Online Dictionary of Idioms

Irina Lobzhanidze

Ilia State University

e-mail: irina_lobzhanidze@iliauni.edu.ge

Abstract

The first edition of this dictionary was prepared under the financial support of the Shota Rustaveli National Science Foundation (Project No Y-04-10). The Dictionary was bilingual in Modern Georgian and Modern Greek, included approximately 12000 entries and reflected how the idioms function. It combined features of translational and learner's dictionaries. The dictionary was not published. Thus, it was decided to compile an online dictionary of idioms both monolingual and bilingual equipped with a friendly environment to find idioms, their meanings and context of use. The project was prolonged with the support of the Shota Rustaveli Science Foundation (Project No LE/17/1-30/13). At this moment, the beta version of dictionary is available at http://idioms.iliauni.edu.ge/.

Present paper is a part of the research project dedicated to the compilation of an online dictionary of Idioms, bidirectional in two languages: Modern Georgian and Modern Greek. The paper is structured as follows: Section 1 provides brief descriptions of Dictionaries and briefly summarizes their prerequisites; Section 2 represents the methodological approach; Section 3 includes findings and hypothesis, which refer to distinguishing features, attributes of dictionaries and characteristics of idioms and Section 4 summarizes the work implemented.

Keywords: online dictionary; idioms; Georgian; Modern Greek

1   Brief Descriptions of Dictionaries

Any kind of dictionary can be considered as a database; generally, its purpose is to store information and to allow user to find appropriate language units. The compilation of any dictionary includes the sequence of stages. In our case, we had to: determine the form of online dictionary and the structure of entry, revise the existing units using the concordance from the corpus of Modern Georgian Language[1] and additional one created in TLex system,[2] add revised and new entries to TLex system, convert the prepared dictionary to .xml format and launch an online version of dictionary.

1.1 Monolingual Dictionary of Idioms

The online edition of monolingual dictionary was compiled on the basis of Modern Georgian-Modern Greek dictionary of Idioms. For the monolingual type of dictionary, we have changed the structure of the dictionary entry and worked out some new entries. At the same time, the decreased number of entries to approximately 8300 allowed us to reflect some changes in Modern Georgian language represented in the works of authors from the twentieth century till today. Some of the idioms have been revised and checked on the basis of language corpora. Citations come not only from the classics of Georgian literature, but from Modern Georgian writers as well. At this moment

---

[1]   http://corpora.iliauni.edu.ge/
[2]   http://tshwanedje.com/

the beta version of dictionary is available at http://idioms.iliauni.edu.ge/.

1.2 Bilingual Dictionary of Idioms

The bilingual bidirectional dictionary of idioms has not been finished yet. This year we have to launch the beta version of its Modern Georgian-Modern Greek part only. The structure of the dictionary entry slightly differs from those presented at the 10th International Conference of Greek Linguistics in 2011 (Lobzhanidze 2012). Especially, each idiom will be accompanied by essential grammar information. We hope that this information will allow user to understand the sense of idiom both for standard and nonstandard contexts.

2   Methods

Following approaches of modern corpus based lexicography (Atkins 2008, Sinclair 1996, Ooi 1988 etc.), a compilation of dictionary is based on the corpus of Modern Georgian Language available at corpora.iliauni.edu.ge and additional one created in TLex system. The online environment is created with php.

3   Findings and Hypothesis

3.1 Idiomaticity

In this dictionary an idiom is considered as a phraseological unit (PU) consisting of two or more words that act as a whole from the semantic point of view. Generally, the meaning of an idiom can't be predicted from the meaning of its components. The dictionary includes idioms, traditional idioms as presented in Chikobava 1950-1964, 2008; Sakhokia 1979; Oniani 1966 etc. and some other PUs whose idiomacity varies according to the hierarchy of 'frozenness'.

3.2 Characteristics of Idioms

Thus, the majority of entries includes idioms that function as the following part of speech:

a) Noun phrase, e.g. *მგლის ლუკმა* 'lit. wolf's bite; ~smth. or smb. who can be dropped dead, wolf's carrion', *სიზიფეს შრომა* 'lit. Sisyphus work; ~a thankless task, Sisyphean task', *შეჭმული ფული* 'lit. eaten money; ~squandered money' etc.

b) Verb phrase, e.g. *თავლაფს ასხამს* 'lit. pours dirt on head; ~to stain smb's good name, to shame smb.', *ჯახანს   გააღებს* 'lit. makes crashing sound; ~will hit out at smb., attack smb.', *წყლივით იცის* 'lit. knows as water; ~to know smth like the back of one's hand' etc.

