



HAL
open science

Kartu-Verbs: A Semantic Web Base of Inflected Verb Forms for the Georgian Language

Mireille Ducassé

► **To cite this version:**

Mireille Ducassé. Kartu-Verbs: A Semantic Web Base of Inflected Verb Forms for the Georgian Language. [Research Report] IRISA. 2019. hal-02460975

HAL Id: hal-02460975

<https://hal.science/hal-02460975>

Submitted on 30 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Kartu-Verbs: A Semantic Web Base of Inflected Verb Forms for the Georgian Language

IRISA Technical Report. SemLIS Team. December 2019

Mireille Ducassé, mireille.ducasse@irisa.fr
Univ Rennes, INSA Rennes, CNRS, IRISA - UMR 6074, France

Abstract

The Georgian language has a complex verbal system, both agglutinative and inflectional, with many exceptions. It is still a controversial issue to determine which lemmas should represent a verb in dictionaries. Verb tables help neophytes to track lemmas starting from inflected forms but if in paper documents they are tedious and error-prone to browse. We propose Kartu-Verbs, a Semantic Web base of inflected Georgian verb forms. For a given verb, all inflected forms are present. Knowledge can easily be traversed in all directions: from Georgian to French and English; from an inflected form to a masdar, and conversely from a masdar to any inflected form; from component(s) to forms and from a form to its components. Users can easily retrieve the lemmas that are relevant to access their preferred dictionary. Kartu-Verbs can be seen as a front-end to any Georgian dictionary, thus bypassing the lemmatization issues. This report illustrates in detail how to use Kartu-Verbs and give some indications about how the base is built. Our base, in its current state, is already a successful proof of concept. It has proven helpful to learn about Georgian verbs.

1 Introduction

Georgian is a Caucasian language, mother tongue of about 5 million people. It has its own alphabet¹. Georgian grammar has an especially complex verbal system². Without going into the details of the linguistic subtleties (we can, for example, refer to [6, 8]), some illustrations of this complexity are given below from a beginner's perspective.

The number of conjugation rules is very important. There are numerous irregular verbs. What is especially confusing for French or English native speakers is that conjugation can modify both the beginning and the termination of verbs. For example, the verb “to work” (mushaoba - მუშაობა), at the first person plural of present

¹We use a transliteration in Latin characters in this article (and in the base) to ease non-native Georgian speakers's reading. The transliteration is currently “French” oriented for historical reasons.

²<https://fr.wikipedia.org/wiki/Géorgien>.

tense gives “vmushaobt” (ვმუშაობთ). Note the “v” at the beginning of the verb and the “t” at the end. Some tenses, such as future and aorist, often but not always, introduce a preverb. For example, for verb “to work”, the first person singular Present is “vmushaob” (ვმუშაობ) and the first person singular future “vimushaveb” (ვიმუშაებ); for verb “to write” (“dats’era”), (დაწერა) the previous two forms are “vts’er” (ვწერ) and “davts’er” (დავწერ). Note that for the verb “to work”, “ob” has changed into “eb” and a post-radical, “v”, has appeared. Indications of movement directions are given by prefixes. For example, “I go” = “mivdivar” (მივდივარ), “I go down” = “chavdivar” (ჩავდივარ), “I come down” = “chamovdivar” (ჩამოვდივარ); note the different prefixes, the preradical “v” for the first person and the common part “divar”. A dozen prefixes can be used. Regarding the variants of the “to come” and “to go” verbs, in addition to all their prefixes, we can note the different forms from one person to another. For example, “I come” = “movdivar” (მოვდივარ) but “you come” = “modikhar” (მოდიხარ), “she comes” = “modis” (მოდის).

We must also count with verbs whose root changes at different tenses. For example, the third person singular, of verb “to change”, Present: “tsvlis” (root, “tsvl”, ცვლ), Aorist: “shetsavla” (root, “tsavl” ცავლ). Similarly for the verb “to punish”, Present: “sdjis” (root, “sdj”, სჯ), Aorist: “dasadjja” (root, “sadj”, საჯ).

The preceding examples do not pretend to make an exhaustive panorama of the complexities. They only want to illustrate the difficulty of morpho-syntactic analysis. This has a significant impact on learning the language. In particular, finding an entry point for a verb in dictionaries is crucial to find translations and definitions. This is correlated with the issues of lemmatization of Georgian verbs in Georgian Dictionaries. Let us briefly recall some aspects (details can be found in [9, 7]). Georgian has no infinitive. Most dictionaries use the “masdar”, a verbal noun, as lemma to represent a verb³. It is thus an entry in the dictionary for a given verb. However, for the neophytes, going from a conjugated form to a masdar can be a real challenge. For example, for “chamodikhar” (ჩამოდიხარ, “you come down”), the masdar is “mosvla” (მოსვლა, “coming”). The only common part is “mo”. Many projects give samples of inflected forms as lemma(s). For example, Daraselia and Sharoff use third person singular future [4]. The “Comprehensive Georgian-English Dictionary” [12], presents, for all verbs, masdar and 3rd person singular in present and future tenses, both active (transitive) form and passive (intransitive) form, with markers for the indirect object in the third person. This is more exhaustive than in any previous bilingual Georgian dictionary. It is, however, still quite hard for a neophyte to track the above-mentioned “chamodikhar”. The Georgian-German dictionary of [15], uses a different approach. It uses the abstract verbal root under which all subparadigms are listed. It can result in an extremely complex structure of entries, which sometimes extend over many pages [7]. While this representation is very informative for linguists, it is too cumbersome for beginners, especially as many roots consist of only one or two characters. To quote T. Margalitadze “Scientific debates over these [lemmatization] principles

³In our system and in this report, a masdar is currently improperly called “Georgian infinitive” because it is easier to understand for the (French or English non-linguist) target users.

still continue right up to this day [9].

