manually. For automating the extraction of dialect lexicographical knowledge and the construction and updating of dialect dictionaries, we involve corpus linguistic tools. The dialect dictionaries are updated with all the words that present dialect characteristics and examples of their real use in the songs. The corpus is dynamic and new songs are added continuously, making the processing also continuous, for the most complete and accurate updating of the dictionaries. The application area is the group of the Dodecanese Islands, south-east Greece, while future work involves the covering of more Greek areas.

1 Introduction

The importance for the preservation of cultural heritage has been recently widely recognized. Making the preservation digital gives easy access to well organized cultural databases which are easy to enhance. Dialects in Greece could gradually disappear since they are nowadays spoken mainly by the eldest. Dialect dictionaries are important for the preservation of dialect lexicographic knowledge. As McEnery et al. (2000) indicate however, language resources on minority languages are still in their infancy, and the case is even worse when it comes to dialects. In this work we present the construction of a folk songs corpus for the extraction of lexicographic knowledge to be used for the construction and updating of dialect dictionaries. As Μπαζιανιάς (1997) notes, folk songs are an important form of art to reveal such information. The folk songs used to be the expression of those sharing the same geographic area and time, given from the elder to the younger orally and expressing peoples' life, feelings, concerns, relations, hopes, beliefs and values.

Till now there has been no digital preservation of folk songs, while their processing has been only attempted manually. We exploit the songs using corpus linguistics tools to reveal knowledge in a complete, precise and quick way (Biber et al. 1998; Ooi, 1998). Our application area is the group of the Dodecanese Islands, South-East Greece, while in future more Greek geographical areas will be considered. The idea is rather simple: Corpus linguistic

tools are applied on songs corpora made from subsets of the Dodecanese islands and/or the islands areas. The words extracted from the corpora that are not found in modern Greek dictionaries are added to electronic dialect dictionaries created for the islands, groups of islands and areas within islands (according to linguistic atlases). The dialect words to be added to the dictionaries present lexicological or other linguistic information. A language expert updates the dictionaries with any relevant information for each new word. Information on the context and use of the dictionary words is also added together with some examples from the corpus. The corpus is continuously enhanced by adding new songs with the scope of the most complete covering of the Dodecanese area. We involve continuous exploitation of the dynamic corpus resulting in the continuous updating of the dialect dictionaries.

2 Dodecanese Folk Songs

Folk songs in the Dodecanese islands survived over the years keeping tradition alive. According to Παπαϊγναντίου (1981), a folk song is a poem that can be sung and often danced. The Dodecanese music has its roots on the Byzantine music and even the ancient Greek music and was in older times mainly used in celebrations. Most of the songs, except some of the narrative ones can be danced. Because of the current way of living however, tourism and the media, the songs are forgotten together with other parts of tradition.

The Dodecanese islands share much of their culture, however each of them has also its own heritage and language characteristics: same songs can be found in many variations. Μακρής (1983) claims that it would be very difficult to produce a collection of the folk songs of the whole of Dodecanese because of the varieties on the language and the tradition among the islands and even among the villages within the same island. Among the most important collectors of the Dodecanese folk songs are: for Rhodes: Βροντής (1930; 1950; 1956-1957), Παπαχριστοδούλου (1959; 1961), Γνευτός (1981), Μπογιάνος (2001); for Kos: Μαυρής (1961-1962); for Karpathos: Μιχαελίδες-Νουάρος (1928; 1969; 2001), Γεοργιάδης (1956-1957), Ιωαννίδου-Μαπαρμπαρίγου, (1960-1961), Αλεξιάδης (1975), Χαλκιάς (1980; 1994); for Kalymnos: Ζεβρός (1956-1957), Χατζηθεοδώρου (1989); for Nisiros: Καζαβής (1940), Σακελλαρίδης (1965); for Kasos: Μαυρής (1928); for Chalki: Bovy-Baud (1935, 1938), Τατσούλη (1947); for Astipalaia: Τατσούλη (1996); for Simi: Χαραβιάς (1884-1885); for Tilos: Καμμάς (1947); for Leros: Dawkins (1950-1951); for Patmos: Σμυρνάκης (1947).

