

The Comprehensive English-Georgian Online Dictionary: Methods, Principles, Modern Technologies

Tina Margalitadze

Keywords: *English-Georgian dictionary, structure content, software.*

Abstract

The aim of the paper is to present the lexicographic project completed in I. Javakhishvili Tbilisi State University, namely the Comprehensive English-Georgian Online Dictionary.

Conceived back in the 1960s of the previous century at the Chair of English Philology of the University, the dictionary project has gone through many difficult stages: erroneous decisions about principles of compilation of dictionary entries, incorrect sources chosen for the dictionary, lack of experience of lexicographic work at an educational institution, no financing, etc. In the 1980s a small team of editors embarked on thorough revision of the dictionary material and started publication of the dictionary in fascicles, on a letter-by-letter basis. The online version of the dictionary, posted in the Internet in 2010, is based on the mentioned fascicles.

The paper discusses the macrostructure of the dictionary, considerations behind the selection of the word-list for the dictionary; principles of presentation of homonyms, converted forms, polysemy; exemplification policy, as illustrative phrases and sentences constitute a very important component of dictionary entries. The paper pays special attention to the treatment of semantic asymmetry between genetically unrelated and systemically completely different languages as is the case with the Georgian and English languages.

The paper elucidates grammatical, as well as other types of labels employed in the dictionary: temporal (archaic, obsolete), regional (American English, Australian English, etc); formal and informal, spoken words, sociolects and connoted vocabulary are also marked by respective labels (formal, informal, colloquial, vulgar, slang, derogatory, contemptuous, pejorative, etc); specialized terminology has subject-specific labels (anatomy, architecture, astronomy, biology, geography, geology, economics, medicine, metallurgy, philosophy, finance, technical, zoology, etc).

The Comprehensive English-Georgian Online Dictionary is a web-application developed in accordance with modern standards and requirements. The engine of the dictionary is written in PHP scripting language. The dictionary vocabulary and systemic bases are located in MySQL database. Interfaces use some JavaScript. The web-application comprises user, administration and billing functions and interfaces, thus creating an integrated and dynamic resource which provides a unique opportunity to simultaneously use, maintain and administer the dictionary.

1. Background. Sources

The idea to create the dictionary was conceived back in the 1960s at the Chair of English Philology of the University on the initiative of a prominent Georgian scholar and translator, then Head of the Chair, Prof. Givi Gachechiladze. The lack of an academic dictionary was especially acutely felt by translators and perhaps not accidentally, the creation of such a dictionary was primarily determined by the need to adequately translate English language literature into Georgian. Since then, the work on the dictionary has gone through many difficult stages. Many mistakes were made too, mainly due to the fact that the specialists and teachers of English participating in the project lacked proper previous experience of lexicographic work.

A few words on why the work on the Comprehensive English-Georgian Dictionary (CEGD) was undertaken by Tbilisi State University.

By the beginning of the 19th century, Georgia became a part of the Russian Empire. As a result of this, the main emphasis was laid on Russian-Georgian lexicography. In the 19th and 20th centuries, Georgian lexicographers compiled and published many general and specialized Russian-Georgian and Georgian-Russian Dictionaries of various volumes. The Georgian Institute of Linguistics was mainly preoccupied with the compilation and

publication of Russian-Georgian dictionaries and the Explanatory Dictionary of the Georgian Language. At the same time, the number of specialists on West European languages, particularly English language specialists, working for the Institute was comparatively small. This is why the Chair of English Philology of the University embarked on the task of creating the Comprehensive English-Georgian Dictionary. It should be noted that in the same period respective chairs of the University began to work on German-Georgian and French-Georgian dictionaries but, unfortunately, the said dictionaries were not completed and published – a fact which underlines once more how difficult it is for an educational institution to work on a comprehensive dictionary.