Some works provide exhaustive tables of inflected forms, for example the *Georgische Verbtabelle* of [2] or the conjugation tables of the “Biliki” series by Nana Shavtvaladze⁴. The latter contain conjugation tables of several types. The first type of tables (called “whole conjugation tables” in the following) concerns the verbs introduced in a given lesson. They are systematically conjugated at all the tenses that have been introduced so far. In these tables, masdar and English translation are also given. The second type of tables (called “sample tables” in the following) gives a list of conjugation samples, one line per verb. A line contains, firstly, an English translation and, for each tense introduced in the book, an inflected form at the third person singular. Those tables are a tremendous help for the neophytes, they contain invaluable information. However, learners have to browse through different books to find relevant information. Finding an inflected form in order to translate from Georgian to English is difficult. Indeed, the lines are sorted by English translation. When searching for an inflected form, learners have to check each one of the more than 10 000 entries. Furthermore, the inflected forms use the Georgian alphabet that is a big hurdle for beginners. Exceptions, which are quite common, cannot always be anticipated from the sample tables. Verbs introduced in the first books do not have a complete “whole conjugation” table because few tenses had been introduced at the time these verbs had been introduced. Searching is thus tedious, it takes time and it is not granted to find an existing entry.

We propose *Kartu-Verbs*, a Semantic Web base of Georgian inflected verb forms that can be seen as a front-end to any dictionary, thus bypassing the lemmatization issues.

Our project builds a comprehensive Semantic Web base of inflected verb forms aiming at foreigners learning Georgian⁵. When a verb is in the base, the aim is that all its inflected form are present and users can retrieve the lemmas relevant to access their preferred dictionary. Knowledge can easily be traversed in all directions: from Georgian to French and English and conversely; from an inflected form to a masdar and from a masdar to any inflected form; from component(s) to forms and from a form to its components. In order to build the base, conjugation rules, taking exceptions into account, are built in *Prolog*, a programming language designed for Language processing. The generated forms are integrated within a Semantic Web tool, *Sparklis*, which can retrieve information from their facets, and allows users to smoothly refine their queries by giving them suggestions [5]. The drawbacks listed above are thus overcome. That is illustrated in Section 2, which also shows how to gain meta-knowledge about the base. Section 3 explains the architecture of the project. Section 4 briefly describes how the inflected forms are systematically generated. Section 5 discusses related work, draws some perspectives and concludes the report.

⁴Biliki, Georgian Language For English Speakers. See <http://lsgeorgia.com>

⁵This base is available for consultation at <https://www-semllis.irisa.fr/software/georgian-verb-inflected-forms-base/>

2 Using Kartu-Verbs Our Georgian Verb Form Base

This section illustrates how to use our base of Georgian verb inflected forms and demonstrates the power of the tool. As the base is primarily meant to be a companion of the “Biliki” books, we use the knowledge structures defined by Nana Shavtvaladze (groups, subgroups, morphological decomposition, etc). We are aware that the morphological decomposition is simplified, for example with respect to the work of Kevin Tuite [16]. Section 2.1 illustrates how to find information about an inflected form, our initial goal for the project. Section 2.2 shows that it is equally easy to get information starting from an English infinitive. Section 2.3 shows how to build a sample of conjugation. Section 2.4 describes how to gain conjugation information from a given tense or a given ending. Section 2.6 shows how to check hypotheses about preradicals using logical operators. Section 2.7 gives more sophisticated queries to gain meta-knowledge about the base using aggregates.

2.1 Finding information about an inflected form

Sparklis SPARQL endpoint <http://servolis.irisa.fr/georgianverbs/sparql> Go Georgian Verb Forms Base Configure Learn

YASGUI view Permalink

Your query and its **current focus** (click on different parts of the query to change it)

give me every **verb**
that has a **Georgian form**
and that has a **person**
and that has a **number**
and that has a **tense**
and that has a **preverb**
and that has a **preradical**
and that has a **root**
and that has a **postradical**
and that has a **pFSF**
and that has an **ending**
and that has an **English infinitive**
and that has a **French infinitive** ✕

Sparklis suggestions to refine your query

Types and Relations of the verb
matches all of
that has a **tense** (118+)
that has a **type** (118+)
that has an **ending** (117+)
that has an **English infinitive** (117+)
that has a **French infinitive** (117+)
that has a **Georgian Infinitive** (117+)
that has a **label** (117+)
that has a **postradical** (117+)
that has a **relation to ...**
decreasing frequency 20+ items

Identities or Values of the verb
matches all of
anything
dardobdnen (2+)
idardebndnen (2+)
inadirebdnen (2+)
nadirobdnen (2+)
shetavazebdnen (2+)
tavazobdnen (2+)
tchkarobdnen (2+)
dardobh (2+)
decreasing frequency 194+ items

Aggregations and Operators
matches all of
for each **verb** give me ...
give me the **number of verb**
give me a **sample of verb**
if ___ else ___
if ___ else ___
is a **blank node**
is a **IRI**
is a **literal**
is a **number**
14 items

Results of your query

Created by Sébastien Ferré © 2013-2019 IRISA
Last modified May 15th, 2019. Visit my homepage for more information.

Figure 1: Defining a query with the help of suggestions

Figure 1 shows a screen shot of the base in Sparklis. The area is roughly divided in 3 parts: the query at the top, suggestions in the middle and results at the bottom (folded on this figure; it is illustrated on following examples). The displayed query enables to find 12 features of inflected forms: its form in Georgian alphabet, person, number, tense, preverb, preradical, root, postradical, present/future stem formant (PFSF), ending, English infinitive and French Infinitive.