The Dodecanese folk songs are divided into categories according to their content. Μιχαηλίδης-Νουάρος (1969) divides the songs into (a) narrative songs, e.g. historical songs, (b) *μαντινάδες* (songs that consist of two lines), e.g. some of the love, marriage, celebration, sorrow songs, (c) satire songs and (d) songs of events, e.g. war, earthquakes. The Dodecanese variations of the *ακριτικά* (songs for the border-defenders), and especially those of Rhodes and Karpathos (Μακρής, 1983; Πολίτης 1991) are among the most important and form a large category. The category of *μαντινάδες* (also a large category) came from Crete and was easily adapted to Dodecanese and especially to the island of Karpathos. These songs were mostly improvised and mainly used for marriages, love, immigration and sorrow.

3 From the Songs Corpus to the Dialect Dictionaries

A database of folk songs of the Dodecanese islands has been created by Frantzi (2005). The songs are obtained from hardcopy collections found mainly in local libraries. During digitization no corrections are made on the original text. The database is continuously updated with new songs as these are acquired. For each song, the following information is kept: title, category, island, village, source, collector, information on the publication, links to other versions of the song and other information available. Table 1 gives a sample of the songs: a love song from the island of Karpathos, and particularly the village of Elympos, a village that dialectically differs from the rest of the island. Table 2 gives the record for that song in the database.

Η αγάπὸ μου φώναζε (Ελύμπος)	
Η αγάπὸ μου φώναζεν - βρ' εσύ σ' απού με τρέλανες βρε, και πάω να την έβρω και τη στρατί δεν ηξέβρω. Και πίνω το στρατί στρατί - γιατί σ' αγάπου πουλι μ' από μικρή βρε, στρατί το μονοπάτι. βήσανα που τα 'χ' η αγάπη. Και το στρατί μ' επόσωσε - βρ' εσύ σ' απού με τρέλανες βρε, στις αγάπὸ τη πόρτα - περκαντί μου παίξε βόρτα. Ερίσκα τις πόρτες ασφαλιστές - μαργιάδες που 'ν' οι κοπελιές βρε και τα κλειδιά περιμένα - κλειώ τα μάτια μου για σένα. Και τις γερόντισσες ρεϊτά: - Που 'ναι η κόρη απ' αγάπὸ. -Έρε μεσά 'ναι και κοιμάται και για σένα συλλοάται. -Κόρη, α κομιάσαι ξόπνησε και μετά μένα μιλάσε βρε, κι αν αγριωνάς κοιμήσου και για μένα συλλοήσου. Κι αν είσαι με τη μένα σου, πάρε κ' εμέν' ανάμα σου βρε και με την περιή σου, πάρε με κ' εμέ μαζί σου. Κι αν είσαι με το κύρη σου, κόψε με νοικοκύρη σου βρε και με τον περιό σου, κόψε με' αγαπητικό σου.	

Table 1. A song from Elympos, Karpathos.

Song Title	Η αγάπὸ μου φώναζε
Island	Κάρπαθος
Village	Ελύμπος
Category	Αγάπη
Number of Words	172
Number of lines	16
Source	Τραγούδια των Δωδεκανήσων
Collector	S.BAUD-BOVY
Publisher	Μουσικό Λαογραφικό Αρχείο
Year of publication	1935
Place of publication	Αθήνα
Variations	N/A
Other info	N/A

Table 2. A record from the songs database.

The size of the database is currently of about 1,800 songs. We construct corpora out of the songs according to the type of information we every time want to extract. Various corpora are constructed, e.g. a corpus of the love songs of the island of Kos, a corpus of all the songs of the island of Nisiros, a corpus of the marriage songs of all the islands, a corpus of all the songs of all the islands, etc. For the construction of the dialect dictionaries we applied corpus linguistic tools on the corpora we created for each island or island area according to linguistic atlases. We extracted word frequency lists. The words of the lists are checked – if they exist or not – in common Greek dictionaries. If they do not exist, they are added to the island's dialect dictionary together with the frequency of occurrence in the corpus. Their meanings are explored and given by the language expert. Words that contain other dialect information, i.e. in their morphology, phonology, syntax, semantics, are also added to the dictionary together with the linguistic characteristic they present. Finally, for each word in the dictionary, examples on their real use in the songs corpus are also given. Table 3 shows a sample of the dialect dictionary of the island of Kastellorizo.