When the work on the CEGD had just started, the question of the sources for the dictionary was under consideration. In the 1960s and the 1970s, English-Russian dictionaries translated into Georgian were believed to be good enough to become such sources. This erroneous decision made all preliminary work and translated materials practically useless. The possibility of making up a bilingual concordance was also considered. This way also proved to have no prospects. A bilingual concordance had to rely upon translated literature, while the literary fiction of that period was rarely translated from the original. Instead, Russian translations from English were used as a source. At the same time, the majority of Georgian translators treated original texts rather freely. From the 1980s a small team of dedicated editors shouldered a complete revision of the dictionary material. They arrived at the decision to regard comprehensive English explanatory dictionaries as a source for the English-Georgian Dictionary. Definitions of explanatory dictionaries reflect social experience with respect to the usage of language norms. Recording collective conception of the meaning of a word by native speakers of a language, generalizing knowledge and experience of many generations mirrored in the language, definitions of comprehensive explanatory dictionaries constitute an extremely valuable source for the identification of the semantic structure of a word. Especially noteworthy in this regard are Oxford English Dictionary on Historical Principles (OED) and Webster's Third New International Dictionary, whose definitions have become the basic source for the semantic investigation of English words. The editorial team also relied on the New English-Russian Dictionary by Ilya Galperin, as well as on other lexicographic publications of the Oxford University Press, Cambridge, Longman, COBUILD and other dictionaries. This stage of the work on the dictionary lasted for 25 years. Currently, printed and published are 13 volumes of the CEGD, which cover 2 200 pages of the dictionary proper. In February 2010, an online version of the English-Georgian Dictionary was placed on the Internet.

2. Macrostructure

The selection of the word-list for the CEGD was determined by the target groups for which it was initially intended – namely translators of English belles-lettres and scientific literature into Georgian, specialists working on specialized, branch-specific texts, learners of English, for whom English is their future speciality, also prospective psychologists, physicians, biologists, etc who study English and need to know foreign special terms and their Georgian equivalents, as well as learners of English in general. The intended users of the dictionary determined the inclusion in it of modern vocabulary of the English language (keeping in mind the highest level of language proficiency) characterized by high frequency of occurrence, as well as the inclusion of less frequently used words, rare, obsolete, archaic, dialectal words or rare, obsolete, archaic and dialectal meanings of modern words. The dictionary has included terms from almost all fields of knowledge. Currently the CEGD comprises 110 000 entries.

3. Semantic asymmetry between English and Georgian words

One of the important issues faced by the editors of the CEGD has been linguistic and cultural anisomorphism between English and Georgian languages, resulting in semantic asymmetry of seemingly similar words. English-Georgian lexicography is not exceptional in this respect, as it is the central problem of bilingual lexicography at large. This problem is well-formulated in the definitions of ‘**equivalence**’ and ‘**equivalent**’ in the Dictionary of Lexicography by Hartmann and James. ‘Because of linguistic and cultural anisomorphism, translation equivalents are typically partial, approximative, non-literal and asymmetrical (rather than full, direct, word-for-word and bidirectional). Their specification in the bilingual dictionary is therefore fraught with difficulties, and recourse must be made to surrogate explanatory equivalents’ (1998: 51).

This semantic asymmetry is even wider between genetically unrelated and systemically completely different languages as is the case with the Georgian and English languages. The Board of Editors developed different techniques for dealing with this problem back in the 1980s independent of the lexicographic theories and practices of European lexicographers formulated in the works of Hartmann (2007), Zgusta (2006) and others (these works are thoroughly reviewed by Arleta Adamska-Salaciak in her plenary lecture at the 14th EUALEX congress, reprinted later in International Journal of Lexicography (2010).

The goal of the editors of the CEGD has always been to describe the equivalence between the English and Georgian languages on a more general, systemic level of the two languages, rather than give only Georgian equivalents of English words found in specific, particular contexts. This is, probably, a novelty in English-Georgian lexicography, introduced by our team. The editors of the CEGD have developed a technique of combining explanatory equivalents of lexical units with translational equivalents presented in different illustrative phrases and sentences selected for entries.