The Suggestions part, in the middle, is divided in 3 areas: “Types and Relations of the verb” that is searched for, “Identities or Values of the verb” and “Aggregations and Operators”. In the Relations area we can see features that have already being used (for example “French Infinitive”); there are also features that could still be used,

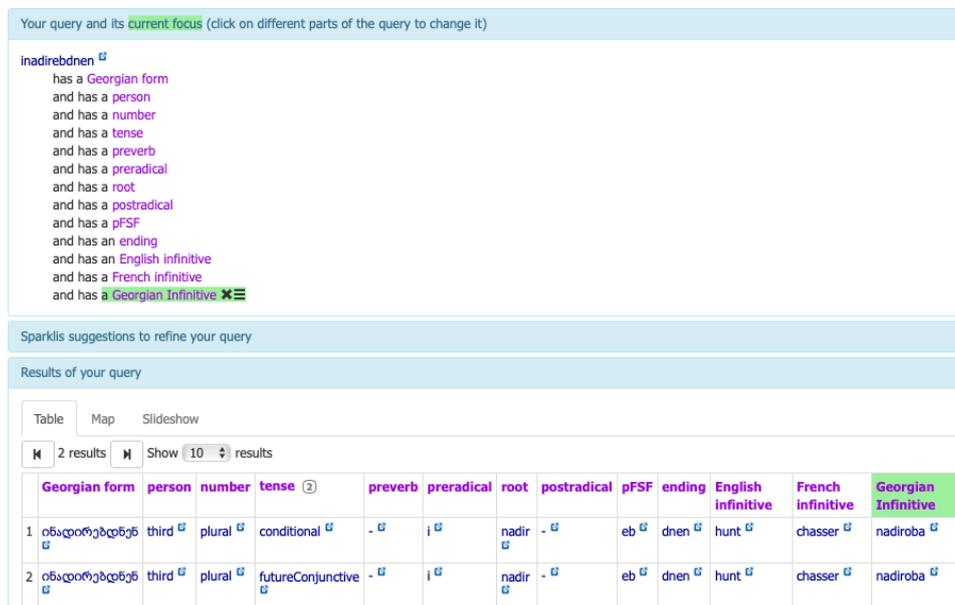


Figure 2: Getting information about an inflected form

for example “Georgian infinitive”. In the Values area we can see possible values of verb form. Let us assume that we are interested in “inadirebdnen” and that we would also like to have information about the “Georgian infinitive”. We can click on both suggestions. The aggregations and Operators of the third area are illustrated in Sections 2.6 and 2.7.

Figure 2 displays the query and results areas after “inadirebdnen” and “Georgian infinitive” have been selected. The suggestions area is folded. The query has been automatically updated. At the top there is no longer “give me every verb” but “inadirebdnen”, and Georgian infinitive has been added. In the result area, the, now, 13 requested features about “inadirebdnen” correspond to 13 columns. We can see its form in Georgian alphabet, “ინადირებდნენ”; it corresponds to the third person plural of both conditional and future conjunctive tenses; it has neither preverb nor postradical (indicated by “-”); its preradical is “i”; its root (or stem) is “nadir”; its present/future stem formant (PFSF) is “eb”; its ending is “dnen”; a possible English translation is “to hunt”; a possible French translation is “chasser”, its Georgian infinitive is “nadiroba”.

Note that the fact that the same inflected form corresponds to 2 different tenses (Conditional and Further Conjunctive) is handled without any problem by Sparklis.

As illustrated above, all queries are built using suggestions. Users do not have to invent anything. They can use filters to help Sparklis propose relevant suggestions but queries are built solely by clicking on suggestions that are necessarily relevant. The benefits are threefold, firstly it is easier to find something in a list than typing it,

secondly users cannot mistype, and lastly, as a direct consequence, the queries can never give an empty result. That is a very strong property.

2.2 Finding Information From an English Infinitive

Your query and its **current focus** (click on different parts of the query to change it)

give me every **verb**
 that has a **Georgian form**
 and that has a **person**
 and that has a **number**
 and whose **tense is present** ✕☰
 and that has an **ending**
 and whose **English infinitive is live**
 and that has a **Georgian Infinitive**

Sparklis suggestions to refine your query

Results of your query

Table Map Slideshow

6 results Show 100 results

	verb ⑥	Georgian form ⑥	person ③	number ②	ending ④	Georgian Infinitive
1	vtskhovrob	ვცხოვრობ	first	singular	-	tskhovreba
2	tskhovrob	ცხოვრობ	second	singular	-	tskhovreba
3	tskhovrobs	ცხოვრობს	third	singular	s	tskhovreba
4	vtskhovrobt	ვცხოვრობთ	first	plural	t	tskhovreba
5	tskhovrobt	ცხოვრობთ	second	plural	t	tskhovreba
6	tskhovroben	ცხოვრობენ	third	plural	en	tskhovreba

Figure 3: The conjugation of the six persons at a given tense for a verb given in English

Figure 3 shows how to search for information starting from an English infinitive (here “to live”). The query also specifies the tense (“present”) and requests information about Georgian form, person, number, ending and Georgian infinitive. The table in the result part gives the conjugation at the 6 persons at the present tense, the specific endings and tells that it corresponds to Georgian verb “tskhovreba”. To get all tenses, users can specify them in turn in the query in order to get them one after the other, or they can leave the tense feature “open” and all tenses will be given. This view is equivalent to a “whole conjugation table” of the Biliki books mentioned in the Introduction.

By “open” features we mean features that do not have a value specified in the query (here “Georgian form”, “person”, “number” and “ending”). They produce the columns of the result area. The features with a value in the query (here “English infinitive” and “Tense”) are not repeated in the result area. One needs the query in order to interpret the results.

Note that, at any moment, users can add or retract any value or feature from the query. The results and suggestions area are then automatically updated by Sparklis.

2.3 Building a Sample of Conjugation

Your query and its **current focus** (click on different parts of the query to change it)

give me every **verb**
 that has a **Georgian form**
 and whose **person** is **second** [↗](#)
 and whose **number** is **singular** [↗](#)
 and that has a **tense**
 and whose **Georgian Infinitive** is **tskhovreba** [↗](#)
 and **that has an English infinitive** [✕](#) [☰](#)

Sparklis suggestions to refine your query

Results of your query

Table Map Slideshow

⏪ 10 results ⏩ Show results

	verb (10)	Georgian form (10)	tense (10)	English infinitive
1	tskhovrob ↗	ცხოვრობ ↗	present ↗	live ↗
2	tskhovrobdi ↗	ცხოვრობდი ↗	imperfect ↗	live ↗
3	itskhovreb ↗	იცხოვრებ ↗	future ↗	live ↗
4	itskhovre ↗	იცხოვრე ↗	aorist ↗	live ↗
5	itskhovro ↗	იცხოვრო ↗	optative ↗	live ↗
6	itskhovrebd ↗	იცხოვრებდი ↗	conditional ↗	live ↗
7	itskhovrebde ↗	იცხოვრებდე ↗	futureConjunctive ↗	live ↗
8	getskhovra ↗	გეცხოვრა ↗	pastPerfect ↗	live ↗
9	gitskhovria ↗	გიცხოვრია ↗	presentPerfect ↗	live ↗
10	tskhovrobde ↗	ცხოვრობდე ↗	presentConjunctive ↗	live ↗