ήρτεν	5	διατήρηση και αναλογική ανάπτυξη του τελικού /n/ σε ονομαστικούς και ρηματικούς τύπους, συχνά για την αποφυγή χιτισμοδίας & ανομοίωση ρθ>rt
ήβγεν	5	διατήρηση και αναλογική ανάπτυξη του τελικού /n/ σε ονομαστικούς και ρηματικούς τύπους, συχνά για την αποφυγή χιτισμοδίας
εισείνη	5	παιτακισμός: η προφορά του /k/ πριν από τα [e] και [i] ως /ts/
δεφ	5	αφομοίωση τελικού /n/ με το σύμφωνο της επόμενης λέξης, όταν είναι ένα από τα παρακάτω: β, γ, δ, φ, χ, θ, λ, ρ, μ, σ
αέρφια	5	αποβολή των μεσοφωνηεντικών /r/, /s/ και /β/
σουντούτσιν	4	διατήρηση και αναλογική ανάπτυξη του τελικού /n/ σε ονομαστικούς και ρηματικούς τύπους, συχνά για την αποφυγή χιτισμοδίας
πουόλι	4	παρατεταμένη προφορά των "διπλών" συμφώνων της γραφής και των στιγμασίων μέσα σε μια λέξη σε λέξεις αρχαίες, δάνεια και νεότερες
πλοίον	4	διατήρηση και αναλογική ανάπτυξη του τελικού /n/ σε ονομαστικούς και ρηματικούς τύπους, συχνά για την αποφυγή χιτισμοδίας
έκλαψεν	4	διατήρηση και αναλογική ανάπτυξη του τελικού /n/ σε ονομαστικούς και ρηματικούς τύπους, συχνά για την αποφυγή χιτισμοδίας
χρουσό	3	υπόλειμμα παλαιάς προφοράς του /n/ ως /ou/ ή επίδραση των περιεκείμενων συμφώνων χαλκίων, υγρίων, έρρινων ή του /s/
τσόβγω	3	παιτακισμός: η προφορά του /k/ πριν από τα [e] και [i] ως /ts/
σεθλι	3	η προφορά των υπερθετικών /k/ και /r/ ως παχιά ουρανικά προ των ασθενών φωνηέντων [e] και [i] και στην συνέχεια την μετατροπή τους σε /ts/ και /s/ αντίστοιχα
μπρέπια	3	Τα ηχηρά /μπ/, /ντ/, /γκ/ ή /γγ/ προφέρονται απερρινωμένα σε όλες τις θέσεις.
καπιτάν	3	τροπή του /e/ σε /i/, καθαρά ιδιωματικό ως πολύ κοντινά στην εκφορά φωνηέντα

Table 3. Sample of the dialect dictionary of Kastellorizo.

The first column on Table 3 gives the words with the dialect information. These are words from the word frequency list extracted from the songs corpus of the island of Kastellorizo. The second column gives the frequency of the words in the corpus in its current size. The third column gives the linguistic characteristic found. So, for example the words *συντούτσιν* and *τούβγω* (lines 6 and 11 respectively) are dialect words that have no entries in common modern Greek dictionaries (Μπαμπινιώτης, 2002, Γρηγορόπουλος, 2002). Their frequencies of occurrence in the current songs corpus are 4 and 3 respectively. Their meanings are explored and added to the dictionary by the language expert.

The word *αέρφια* (line 5, meaning “brothers” and/or “sisters”) appeared 5 times in the current corpus of the Kastellorizo songs. Its meaning is known since the word is used in common modern Greek, however it presents two dialect characteristics on its morphology: (a) that of the omission of the consonants *γ* (pronounced as *w* in *win*), *δ* (pronounced as *th* in *then* – our case) and *β* (pronounced as *v* in *van*) when within words, and (b) that of the alteration of *λφ* (pronounced *lf*) to *ρφ* (pronounced *rf*). That is, in common modern Greek the word *αέρφια* would be *αδέλφια*. The word *πλοίον* (line 8, meaning “ship”) appeared 4 times in the current corpus, and shows the characteristic of the preservation of the archaic final *_* (pronounced *n*) on nouns, i.e. in common modern Greek the word *πλοίον* would be *πλοίο*. The dictionary gives also examples on the use of the words, e.g. for the word *ετσείνη* (line 3, meaning “she”, in common modern Greek it would be *εκείνη*), we keep the line “Αουρος κόρην απαά τσ’ ετσείνη δεν το ξέρει” (meaning “a young man loves a young woman but she does not know it”) from the love song titled ‘του ετοιμοθάνατου νέου από επωτά’ (meaning “For the young man who is dying of love”).