There are numerous instances of English-language concepts not lexicalized in Georgian and vice versa. For example:

brocket ‘a stag in its second year’
hearst ‘a hind of the second or third year’
hart ‘a male deer after its fifth year’

(cf. Georgian **nukri** ‘a young of a deer, gazelle and some other wild animals’; **duraqi** ‘one-year-old ibex’; **bitla** ‘ibex buck under 5 years old’, etc). Such non-lexicalized concepts may be lexical units of polysemous words. For example:

elastic
(a) ‘capable of recovering quickly from low spirits, disappointment, or misfortune’
(b) ‘capable of ready change or easy expansion; not rigid or constricted’

There is no Georgian word, matching or approximating the above-cited definitions. Such English words or lexical units have been rendered into Georgian by means of Georgian definitions, definitional / explanatory equivalents. This type of explanatory equivalents was also widely employed by the team of the Comprehensive Georgian-English Dictionary under the general editorship of Donald Rayfield (2006). For example:

zedsidze ‘son-in-law leaving with wife’s parents or working for them’.

Concerning semantic asymmetry of seemingly similar words between English and Georgian, this asymmetry may be manifested in many different ways. Very often the semantic structure of an English word corresponds to some part of that of a Georgian word, whereas its other parts relate to completely different Georgian words, requiring the use of different Georgian words in process of translation. The proportion of semantic overlapping varies to different degrees.

There are numerous cases when English words have more general meanings than their Georgian counterparts, comprising several Georgian words. For example:

squawk ‘a loud grating call or cry of birds’.

This cry can be uttered by parrots, peacocks, seagulls, etc. Such cases have been usually treated with a short definition in Georgian (e.g. ‘a loud cry’), followed by specific, contextual equivalents (in this case onomatopoetic words: *qiqini, kivili, tchqivili, chkhavili*), or with a short definition in Georgian and translational equivalents given in different illustrative phrases and sentences.

It should be noted that illustrative phrases and sentences are considered to be a very important component of entries in CEGD. The editors of the CEGD have always regarded illustrative material as an important component of highlighting the meaning of English words. Meaning has many facets. Illustrative phrases and sentences are selected so as to reveal all shades of the meaning of a lexical unit, all its facets, which require different translational equivalents in Georgian.

Explanatory equivalents are not always definitions. Often more general equivalents of English words have been selected, one or several. For example:

to mar is rendered in Georgian by general equivalents: *gaputcheba, dazianeba* (‘to spoil’, ‘to damage’). In the illustrative phrases: *to mar one’s joy; to mar one’s happiness; small-pox marred her face; smth. mars the beauty of the landscape* - the entry word **to mar** is translated by different contextual equivalents: *chamtsareba* (lit. ‘embittering’); *chrdilis miqeneba* (lit. ‘casting shadow’); *dakenkva* (lit. ‘disfiguring with pock-marks’); *daushnoeba* (‘making smb., smth. ugly’). These translational equivalents provide good, literary translations of English phrases and sentences into Georgian in particular contexts, thus providing different single lexical equivalents for the English word **to mar**.

There are few instances of parallel polysemy between Georgian and English words. If one compares European languages, from this point of view, more parallelism can be observed between polysemous words, due to more convergences in related languages, mutual influence of these languages on one another, etc. For example:

French **retirer** and English **withdraw** have more parallel forms than Georgian **gamoqvana** and its English counterpart.

Retirer (*troupes*) – **withdraw** (*troops*);

retirer (*soutien*) – **withdraw** (*support*);

retirer (*projet, candidature*) – **withdraw** (*candidature*);

retirer (*argents*) – **withdraw** (*money*), etc.

Georgian translational equivalents for English **withdraw** will be *gamoqvana* (lit. ‘bringing / taking out’), *tsamoqvana* (lit. ‘bringing over’), *gamosvla* (lit. ‘coming out’), *ukanve tsagheba* (lit. ‘taking back’), *gamotana* (lit. ‘drawing from’), *amogheba* (lit. ‘taking out’), etc. This fact accounts for the decision of the editors to present polysemous meanings of English words to the maximum extent possible with the whole panoply of their equivalents in Georgian.