Figure 4: All 10 tenses at the second person singular of a verb given in Georgian

Figure 4 shows how to conjugate Georgian verb “tskhovreba” at all the tenses known by the base for the second person singular. This is equivalent to a “Sample table” of the “Biliki” books. The advantage is that users can chose the person(s) they want. Note that this time, we have specified the verb by its Georgian infinitive but we could have given one of its English or French infinitives.

2.4 Learning About Tenses

Your query and its **current focus** (click on different parts of the query to change it)

give me every **verb**
 that has a **Georgian form**
 and whose **person** is **first** [↗](#)
 and whose **number** is **singular** [↗](#)
 and whose **tense** is **aorist** [↗](#)
 and that has a **preverb**
 and that has a **pFSF**
 and that has an **ending**
 and that has an **English infinitive**
 and for each **ending**
 give me
 a **sample of verb**
 and **the number of verb** [✕](#) [☰](#)

Sparklis suggestions to refine your query

Results of your query

Table Map Slideshow

⏪ 6 results Show 100 results

	ending ↗	sample of verb ↗	number of verb ↗
1	e ↗	shevtavaze ↗	182
2	a ↗	meqola ↗	8
3	- ↗	vikirave ↗	6
4	da ↗	mkonda ↗	1
5	i ↗	gavushvi ↗	66
6	edi ↗	ts'avedi ↗	10

Figure 5: Learning the possible endings at a given tense and a given person

Let us assume that we want to learn about the possible endings at the aorist tense for the first person singular. Figure 5 shows a query that produces all the possible endings. Six of them are found in the base (“e”, “a”, none (“-”), “da” “i” and “edi”); for each ending the tool gives a sample of inflected form using this ending and the

total number of forms using this ending. One can note that “mkonda” is the only verb currently in the base with ending “da” at aorist first person singular. It is a form of verb to have (kona, ქონა), a very special verb in many languages. The singular point is correct. On the contrary, the endings of “vikirave” and “ts’avedi” are highly suspect. Mistakes in the morphological analysis seem to have been made in the process of making the base on them and probably on the other verbs of their cluster (6+10). While the problems remain to be fixed, this is an illustration that Sparklis is also a powerful tool to test the quality of the base.

Your query and its **current focus** (click on different parts of the query to change it)

give me every **verb**
 that has a **Georgian form**
 and that has a **person**
 and that has a **number**
 and that has a **tense**
 and that has a **preverb**
 and that has a **pFSF**
 and whose **ending is da** ❌ ❌ ❌
 and that has an **English infinitive**
 and that has a **Georgian Infinitive**

Sparklis suggestions to refine your query

Results of your query

Table Map Slideshow

11 results 1 - 10 of 200+ Show 10 results

	verb (147+)	Georgian form (147+)	person	number	tense (2+)	preverb (9+)	pFSF (4+)	English infinitive (94+)	Georgian Infinitive (77+)
1	tavazobda	თავაზობდა	third	singular	imperfect	-	ob	offer	shetavazeba
2	shetavazebda	შეთავაზებდა	third	singular	conditional	she	eb	offer	shetavazeba
3	dardobda	დარდობდა	third	singular	imperfect	-	ob	feel_sorrow	dardi
4	idardebda	იდარდებდა	third	singular	conditional	-	eb	feel_sorrow	dardi
5	nadirobda	ნადირობდა	third	singular	imperfect	-	ob	hunt	nadiroba
6	inadirebda	ინადირობდა	third	singular	conditional	-	eb	hunt	nadiroba
7	tchkarobda	ჩქარობდა	third	singular	imperfect	-	ob	hurry	sitchkare
8	itchkarebda	იტყვარებდა	third	singular	conditional	-	eb	hurry	sitchkare
9	khumrobda	ხუმრობდა	third	singular	imperfect	-	ob	joke	khumroba
10	ikhumrebda	იხუმრებდა	third	singular	conditional	-	eb	joke	khumroba

Figure 6: Finding possible tenses from a given ending

Let us, now, assume that we have a verb form that is not present in the base but the user thinks that the ending is “da”. It could already be interesting to know the possible tenses. Figure 6 shows a query that sets the ending and asks for many features (“Georgian Form”, person, number, tense, preverb, pFSF and English infinitive), in order to try to map the searched verb to what is currently in the base. The result part shows 10 forms out of the more than 200 that match the query. One can already see that the ending most probably corresponds to a third person singular and that the tense is most probably imperfect or conditional. One can also see that there seems to be a correlation between the pFSF and the tense.

Your query and its **current focus** (click on different parts of the query to change it)

give me every **verb**
 that has a **Georgian form**
 and that has a **person**
 and that has a **number**
 and that has a **tense**
 and that has a **preverb**
 and whose **pFSF is ob** **X**
 and whose **ending is da**
 and that has an **English infinitive**
 and that has a **Georgian Infinitive**