4 Future Plans

We presented the use of a folk songs corpus in electronic form for the construction and updating of dialect dictionaries. The corpus is dynamic (new songs are added as acquired) resulting to dynamic dictionaries. The dictionaries are updated with new words, new meanings and new dialect information. Regarding future work we aim at the following directions:

- Firstly, the corpus will be continuously updated, and so will be the dictionaries. This is part of both current and future work.
- The dictionaries will be enriched with collocations as well.
- More dialect text types will be considered (i.e. stories, oral speech).
- More Greek geographical areas will be considered and new corpora will be built and processed for the construction of their dialect dictionaries.

5 Acknowledgement

I wish to thank all of my students who helped with data collection.

References

A. Dictionaries

- Μπαμπινιώτης, Γ. Δ. (2002), *Λεξικό της νέας ελληνικής γλώσσας: με σχόλια για τη σωστή χρήση των λέξεων*. (πρώτη έκδοση 1998) Κέντρο Λεξικολογίας, Αθήνα (in Greek).

Γρηγορόπουλος, Δ. (2002), *Λεξικό της νέας ελληνικής γλώσσας: με αναγωγή στην αρχαία ελληνική*. Έννοια, Αθήνα (in Greek).

B. Other literature

- Biber, D., Conrad, S., Reppen, R. (1998), *Corpus Linguistics - Investigating Language Structure and Use*, Cambridge University Press, Cambridge, U.K.
- Bovy-Baud, S., (1935,1938), *Τραγούδια των Δωδεκανήσων, τόμ. Α'-Β'*, Αθήνα.
- Frantzi, K.T. (2005), 'Preserving and Exploiting the Dodecanese Traditional Songs', in *Book of Extended Abstracts of the 1st South-Eastern European Digitization Initiative Conference (SEEDI), Digital (Re)Discovery of Culture*, Ohrid, F.Y.R.O.M. (Republic of Macedonia), 2005, pp. 41-45.
- McEnery, A.M., Baker, P., Burnard, L. (2000), 'Corpus Resources and Minority Language Engineering', in Gavrilidou, M., Carayannis, G., Markantontou, S., Piperidis, S. and Stainhauer, G. (eds.), *Proceedings of the Second International Conference on Language Resources and Evaluation*, Athens, Greece, 2000, pp. 801-806.
- Οοι, V. (1998), *Computer Corpus Lexicography*. Edinburgh, University Press, Edinburgh.
- Αλεξιάδης, Μ. Α. (1975), *Συμβολή στην Έρευνα του Καρπαθιακού Δημοτικού Τραγουδιού (Ανέκδοτες Παραλλαγές)*, Αθήνα (in Greek).
- Βροντής, Μ. Α. (1930), *Της Ρόδου Παραδόσεις και Τραγούδια*, Ρόδος (in Greek).
- Βροντής, Μ. Α. (1950), *Η αγάπη στη Ρόδο, (Δημοτικά Τραγούδια της Αγάπης)*, Ρόδος (in Greek).
- Βροντής, Μ. Α. (1956-1957), 'Δημοτικά Τραγούδια της Ρόδου' στο *Δωδεκανησιακόν Αρχεῖον τόμ. Β'*, pp. 267-277 (in Greek).
- Γεωργιάδης, Γ. Μ. (1956-1957), 'Δημοτικά Τραγούδια Μεσοχωρίου Καρπάθου' στο *Δωδεκανησιακόν Αρχεῖον τόμ. Β'*, pp. 278-288 (in Greek).
- Γνευτός, Π. (1981), *Δημοτικά Τραγούδια της Ρόδου, (Δεύτερη έκδοση)*. Πρίσμα, Ρόδος (in Greek).
- Ζερβός, Γ. Κ. (1956-1957), 'Δημοτικά Τραγούδια και Λαϊκά ή Στιχοπλοκές Καλύμνου' στο *Δωδεκανησιακόν Αρχεῖον τόμ. Β'*, pp. 241-266 (in Greek).
- Ιωαννίδου-Μαπαρμπαρίγου, Μ. Χ. (1960-1961), 'Νεώτερες Μορφές Δημοτικών Τραγουδιών στην Κάρπαθο' στο *Δωδεκανησιακόν Αρχεῖον τόμ. Δ'*, 95-101 (in Greek).
- Καζαβής, Γ. (1940), *Νισύρου Λαογραφικά*, Νέα Υόρκη (in Greek).
- Καμμάς, Π. (1947), 'Δημοτικά Τραγούδια της Τήλου' στο *Δωδεκανησιακή Επιθεώρηση I* (in Greek).
- Καμμάς, Π. (1963), 'Δημοτικά Τραγούδια της Τήλου' στο *Δωδεκανησιακόν Αρχεῖον τόμ. Δ'*, pp. 270-274 (in Greek).
- Κοντοσόπουλος, Γ. Ν. (2001), *Διάλεκτοι και ιδιώματα της Νέας Ελληνικής, Εκδόσεις Γρηγόρη, 3η έκδοση*, Αθήνα (in Greek).
- Μακρής, Μ. (1983), *Δωδεκανησιακά Δημοτικά Τραγούδια Ανθολογία*, εκδόσεις Πρίσμα, Ρόδος (in Greek).
- Μαυρής, Ν. Μ. (1961-1962), 'Δημοτικά Τραγούδια της Κω' στο *Δωδεκανησιακόν Αρχεῖον τόμ. Δ'*, pp. 245-269 (in Greek).
- Μαυρής, Ν.; Παπαδόπουλος, Ε. (1950), *Δωδεκανησιακή Λύρα, τομ. Α', Κασιακή Λύρα*, Πορτ Σαΐδ (in Greek).
- Μιχαηλίδη-Νουάρος, Μ. Γ. (2001), *Καρπαθιακά Δημοτικά Άσματα*, εκδόσεις επιστημ. βιβλίων & περιοδικών, Θεσσαλονίκη (in Greek).
- Μπαζιανάς, Ν. (1997), *Γιά τη Λαϊκή μας Παράδοση Μικρά Μελετήματα*, εκδόσεις τυπωθήτω, Αθήνα (in Greek).
- Μπογιανός, Θ. (2001), *Τραγουδώντας στα Δωδεκάνησα (Γνωστά & άγνωστα, όμορφα & παλιά τραγούδια του γάμου)*, Ρόδος (in Greek).
- Παπαϊγναντίου, Δ. (1981), 'Δωδεκανησιακό Τραγούδι και Ποίηση, Ανεπιτήδευτος στίχος, Πρωτοποριακή Μουσική' στα *Πολιτιστικά Δωδεκάνησου, Εισηγήσεις και Πορίσματα Α' Πολιτιστικού Συμποσίου Δωδεκάνησου '78*, εκδ. Στέγη Γραμμάτων και Τεχνών Δωδεκάνησου, σειρά αυτοτελών εκδόσεων, II, Αθήνα, pp. 186-192 (in Greek).

