

**Khatuna Beridze**

Batumi State University

**M. Tandaschvili,**

Frankfurt Goethe University

**M. Giorgadze, R. Khalvashi**

Batumi State University

## **Translation of the Batumi Linguocultural Digital Archive: The methodology of Glossing and Tagging Social, Political and Cultural Information in Translation of DRSP**

### **ABSTRACT**

The Georgian dialects' translation methodology (GDTM) for the digitized recorded speech patterns DRSP in the BalDAR was adopted for translation of the stratified language, which in this specific case is the Georgian dialect containing residues of Turkish, Persian, and Arabic, and Pontic Greek borrowings. However, GDTM can be extended as a general approach to the translation of other Georgian dialects into Standard English (SE) for the language documentation and archiving purposes. The justification for the Georgian Dialect Translation Methodology (GDTM) is that Batumi Linguocultural Digital Archive is a source of the cross-cultural validation of the Georgian and its dialects. The digitized, documented speech patterns contain many facts which are, at present, and will be in the future too, worth researching in various fields of study: Glossing and Tagging in translation can enhance the potential of the DigiArchive for the global research by the international scholars.

**Keywords:** translation methodology, glossing, tagging, Batumi Linguocultural Digital Archive

### **Introduction:**

The paper dwells on the methodology of adding glosses as annotations in the process of translating Adjarian dialect<sup>1</sup> of the Georgian language into English. The texts are digitized

recorded speech patterns (DRSP) of the elderly population living in the borderline territories of Adjara. The preceding research paper Batumi Linguocultural Digital Archive (Contemporary Technological Achievements for the Database

---

<sup>1</sup>. See more about the areas in Adjara where the respondents were recorded at

[http://www.multilinguaeducation.org/storage/uploads/articles\\_contents/6-tandashvili.pdf](http://www.multilinguaeducation.org/storage/uploads/articles_contents/6-tandashvili.pdf)

Arrangement of the Folklore Resources)<sup>2</sup> mentions that the respondents used Russian and Turkish barbarisms.

The Georgian dialects' translation methodology (GDTM) for the DRSP was adopted for translation of the stratified language, which in this specific case is the Georgian dialect containing residues of Turkish, Persian, and Arabic. However, GDTM can be extended as a general approach to the translation of other Georgian dialects into Standard English (SE) for the documenting and archiving purposes. We followed the international experience of using SE for translation purposes, recommended for translating dialects especially for the language documenting. Meanwhile, the pieces of socio-cultural and political information encoded in the informal register patterns have been explained, clarified, defined (glossed) and tagged.

The goal of the research is to justify the introduction of simple glosses in the DRSP translation, to outline the design part of speech (POS) tags and specify the custom-tailored Tagset.

### **Brief Overview of Dialects in Adjara**

Adjarian is one of the oldest and most important dialects of the Georgian language. It reflects the historical past of Georgia. Since the

VIII century, BC Adjara is part of Colchis, then of the Egrisi Kingdom. Since the VIII century AD, Adjara is a part of the strongest principality of Tao-Klarjeti till the second half of the XIII century, when it became part of the Samtskhe-Saatabago Principality, from the second half of the XIV century of Guria Principality, returning half a century later (1512 ) together with Chaneti to the Atabags and was declared part of the Meskhetian Principality. Since the mid-XVI century, Adjara was under the control of Turks for over three centuries, which resulted in the hardest consequences not only for the people but also for the spiritual and material culture, language and religion.

The previous studies claim that the peculiarities of Adjarian dialects stem from of Kolkhuri, then Meskhuri-Klarjuri and Guruli Dialects, and finally by the Turkish influence.

The paper points out that the Russian language influence is visible along with above mentioned in the DRSP of the Adjarian dialects.

However, if taken into consideration the paper by M. Giorgadze and T. Abuladze, "Diaspora and Migration Studies: Pontic Greeks in the Post Soviet Period" (Giorgadze, Abuladze, 241:2018), it is also worth examining in the further studies, whether the Adjarian dialects borrow from the speech of the Greeks located at the Black Sea area. The

<sup>2</sup> The research is carried out in the framework of the project "Batumi Linguocultural Digital Archive (DigiArchive)",

funded by the Shota Rustaveli National Science Foundation of Georgia.

Greeks were the immediate socio-cultural setting in this region. As the researchers point out, they analyzed the borrowings in the Pontic Greek from Turkish, Russian, Georgian, and Laz. Besides, they argue that Georgian speech by the Pontian Greeks contains many Turkisms. The process of cultural blending is a long process which is reflected in the language as well (Giorgadze, 2017: 10-12; 312-315). Hence, further study will investigate whether residues of the Pontic Greek are reflected in the DRSP containing Adjarian dialects.