Sparklis suggestions to refine your query

Results of your query

Table Map Slideshow

results 21 - 30 of 39 Show 10 results

	verb ⁽³⁴⁾	Georgian form ⁽³⁴⁾	person	number	tense ⁽²⁾	preverb ⁽³⁾	English infinitive ⁽³⁴⁾	Georgian Infinitive ⁽³¹⁾
21	akrobdā	აკრობდა	third	singular	imperfect	-	swich_off_(light)	tchakroba
22	akrobdā	აკრობდა	third	singular	imperfect	-	put_out_(fire)	tchakroba
23	akrobdā	აკრობდა	third	singular	imperfect	-	blow_out_(candel)	tchakroba
24	tchaakrobdā	ჩააკრობდა	third	singular	conditional	tcha	swich_off_(light)	tchakroba
25	tchaakrobdā	ჩააკრობდა	third	singular	conditional	tcha	put_out_(fire)	tchakroba
26	tchaakrobdā	ჩააკრობდა	third	singular	conditional	tcha	blow_out_(candel)	tchakroba
27	mushaobda	მუშაობდა	third	singular	imperfect	-	work	mushaoba
28	tevzaobda	თევზაობდა	third	singular	imperfect	-	fish	tevzaoba
29	k'itkhulobda	კითხულობდა	third	singular	imperfect	-	read	k'itkhva
30	ats'qobda	აწყობდა	third	singular	imperfect	-	set_table	gats'qoba

Figure 7: Refining the query with the “ob” PSFS to try to identify rules

Figure 7 shows a refinement of the query, specifying “ob” for the PFSF. There are indeed mostly imperfect forms except some cases that are conditional. Those cases are the only one with a preverb. Users could investigate further to confirm or infirm whether the correlation really exists.

2.5 Comparing Similar Forms

When learning, it is often useful to confront similar forms. For example, let us assume that we realize that we are confused about “to have someone” and “to resemble”. Figure 8 illustrates how to display the third singular present form for both verbs, using feature “French infinitive” and the ‘ðr’ logical operator. The result part shows that there is only one character difference. Now we know that “to have someone” has a “q” (“ყ”) as second letter and “to resemble” a “g” (“გ”).

09:23 Mar. 17 déc. Non sécurisé — irisa.fr

YASGUI view Permalink

Your query and its **current focus** (click on different parts of the query to change it)

give me everything
 that has a **Georgian form**
 and that is a **verb**
 and whose **person** is **third**
 and whose **number** is **singular**
 and whose **tense** is **present**
 and that is something
 and whose **French infinitive** is
resembler
 or **avoir_quelqu_un**
 and that has an **English infinitive**
 and that has a **Georgian Infinitive**

[Sparklis suggestions to refine your query.](#)

Results of your query

Table Map Slideshow

2 results Show 100 results

	verb ②	English infinitive ②	Georgian Infinitive ②	Georgian form ②
1	hqavs	have_someone	qola	ჰყავს
2	hgavs	resemble	damgvaneba	ჰგავს

Figure 8: Comparing two similar forms

Similarly, let us assume that we are confused about verbs containing “gheb” (“ღებ”). Figure 9 illustrates how to use the suggestion area for the “Georgian form” features at third singular present that contain “ღებ”. Now we know that for verb “to dye” “eb” is part of the stem/root and the PFSF is “av”, while for verbs of the family of “to take” “eb” is not part of the stem but it is the PFSF.

09:44 Mar. 7 janv. Non sécurisé — irisa.fr 80 %

YASGUI view Permalink (parts of the query to change it)

give me everything
 that has a **Georgian form** **x**
 and that is a **verb**
 and whose **person** is **third**
 and whose **number** is **singular**
 and whose **tense** is **present**
 and that is something
 and that has an **English infinitive**
 and that has a **Georgian Infinitive**
 and that has a **root**

Sparklis suggestions to refine your query

Types and Relations of the Georgian form

matches :

- that is the **Georgian form of ...** (3)
- that has a **label** (3)
- that has a **relation to ...**
- that has a **relation from ...**
- according to which there is ...**

decreasing fr 5 items

Identities or Values of the Georgian form

matches :

- anything**
- იღებს** (4)
- უღებს** (3)
- იღებავს**

decreasing fr 4 items

Aggregations and Operators

matches all of :

- for each thing give me ...
- give me the **number** of thing
- give me a **sample** of thing
- if ... else ...**
- is a blank node**
- is a IRI**
- is a literal**
- is a number**

14 items

Results of your query

Table Map Slideshow

8 results Show 20 results

	verb (3)	English infinitive (6)	Georgian form (3)	Georgian Infinitive (4)	root (2)
1	ighebs	take	იღებს	agheba	gh
2	ighebs	take	იღებს	migheba	gh
3	ighebs	receive	იღებს	agheba	gh
4	ighebs	receive	იღებს	migheba	gh
5	ughebs	make_copy	უღებს	gadagheba	gh
6	ughebs	take_a_picture	უღებს	gadagheba	gh
7	ughebs	give_another_helping	უღებს	gadagheba	gh
8	ighebavs	dye	იღებავს	shegheba	gheb

Figure 9: Investigating similar stems

2.6 Learning About Preradicals

Your query and its **current focus** (click on different parts of the query to change it)

give me every **verb**
that has a **Georgian form**
and whose **person** is **first**
and whose **preradical** is
not v
and **not vi**
and **something**

Sparklis suggestions to refine your query

Types and Relations of **the preradical**

matches all of ...

- that is the **preradical of** ...
- that is the **postradical of** ...
- that is the **ending of** ...
- that is the **pFSF of** ...
- that has a **relation to** ...
- that has a **relation from** ...
- according to** which there is ...

decreasing frequency 9+ items

Identities or Values of **the preradical**

matches all of ...

- anything**
- va**
- gvi**
- mi**

decreasing frequency 7+ items

Figure 10: Learning preradicals at first person, using logical operators “and” and “not”

Let us assume that a user believes that a first person always has preradical “v” or “vi”. Figure 10 shows a query that searches for forms at first person (singular or plural, as number is not specified) and whose preradical is neither “v” nor “vi”. The suggestions area immediately shows that there are at least 6 other possibilities. The user has to refine his/her knowledge!

Note that the query uses logical operators “and” and “not”. Other operators such as “or” are also available by clicking on the small menu logo at the right of the focus (in green on the figures).