c) Adverb phrase, e.g. *კოტაზედ (ნურას) უკაცრავად* 'lit. no pardon; ~you are not right', *უკან მიუხედავად* 'lit. without looking back; ~as fast as one can' etc.

d) Idioms with postpositions, e.g. *ბედნიერ ხელზე* 'lit. on the happy hand; ~bringing good luck', *ქვეყნის შარაზე* 'lit. on the main road of country; ~to make smth. public',*ენაზე აკერია* 'lit. smth. is sewn on the tongue; ~smb. cannot stop talking about smth.' etc.;conjunctions, e.g. *დღეს თუ ხვალე* 'lit. today or tommorow; in the near future', *დიდი თუ პატარა* 'lit. adult or tiny; all', *დუღს და გადმოდუღს* 'lit. boils and boils over; has more money than one knows what to do with' etc.

e) Formula phrases used in different communication situations like greetings, blessings, curse etc. e.g. *შენი რისხვა არ მომეცეს* 'lit. not give your anger to me; ~may you not be angry with me, I swear to you', *ვაი ჩემი ბრალი* 'lit. woe, my fault; I will be blamed', *ღვთის ცეცხლი!* 'lit. fire of God; ~the wrath of God' etc.

f) and some other PUs, which are not strictly idiomatic, but according to the Georgian lexicographic tradition are considered as idioms, e.g. *ცუდს იზამს* 'will harm smb.', *კარგად გახდება* 'will become well' etc.

Generally, the above-mentioned groups have the following features distinguishing them from other PUs:

a) Defective paradigms in case (*ქუდზე კაცი* 'lit. a man for hat; ~all able-bodied man', *ცისაბრისა* 'lit. sky dust; ~this and that' etc.), number (*ბრმა იარაღი* 'lit. blind weapon; ~blind force', *ნიანგის ცრემლები* 'lit. crocodile tears; ~a false, insincere display of emotion' etc.), TAM (*ტყემალზე ზის* 'lit. is sitting on a plum tree; ~not to have a clue what's going on', *ჭურში ზის* 'lit. is sitting in the wine-jar; ~to be in a world of one's own' etc.), version (*გულისყურს ადევნებს* 'lit. is keeping an eye of heart; ~to pay attention to smth.', *თავს იგდებს* 'lit. is throwing a head; ~to be insolent', *ყურს უგდებს* 'lit. is throwing an ear; ~to listen to smb.' etc.), aspect (*სულს ამოუღებს* 'lit. removes a soul; ~to despair, to trouble smb.', *ძალას ჩაუდგამს* 'lit. will give strength to smb.; ~to give strength to smb.' etc.)etc;

b) Some idioms preserve archaic forms e.g. *ხე ცნობადისა* 'Tree of Knowledge', *მარილი ამა ქვეყანისა* 'the salt of the earth' etc.;

c) The majority of verbal idioms can be used with negative particles არ or ვერ, but in some cases the idiomacity is caused only by their use, e.g. *თავზე ბუზს არ ისვამს* 'lit. won't let a fly land on one's head; ~to be haughty', *ვერ მიართვეს* 'lit. didn't give; ~to fail', *თავი არა მაქვს* 'lit. I haven't a head; ~not to be able to do smth.' etc.

3.3 Grammar

Generally, it is supposed that the user knows about Modern Georgian and Modern Greek grammar. But, the several points are worth mentioning, especially:

- The search criteria for the majority of dictionaries worldwide are based on alphabet order of entries and the initial point of word. Such kind of criteria are very useful for the languages, where the inflectional changes don't affect the initial point of word. Generally, it is sufficient to type initial letters and to find the concrete word or word with similar stem. Georgian Language is similar to European languages regarding the structure of nouns, adjectives etc., but there is a big problem with such kind of search in case of verb, where a variety of prefixes play an active role in the formation of verbal structure (e.g. personal markers, version markers etc.);

- The majority of entries can be considered as verbal idioms (6704 entries). The traditional Georgian Lexicography accepts two possibilities of representing verbal idioms: either masdar (Sakhokia 1979) or the third person singular present (Chikobava 1950-1964). We have decided that it would be better to use the third person singular for two reasons: a) in the majority of idioms Georgian language never uses masdar; b) the third person singular is considered as a lemma for the analyzer of Modern Georgian language;

- The Georgian verb is polypersonal. The grammatical subject may be in the nominative, the ergative or the dative cases, the grammatical direct or indirect object may be in the nominative or the dative cases. Thus, the dictionary includes different verbal patterns;

- The online version of dictionary is connected to the analyser of Modern Georgian language and equipped with a new search function, which allows user to find all those idioms, which undergo concrete grammatical changes and to avoid the problem mentioned above.