- Παπαχριστοδούλου, Χ. Ι. (1959, 1961), 'Δημοτικά Τραγούδια της Ρόδου', στα *Λαογραφία* 18, 257-232, 19, pp. 96-158 και ανάτυπο (in Greek).
- Πολίτης, Ν. Γ. (1991), *Εκλογαί από τα Τραγούδια του Ελληνικού Λαού*, εκδόσεις γράμματα, Αθήνα (in Greek).
- Σακελλαρίδης, Κ. (1965), 'Το Νισυρικό Δημοτικό Τραγούδι', στο *Νισυριακά τόμος Β'* 13-202 (in Greek).
- Σμυρνάκης, Γ. (1947), 'Δημώδη πατμιακά Άσματα (εκ των καταλοίπων)', στο *Δωδεκανησιακή Επιθεώρηση* 1, 26-29, 64-68 (in Greek).
- Ταρσούλη, Α. (1947), *Τά Δωδεκάνησα*, τόμος Α', Β', Γ', Αθήνα (in Greek).
- Ταρσούλη, Α. (1996), *Αστυπάλαια, ανάτυπο της «Αστυπάλαιας» 2ο βιβλίο. «Τα Δωδεκάνησα»* 1. Μ. Σκαζίκη, εκδόσεις Άλφα (in Greek).
- Χαβιράς, Μ. Δ. (1884-1885), *Δημώδης Ποίηση της Σύμης η' Συλλογή Συμαϊκών Δημοδών Ασιμάτων*, *Κων. Φιλολ.* 19, pp. 208-247 (in Greek).
- Χαλκιάς, Γ. Α. (1980), *Μούσα Ολύμπου Καρπάθου*, Αθήνα (in Greek).
- Χαλκιάς, Ν. Ι. (1994), *Οι άνθρωποι και τα Τραγούδια της Ολύμπου*, εκδόσεις Χαλκιά Ν. Ι., Ρόδος (in Greek).
- Χατζηθεοδώρου, Γ. (1989), *Τραγούδια και Σκοποί στην Κάλυμνο*, αναγν. Καλύμνου «Αι Μούσαι», σειρά αυτοτελών εκδόσεων, Κάλυμνος (in Greek).