2.7 Meta-Knowledge about the base

Your query and its **current focus** (click on different parts of the query to change it)

give me every **verb**
 that has a **group**
 and that has an **ending**
 and that has a **pFSF**
 and that has a **preverb**
 and that has a **Georgian Infinitive**
 and for each **group**
 give me
 a **sample of verb**
 and the **number of verb**
 and the **number of pFSF**
 and **the number of preverb** ✕☰
 and the **number of Georgian Infinitive**

Sparklis suggestions to refine your query

Results of your query

Table Map Slideshow

4 results Show 100 results

	group ④	sample of verb ④	number of verb ④	number of pFSF ④	number of preverb ④	number of Georgian Infinitive ④
1	g4	khar	1626	13	18	32
2	g2	mqopnis	599	3	7	11
3	g3	movip'ove	3051	7	8	56
4	g1	vtavazob	9988	15	13	179

Figure 11: Gaining knowledge about the group distribution, their numbers of PFSF, preverbs and Georgian infinitives using aggregation operators

Sparklis also enables to gain knowledge about the current base. For example, Figure 11 shows a more sophisticated query, using aggregates, to gain information about the distribution of the 4 Biliki groups. On the results area, we can see that group g1 is the largest one. It gathers 179 Georgian infinitives, 9 988 inflected forms, 15 different PFSF and 13 preverbs. The results area also shows a sample of each group.

Note that this query, as all the previous ones, was built by clicking. Here the right hand part of the suggestion area had been used (not illustrated here).

Your query and its **current focus** (click on different parts of the query to change it)

give me every **verb**
that has a **group**
and that has a **pFSF**
and whose **tense** is **present** [↗](#)
and whose **number** is **singular** [↗](#)
and whose **person** is **second** [↗](#)
and for each **pFSF**
give me
a sample of verb [✕](#) [≡](#)
and the **number of verb**

[Sparklis suggestions to refine your query](#)

Results of your query

Table Map Slideshow

⏪ 8 results ⏩ Show 100 results

	pFSF ⑧	sample of verb ⑧	number of verb ⑦
1	av ↗	ip'arsav ↗	33
2	a ↗	ggonia ↗	1
3	ev ↗	ukhvev ↗	1
4	ob ↗	stavazob ↗	31
5	- ↗	khar ↗	47
6	eb ↗	aukmeb ↗	130
7	i ↗	ivartskhni ↗	24
8	am ↗	abam ↗	6

Figure 12: Gaining knowledge of PFSF at the present tense

Similarly, Figure 12 shows information about the distribution of PFSF at present tense. Again, PFSF “a” and “ev” with only one representative could be questioned: is the base incomplete? Is it a mistake? That is again an illustration of how Sparklis helps to test the quality of the base.

3 Project Architecture

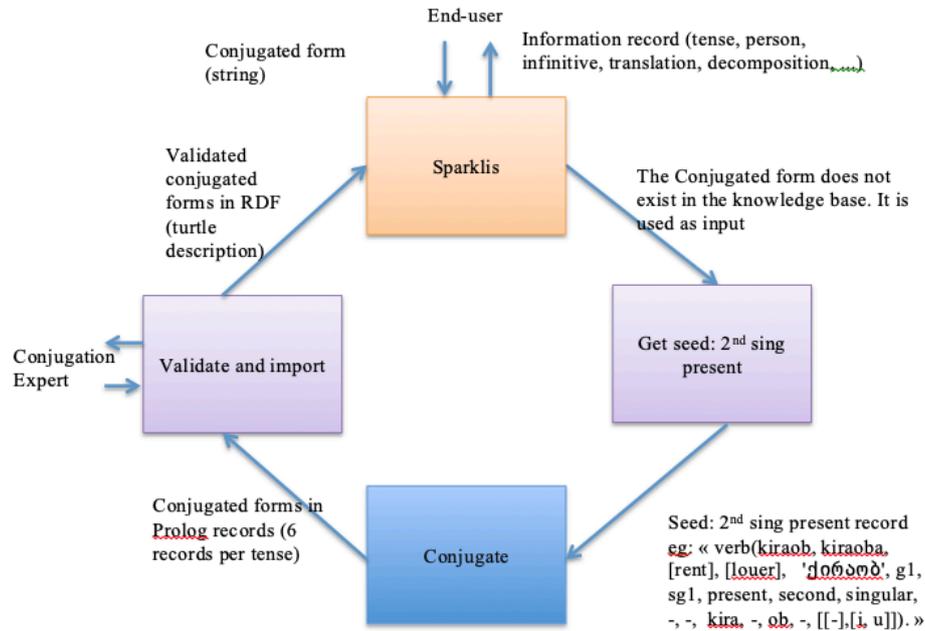


Figure 13: Project Architecture

Figure 13 represents the global architecture of the project. A user, for example a foreigner learning Georgian, enters a conjugated form. If this form exists in the base, a structure giving information about this form is accessible and allows the user to find the tense, the person, possible translations, the infinitive form, etc. If the form does not exist, someone can provide the seed to make conjugated forms: the second person singular at present tense, with its morphological decomposition and other information, for example “verb (kiraob, kiraoba, [rent], [louer], its form in Georgian characters, 'ქირაობ', g1, sg1, present, second, singular, -, -, kira, -, ob, -, [[-], [i, u]])” which indicates that form “kiraob” comes from verb “kiraoba”, a possible translation into English is “to rent”, a possible translation into French is “louer”, the verb is of the 1st group, 1st subgroup, the form is in the Present tense, 2nd person singular, it has neither preverb nor pre-radical, its root is “kira”, it has no postradical, its PFSF is “ob”, it has no ending, the possible preverbs for the future and other times that require it are “i” and “u”. From this seed, the conjugation module makes inflected forms, 6 structures per tense with the same syntax as the seed. There are too many irregularities to trust the rules of conjugation. The generated forms have to be validated by experts. The validated forms are then transcribed in RDF Turtle format to be integrated into the base under Sparklis.

At present, the “Sparklis” and “Conjugate” boxes are working. The Sparklis process has been illustrated in Section 2. Generating the inflected forms is briefly described in Section 4. They both operate independently. The actions of the other boxes currently have to be done by hand.

4 Building Inflected Forms

To systematically produce the inflected forms, we use Prolog, a programming language that has been designed by Alain Colmerauer for natural language processing [3, 14]. Our “Conjugate” program allows both to easily program grammar rules and code exceptions.