3.4 The Dictionary Entries

The dictionary entries for monolingual and bilingual parts differ in the elements represented. Most entries include information on lemma sign, derivational variants of use, etymological notes for some entries, definitions, literary citation with indication of literary source. The polysemous meanings of idioms are listed as well. Noun phrases, generally, are used in singular, but in case of defective paradigms – in plural. Verbal phrases are used in finite forms as it was mentioned above.

3.4.1 Ordering of Entries

In the TLex entries are ordered alphabetically. The similar ordering is provided online. At the same time, if a user wants to find any of constituents (not the first word of multiword expression), it is allowed directly from the web, e.g.
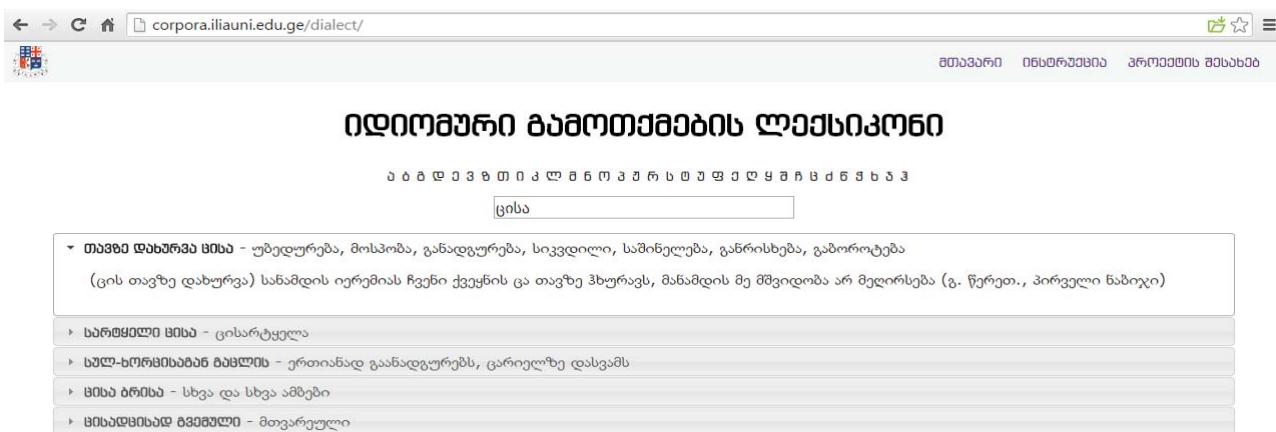

Figure 1: Dictionary entry.

One of the problems that should be considered additionally is the problem of different verbal forms, which are allowed in idioms, e.g. *გულს გაუტეხს* 'lit. will broke smb. heart; ~will disappoint, discourage smb.' can be reached by entering initial letters of words *გული* 'heart' or *გაუტეხს* 'to break' or directly by entering words:
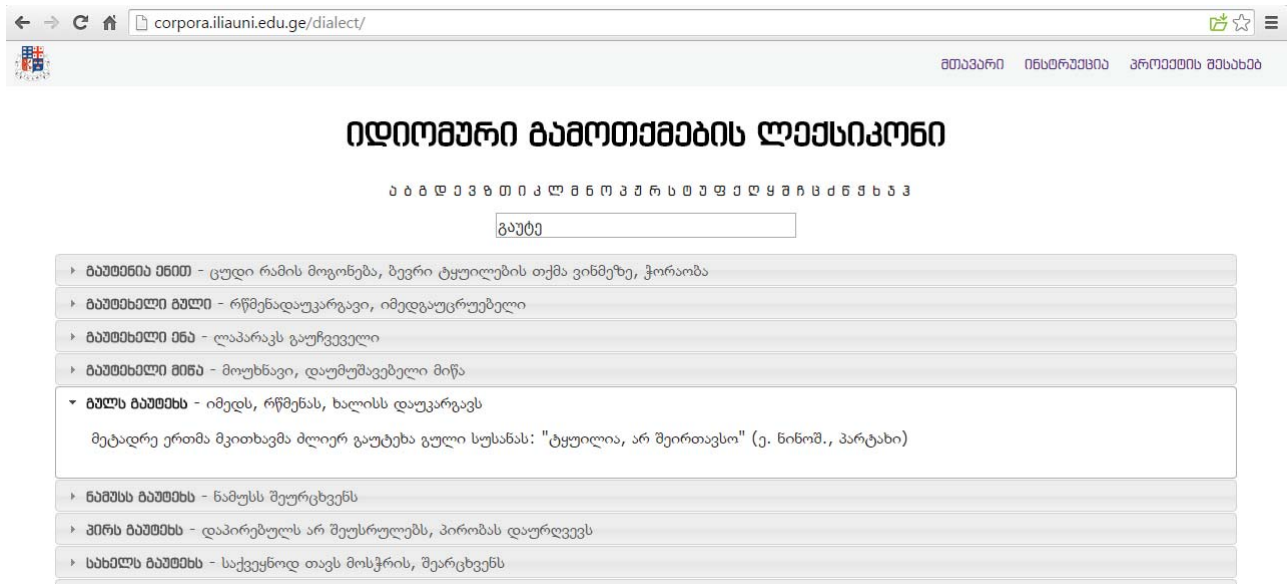
Figure 2: Dictionary entry.