English adjectives and verbs often have specifying notes like: *said of the sea, weather, wind, ground, journey, places, persons, language, etc* in the entries of the CEGD.

Combination of explanatory equivalents with contextual, translational equivalents, careful selection of illustrative phrases and sentences help editors of the CEGD to reveal all shades of meanings of English words. Perhaps, such approach explains the positive feedback the editorial board of the CEGD constantly receives from the readers (especially translators), who lay special emphasis on the fact that the CEGD greatly helps them in perceiving even the slightest nuances of meaning and in rendering them into Georgian. As for learners of English, CEGD serves as a kind of bilingualized dictionary for them and, as students of different universities point out, the CEGD facilitates not only proper comprehension and production of English words but also assists them in improving their command of the literary Georgian language.

4. Homonymy. Conversion. Polysemy

Homonyms are treated as separate entries in the CEGD and are marked with small homonym numbers, for example:

bay¹ [beɪ] *n* **1.** უბე, ყურე; ...

bay² [beɪ] *n* **1.** მშენ. 1) მალი (სვეტებს შორის და ა.შ.); ...

Forms produced by conversion are also represented in the dictionary as separate entries and are distinguished by Roman numerals, for example:

bay³ I [beɪ] *n* **1.** ხმამალალო ყეფა; ...

bay³ II [beɪ] *v* **1.** ხმამალლა ყეფა (ყეფს); ...

bay¹ *noun* უბე, ყურე (ზღვისა); ◊Bay State ამერ. "შტატი..."

bay² *noun* 1. მშენ. 1) მალი (სვეტებს შორის და ა.შ.); 2) ხილის...

bay³ I *noun* 1. ხმამალალო ყეფა (ნადირის ღვენაში მყოფი ძალღებისა);...

bay³ II *verb* 1. 1) ხმამალლა ყეფა (ყეფს) იოქმის ნადირის ღვენაში...

As mentioned above, CEGD fully represents the polysemy of English words. In order to distinguish between polysemantic meanings of words, a double numeration system is used: there are different modes of numbering for main meanings [1. 2. 3.] and sub-meanings [(1) 2) 3)]. Cases of polysemy of phraseological units are marked out by letters of the Georgian alphabet.

5. Grammatical and Other Labels

Headword is followed by a phonetic transcription and a part of speech label. Entries contain irregular forms of nouns, adjectives, adverbs, verbs. These irregular forms are also entered as headwords and cross-referenced to major forms. Entries also have other types of labels: temporal (*archaic, obsolete*), regional (*American English, Australian English, etc*). Dialectal words, as well as poetic and literary words are marked in the dictionary. Formal and informal, spoken words, sociolects and connoted vocabulary are also marked by respective labels

(formal, informal, colloquial, vulgar, slang, derogatory, contemptuous, pejorative, etc). All foreign words have corresponding labels. Specialized terminology has subject-specific labels (*anatomy, architecture, astronomy, biology, geography, geology, economics, zoology, horticulture, medicine, metallurgy, apiculture, philosophy, psychology, finance, technical, etc*).

6. Software

The CEGD is a web-application developed in accordance with modern standards and requirements with Internet connection and standard web-browser being all one needs to use it. The engine of the dictionary is written in PHP scripting language. The dictionary vocabulary and systemic bases are located in MySQL database. Interfaces use some JavaScript. The web-application comprises user, administration and billing functions and interfaces, thus creating an integrated and dynamic resource which provides a unique opportunity to simultaneously use, maintain and administer the dictionary.

In addition to standard features and pages, the user side includes the search formula which enables to effectively look up both English and Georgian words despite the fact that the dictionary vocabulary has one-way – English-Georgian base. Search functions also take into consideration possible spelling errors and in case of an actual spelling mistake the search system will suggest to the user a supposed correct spelling version of the word.

After a long period of planning and testing, the system functions in the way which enables to look up and retrieve the desired data from the vocabulary base, including up to 200 000 rows and weighing up to 100MB, in a very short period of time, in most cases – virtually in a split second.