Present		Imperfect		Future	
vkiraob	ვექირაობ	vkiraobdi	ვექირაობდი	vikiraveb	ვიქირავებ
kiraob	ქირაობ	kiraobdi	ქირაობდი	ikiraveb	იქირავებ
kiraobs	ქირაობს	kiraobda	ქირაობდა	ikiravebs	იქირავებს
vkiraobt	ვექირაობთ	vkiraobdit	ვექირაობდით	vikiravebt	ვიქირავებთ
kiraobt	ქირაობთ	kiraobdit	ქირაობდით	ikiravebt	იქირავებთ
kiraoben	ქირაობენ	kiraobdnen	ქირაობდნენ	ikiraveben	იქირავებენ

Figure 14: Conjugation of a verb from the first group, at present, imperfect and future tenses, starting from the 2nd person singular present

As mentioned earlier, we use the groups and subgroups defined by Nana Shavtvaladze in her series of “Biliki” books. We also rely on her morphological decomposition and on the tables of verbs given in the books.

Figure 14 shows the conjugation of a verb from the first group at present, imperfect and future tenses. To build the forms at the Present tense for verbs of the first group, we start from the second person singular present (here “kiraob”) for all the singular persons. In general the first person singular is constructed by adding a “v” preradical, the 3rd person singular by adding a “s” ending. The plural persons are in general built from their singular person. The first person plural by adding a “t” ending to the first person singular; the second person plural by adding a “t” ending, and the 3rd person plural by adding most often a “en”, “an” or “ian” ending. For imperfect, in general, the form is constructed from the same person at present tense and the endings are respectively replaced by “di”, “di”, “da”, “dit”, “dit” and “dnen”. For Future, we also start from Present at the same person, and it suffices, in general, to add a preverb (here “i”). For the displayed example, in addition, as the verb is irregular, it is necessary to modify the PFSF (“ob” changed to “eb”) and add a postradical (“v”) after the root.

For other tenses and other groups, rules are slightly different but the same principle is used. See [10] for more details

Figure 15 shows an excerpt from the Prolog program to build, for the present tense, the 3rd person plural from the 3rd person singular. Predicate `build_third_plu`

```

build_third_plu(GeorgianInfinitive, SubGroup, ThirdSing, ThirdPlu) :-
    change_number(singular, plural, ThirdSing, Form),
    build_third_plu_do(GeorgianInfinitive, SubGroup, Form, ThirdPlu).

% Exceptions
% ...
build_third_plu_do(djdoma, sg15, _, Form, ThirdPlu) :-
    change_root(zi, skhed, Form, Form1),
    change_ending(s, an, Form1, ThirdPlu),
    !.

%regular rules
build_third_plu_do(_GeorgianInfinitive, sg7, Form, ThirdPlu) :-
    change_ending(_, t, Form, ThirdPlu),
    !.
build_third_plu_do(_GeorgianInfinitive, sg8, Form, ThirdPlu) :-
    change_ending(_, at, Form, ThirdPlu),
    !.
build_third_plu_do(_GeorgianInfinitive, SubGroup, Form, ThirdPlu) :-
    member(SubGroup, [sg5, sg15]),
    change_ending(_, an, Form, ThirdPlu),
    !.
build_third_plu_do(_GeorgianInfinitive, SubGroup, Form, ThirdPlu) :-
    member(SubGroup, [sg9, sg10, sg11]),
    change_ending(_, ian, Form, ThirdPlu),
    !.
build_third_plu_do(_GeorgianInfinitive, SubGroup, Form, ThirdPlu) :-
    member(SubGroup, [sg1, sg2, sg3, sg4, sg6, sg12, sg13, sg14]),
    change_ending(s, en, Form, ThirdPlu).

```

Figure 15: An excerpt from the Prolog predicate to build for the Present tense the 3rd person plural from 3rd person singular

takes as parameter the Georgian infinitive, the subgroup, the form for 3rd person singular and returns the form for the 3rd person plural. It first changes the number from singular to plural and then make the others changes depending on the subgroup by calling predicate `build_third_plu_do`.

Predicate `build_third_plu_do` deals with exceptions first. We have shown only one of the exceptions on the figure: verb “*djdoma*” (ჯდობა, “to sit down”), of subgroup `sg15`, for which the root has to be changed from “*zi*” to “*skhed*”. Then regular rules address each subgroup. Five rules respectively address subgroups 7, 8, 5&15, 9&10&11 and the remaining ones by changing the ending, respectively to “*t*”, “*at*”, “*an*” “*ian*” and “*en*”.

For each of the 6 persons in each of the tenses, a set of rules is built according to this model with more or less exceptions. One can see that Prolog is especially well suited to this problem. The rules are relatively explicit and for each detected exception, a clause is simply added in front of the general rules.

5 Discussion, Perspectives and Conclusion

To our best knowledge, our tool is the only one of its type. We have, for example, found nothing specific to the Georgian language on the MultiTAL platform⁶, expert in automatic language processing (TAL) focused on Eastern and/or poorly endowed languages [13]. The Georgian Wiktionary⁷ is aimed at Georgian speaking people. It is of no help to people who are beginning to learn the language. Google translate⁸ is till doing quite poorly to translate Georgian verbs. INESS::XLE-Web⁹, the system of Paul Maurer, is dedicated to linguists. It is able to parse sentences and produce syntax tree of a number of languages, including Georgian [11]. While its linguistic power is much larger than what we offer, the information that we need is buried in the syntax tree and not really accessible to beginners. Furthermore, there are no transliterations, no translations, and last but not least none of our querying possibilities.

Our project is still under development. Currently, the base contains 15235 inflected forms related to 253 verbs for 10 tenses (present, imperfect, conjunctive, future, conditional, future conjunctive, aorist, optative, present perfect, past perfect). The forms have been generated and tested by students who are native Georgian speakers. At least all the verbs of the “*Biliki*” books are covered. One can expect that the most useful verbs for every day’s life are already present. According to Tuite [16], 5 tenses are missing: present iterative, imperative, permansive, mixed conjunctive present, perfect conjunctive.