But Modern Georgian allows forms like: გული გამიტეხა 'he/she disappointed me', გული გავუტეხე 'I disappointed him' etc. Such kind of forms can be seen in literary sources as well and the user whose knowledge of Modern Georgian is not very high will not be able to find them. The decision to the above-mentioned problem is under consideration yet, but there is a possibility to find lemma sign for a words not listed in the dictionary, which can be achieved by using the morphological analyzer of Modern Georgian language.

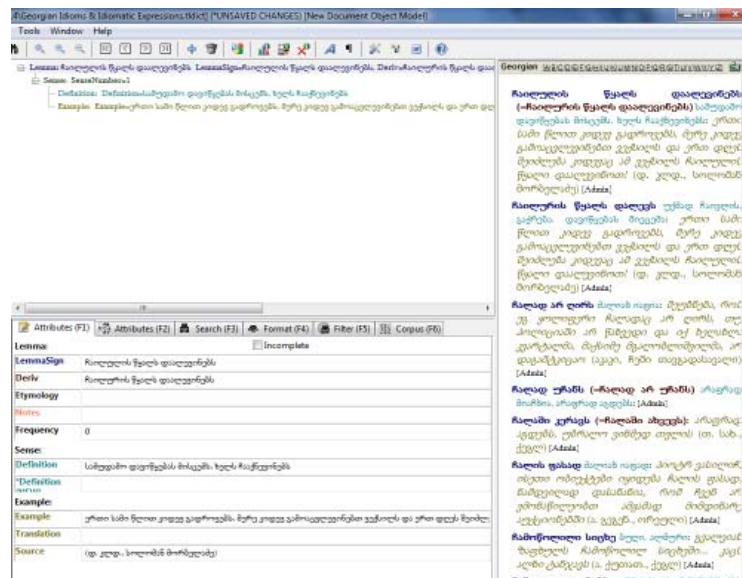***Sample Entries for Monolingual Dictionary***

Figure 3: Monolingual Dictionary entry in TLex.

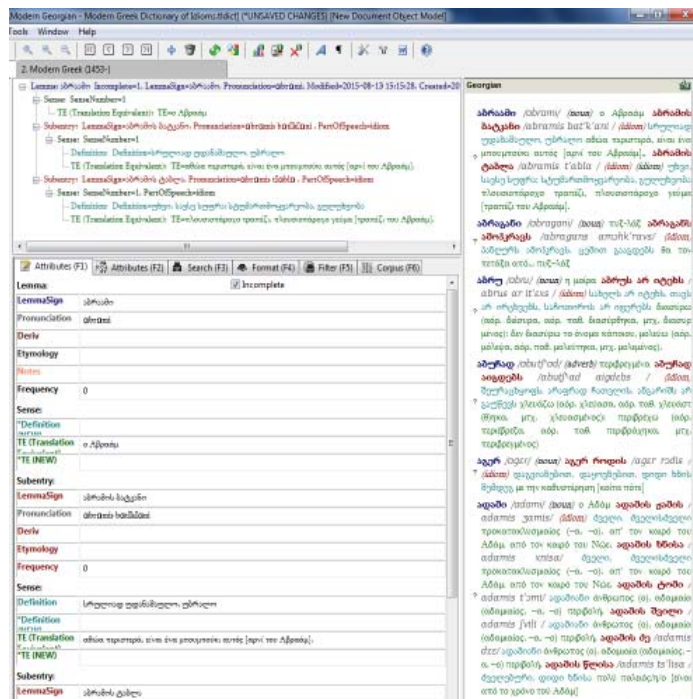*Sample Entries for Bilingual Dictionary*

Figure 4: Bilingual Dictionary entry in TLex.