Control Panel of the dictionary performs multiple basic functions, including:

- The functions of viewing and editing the dictionary vocabulary, as well as the function of adding new entries;
- The logs of searches done by the dictionary users;
- The tools necessary for editors, such as generators and converters; and
- Tables showing registrations, payments and statistics, as well as various informative tables.

Also embedded in the System is a specially designed billing system enabling both users and administrators to view and manage subscription fee payments. The dictionary billing system interacts with paybox terminals and plastic card payment systems.

Both the database and the engine of the dictionary are in the process of constant upgrading and improvement in order to provide the users with a renewable, up-to-date, user-friendly, safe and perfect product.

7. User's Guide

The dictionary is supplied with a user's guide, explaining in details the structure and organization of the entries. It also contains a technical guide with four videos, providing instructions for the search of English words, collocations, phrasal verbs, idioms, as well as Georgian words and phrases.

References

A. Dictionaries

- Cambridge Dictionaries Online*. <http://dictionary.cambridge.org>.
- Chikobava, A. et al. (eds.) 1950-1964.** *Explanatory Dictionary of the Georgian Language*. Tbilisi: Publishing House of the Georgian Academy of Sciences. (EDGL)
- Collins Dictionaries Online*. <http://www.collinsdictionary.com>.
- Comprehensive English-Georgian Dictionary*. (Online version.) February 2010. <http://dictionary.ge>. (CEGD)
- English Cobuild Dictionary Online*. <http://dictionary.reverso.net/english-cobuild>.
- Hartmann, R. R. K. and G. James. 1998.** *Dictionary of Lexicography*. London: Routledge.
- Galperin, I. R. et al. (eds.) 1972.** *New English-Russian Dictionary*. (In two volumes.) Moscow: Soviet Encyclopaedia Publishing House
- Longman Dictionary of Contemporary English*. <http://www.ldoceonline.com>. (LDOCE)
- Macmillan Dictionary and Thesaurus: Free English Dictionary Online*. <http://www.macmillandictionary.com>.
- Margalitadze, T. (ed.) 1995-2008.** *Comprehensive English-Georgian Dictionary*. (Fascicles I-XIII) Tbilisi: Lexicographic Centre at Tbilisi State University. (CEGD)
- Merriam-Webster Dictionary Online*. <http://www.merriam-webster.com>.
- Oxford Advanced Learner's Dictionary*. (7th Edition.) Oxford: Oxford University Press. (OALD 7)
- Oxford English Dictionary*. (Second edition on CD-ROM. Version 2.0) (OED)
- Oxford English Dictionary on Historical Principles*. (Vol. I – XIII; suppl. I – IV.) Oxford: Oxford University Press. (OED)
- Procter, P. (ed.) 1995.** *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.
- Shorter Oxford English Dictionary 2002, 2007*. (Sixth Edition.) Oxford: Oxford University Press. (SOED)
- Shorter Oxford English Dictionary 2002, 2007*. (Sixth Edition, electronic version.) Oxford: Oxford University Press. (SOED)
- Sinclair, J. (ed.) 1987.** *Collins COBUILD English Language Dictionary*. London: Collins.
- Stevenson, A., C. Soanes, J. Pearsall and P. Hanks 2005.** *Oxford Dictionary of English*. (Second Edition, Revised.) Oxford: Oxford University Press.
- Summers, D. 1992.** *Longman Dictionary of English Language and Culture*. Harlow: Longman.
- Webster's Third New International Dictionary 1981*. (Unabridged.) Merriam Webster Inc.
- ### B. Other literature
- Adamska-Salaciak, A. 2010.** 'Examining Equivalence.' *International Journal of Lexicography* 23.4: 387–409.
- Hartmann, R. R. K. 2007.** *Interlingual Lexicography: Selected Essays on Translation Equivalence, Contrastive Linguistics and the Bilingual Dictionary*. Tübingen: Max Niemeyer.
- Zgusta, L. 2006.** *Lexicography Then and Now*. Tübingen: Max Niemeyer.