The short-term perspectives are as follows. The current verb forms still have to be systematically tested (using Sparklis as in Section 2.7 for example). We are still

⁶<http://multital.inalco.fr>

⁷[https://meta.wikimedia.org/wiki/List_of_Wiktionaries_by_language_family#Kartvelian_languages_\(8,074_-_1_-_1\)](https://meta.wikimedia.org/wiki/List_of_Wiktionaries_by_language_family#Kartvelian_languages_(8,074_-_1_-_1))

⁸<https://translate.google.fr>

⁹<http://clarino.uib.no/iness/xle-web>

in the process of analyzing exceptions and irregularities. A library of usual queries is under construction. Phonetic and English-oriented transliterations are planned in order to help non-French users. More verbs will be added.

In the medium term, we have to slightly revise the ontology that is structuring our form description in order to use a vocabulary more standard in linguistics and to be able to accommodate other types of words (past participles, nouns, adjectives, ...). We have to adapt the generation rules in order to be able to build forms with direct and indirect object markers (see for example [1]), a feature that is especially confusing for French and English speakers. For example, “I do (without other indication)”=“vak’eteb”(ვაკეტებ), “I do for me” = “vik’eteb”(ვიკეტებ), “I do for you” =“gik’eteb” (გიკეტებ) [note the disappearance of the first person marker, “v”], “you do for me”=“mik’eteb”(მიკეტებ), etc.

In a longer term, we want to complete the system with the 2 modules presented above that are not yet designed: 1) to help users enter a seed, especially to help them make the decomposition into morphemes and decide upon their group relying on Sparklis’ suggestion mechanism; 2) to help users validate and update the Conjugate program when exceptions are detected by experts. Users should not have to see the Prolog code. Meta-facilities of Prolog allow programs to build Prolog code. We hope to be able to build on top of them.

In the long term, links to an actual electronic dictionary would be a big asset. At some point, it will be important to ensure that the tool is collaborative, and that any user can suggest modifications and new entries in the database in a safe way.

In conclusion, our base, in its current state, is already a successful proof of concept. It can help foreign learners to easily get information about verbs that they encounter in Georgian texts. This is an ongoing project with numerous perspectives and we are open to collaborations.

Acknowledgment

Thanks to the Erasmus+ International Credit Mobility program, the following Georgian students contributed to the project: Keti Meipariani, Mariam Asatiani, Ana Idadze, Mikheil Maisuradze, Veriko Nikuradze, Tornike Tchanturia, Ana Elchishvili, and Aleksandre Jajanidze. I am especially grateful to Keti Meipariani who managed the working teams of two semesters. She proofread this report and made contributed many useful remarks.

References

- [1] Irène Assatiani and Michel Malherbe. *Parlons Géorgien*. l’Harmattan, 2011.
- [2] Steffi Chotiwari-Jünger, Damana Melik’išvili, and Lia Wittek. *Georgische Verbtabellen*. Buske, 2010.

- [3] Alain Colmerauer. From natural language processing to Prolog. Presentation at the Academy of Mathematics and Systems Science, Beijing, April 8, 2011, April 2011. <http://alain.colmerauer.free.fr>.
- [4] Sophiko Daraselia and Serge Sharoff. Enriching Georgian dictionary entries with frequency information. In Tinatin Margalitadze and George Meladze, editors, Proceedings of the 17th EURALEX International Congress, pages 321–327, Tbilisi, Georgia, sep 2016. Ivane Javakhishvili Tbilisi University Press.
- [5] Sébastien Ferré. Sparklis: An expressive query builder for SPARQL endpoints with guidance in natural language. *Semantic Web: Interoperability, Usability, Applicability*, 8(3):405–418, 2017.
- [6] Hélène Gérardin. Les valeurs modales de la diathèse passive en Géorgien. *PhiN, Philologie im Netz*, 2016. *PhiN-Beiheft 9/2016*: 26, <http://www.fu-berlin.de/phin>.
- [7] Jost Gippert. Complex morphology and its impact on lexicology: the Kartvelian case. In Tinatin Margalitadze and George Meladze, editors, Proceedings of the 17th EURALEX International Congress, pages 16–36, Tbilisi, Georgia, sep 2016. Ivane Javakhishvili Tbilisi University Press.
- [8] Hélène Gérardin. Les verbes intransitifs primaires et dérivés en Géorgien : Description morphosyntaxique, sémantique et dérivationnelle. PhD thesis, Institut National des Langues et Civilisations Orientales, December 2016. UMR 7192 – « Proche-Orient-Caucase : langues, archéologie, cultures ».
- [9] Tina Margalitadze. Lexicography of Georgian: a brief overview. *Language@Leeds Working Papers in Linguistics*, 1, 2020. Forthcoming.
- [10] Ketevani Meipariani and Mariam Asatiani. A database of Georgian verb forms: Conjugation part. Master’s thesis, INSA Rennes, France and Tbilisi State University, Georgia, June 2019. Supervisor: Ducassé Mireille.
- [11] Paul Meurer. A computational grammar for Georgian. In *International Tbilisi Symposium on Logic, Language, and Computation*, pages 1–15. Springer, 2007.
- [12] Donald Rayfield, Shukia Apridonze, Ariane Chanturia, Rusudan Amirejibi, Laurence Broers, Levan Chkhaidze, and Tina Margalitadze, editors. *A Comprehensive Georgian-English Dictionary*. Garnett Press, 2006.
- [13] Driss Sadoun, Satenik Mkhitarian, Damien Nouvel, and Mathieu Valette. The MultiTal NLP tool infrastructure. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pages 156–163, 2016.

- [14] L. Sterling and E. Shapiro. *The Art of Prolog*, second edition. MIT Press, Cambridge, Massachusetts, 1994. ISBN 0-262-19338-8.
- [15] Kita Tschenkéli, Yolanda Marchev, and Lea Flury. *Georgisch-deutsches Wörterbuch*, volume 2. Amirani-Verlag Zürich, 1965.
- [16] Kevin Tuite. *Kartvelian morphosyntax: Number agreement and morphosyntactic orientation in the South Caucasian languages*. Lincom Europa Munich, 1998.