3.4.2 Greek Equivalents

The majority of entries contain more than one Modern Greek equivalents, since different contexts require different translations of a given idiom. If possible, equivalents are presented in the same form as Georgian idiom, i.e. Modern Georgian-Modern Greek noun phrase, verb phrase etc. In cases when it is impossible to present the Modern Greek equivalent of a Modern Georgian idiom, the translation equivalent field gives explanation of idiom.

3.5 Citations and Examples

The monolingual part of dictionary is illustrated by citations from Modern Georgian literature and information about its source. Sources of Georgian idioms include approximately 256 Georgian writers; numerous contexts provided by the Corpus of Modern Georgian language; all available monolingual phraseological dictionaries.

4   Conclusions

Taking into account that we have finished compilation of the monolingual part of dictionary, we present the macrostructure and microstructure of the dictionary and describe the further stages of its development. The beta-version of dictionary is available at http://idioms.iliauni.edu.ge/.

5   References

Atkins, S.B.T., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography.* Oxford: Oxford University Press.

Carter, J. M. (1996). Corpus to Corpus: A Study of Translational Equivalence. In *International*

Journal of Lexicography, 9(3), Oxford: Oxford University Press, pp. 171-178.

Ooi, V. B. (1998). *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.

Lobzhanidze, I. (2012). Περί προβλημάτων της κατάρτισης του Γεωργιανό-Νεοελληνικού Λεξικού Ιδιωτισμών [To the compilation of Georgian – Modern Greek Dictionary of Idioms]. In *Proceedings of the 10th International Conference of Greek Linguistics.* Komotini: The Democritus University of Thrace, pp. 899-904.

Rayfield, D., Apridonidze, Sh., Margalitadze, T., etc. (2006). *A comprehensive Georgian-English Dictionary*. Garnett Press.

Stavropoulos, A. H. (1998). *Oxford English-Greek Learner's Dictionary.* Oxford University Press.

Βλαχόπουλος, S. (2007). *Λεξικό των ιδιωτισμών της νέας ελληνικής* [Modern Greek Dictionary of Idioms]. Αθήνα: Κλειδάριθμος.

Γιαννουλέλλης, Γ. (1982). *Νεοελληνικές ιδιωματικές λέξεις δάνειες από ξένες γλώσσες* [Modern Greek Idioms borrowed from Foreign Languages]. Αθήνα: Κείμενα.

Δημητρίου, Α. (1995). *Λεξικό νεοελληνισμών: Ιδιωτισμοί, στερεότυπες μεταφορές και παρομοιώσεις, λέξεις και φράσεις από την καθαρεύουσα* [Modern Greek Dictionary: idioms, standard metaphors and proverbs, words and expressions from Kathareousa].Αθήνα: Γρηγόρη.

Κριαράς, E. (1995). *Νέο ελληνικό λεξικό της σύγχρονης δημοτικής γλώσσας, γραπτής και προφορικής* [Modern Greek Dictionary of written and spoken language]. Αθήνα: Εκδοτική Αθηνών.

Παπαλέξης, Αντ. Α. (1988). *Λεξικό ομόηχων λέξεων και τονικών παρωνύμων: απο το Ομηρο μέχρι σήμερα* [Dictionary of similar words and proverbs: from Homer since today]. Αθήνα: Δημόκριτος.

Σκούρτης, Α. (1988). *Λεξικό ξενόφερτων λέξεων στην ελληνική γλώσσα* [Modern Greek Dictionary of unusual words]. Αθήνα: Πύρινος Κόσμος.

Δαγκίτσης, Κ. (1970). *Λεξικό των συνωνύμων της νεοελληνικής* [Modern Greek Dictionary of Synonyms]. Αθήνα: Βασιλείου.

Μαρκαντωνάτος, Γ. (2002). *Λεξικό Αρχαίων, βυζαντινών και λογίων φράσεων της Νέας Ελληνικής* [Modern Greek Dictionary of ancient, Byzantine and scholar words]. Αθήνα: Gutenberg – Γιώργος & Κώστας Δαρδανός.

Μπαμπινιώτης, Γ. (1998). *Λεξικό της Νέας Ελληνικής γλώσσας* [Modern Greek Dictionary]. Αθήνα: Κέντρο λεξικολογίας.

Acknowledgements