### **A brief overview of the Project BaLDAR**

The joint scientific project Batumi Linguocultural Digital Archive (BaLDAR) implemented jointly by Batumi Shota Rustaveli State University and Goethe Frankfurt University is sponsored by the Shota Rustaveli National Science Foundation. The project is a result of international cooperation and it aims to introduce new forms of scientific communication, which will support multidisciplinary research development. Our team of scholars used the advanced software: International Standard for Language Engineering (ISLE)<sup>1</sup>, Text Encoding Initiative (TEI)<sup>2</sup>, The ISLE Meta Data Initiative (IMDI)<sup>3</sup>, ELAN<sup>4</sup>. The project team mainly use TLA (<https://corpus1.mpi.nl/ds/asv/>) and DOBES (<http://dobes.mpi.nl/dobesprogramme/>).

The BaLDAR consists of 4 types of resources: A type resources: archived audio and video resources (MP3 and AVI format) B type resources: archived audio and video resources with transcribed text (in ELAN); C type resources: digitally documented and archived video resources with multimedia annotation – transcribed, glossed and structured interlinearly (in FLEX); D type resources: digitally documented and archived video resources with multimedia annotation and English translation. The translated digitized texts comprise 5 hours of the recordings.

### **Primary Georgian Dialect Translation**

#### **Methodology (GDTM)**

The pivotal argument for the Georgian Dialect Translation Methodology (GDTM) was the argument, that Batumi Linguocultural Digital Archive is a source of the cross-cultural validation of the Georgian and its dialects.

The Batumi Linguocultural Digital Archive contains Adjarian dialects and colloquial units of language translated into standard English. Adjarian dialects are sets of the spoken varieties of the Georgian language. They are influenced by the Turkish, due to the geopolitical influences over the five centuries. We noticed that the Adjarian speech patterns pivot on the Georgian syntactic structures. We observed that they tend to apply various dialectal borrowings from the Ottoman Turkish, which, in its own right, contains

Iranian (Persian), and Arabic. However, the urgency of building the digital archive was due to the great risk of aging and passing away of the vernaculars.

Some examples collected from the Batumi Linguocultural Digital Archive are presented in the paper.

In the previous research we established the following guidelines and methods for translation:

- a) The BaLDAR resources are part of the global language space, and the collections shall be accessible for non-speakers of 1. Georgian; 2. Georgian dialects;
- b) The translated patterns shall retain the cultural and historical flavour of the dialect wherever possible, i.e. if the modern Georgian can linguistically allow;
- c) A dialect borrowing shall not occur in translation from the colloquial speech into the standard Georgian and then into the standard English;
- d) The translation for the BaLDAR, and in general for the archived electronic texts, shall support the building of the unified translation system;
- e) The texts may use glosses/isoglosses, which will be added at for the later stage of work. Currently, the translations are proper

reflections of the characteristic speech patterns stored in the BaLDAR.

We expressed hopes, that the translation could help international researchers in their synchronic and diachronic analysis of the Adjarian speech patterns preserved now in the e-archives.

Therefore, our task was to develop an approach to the digitized recorded speech patterns (DRSP). We opted for the method that a) the syntax of the Modern Standard English (MSE) would be used on one hand, and 2) the other hand, the dialects, as well as all the cases of linguistic variation would be made available for international research purposes.

We had to figure out the criteria for translation of the social and situational variation of language. The search for the translation methodology for the DRSP did not yield relevant results.

### **The Linguistic Research Potential of the Archived Digitized Records through the Translation Methodology GDTM**

The digitized recorded speech patterns are historical documents in the DigiArchives. In this particular case of BaLDAR, if these patterns retain authentic markedness in translation, they can be of great potential for the international scholars of linguistics. Translated DRSP (digitized recorded speech patterns ) have significance their uniqueness in the native language makes them especially valuable for

the analysis. The digitized, documented speech patterns contain many facts which are, at present, and will be worth researching in various fields of study:

- Anthropology (social and cultural)
- History
- Culture
- Language use
- Politics

Hence, for the digitized recorded speech patterns in the databases of the DigiArchives, which a. originate from the languages of lesser diffusion and b. are intended for the wider international research opportunities, are to be

1) aligned; 2) translated; 3) glossed; 4) tagged

Justification of glossing, tagging and thus retaining such forms in the DRSP translation is that the speakers narrate about the Great Purge of 37ies, sharing their private dramas and the injustice suffered by their families. However, the speech patterns are not composed of negative judgments and do not contain negatively charged lexical units. The narratives are rather descriptive, the horrors of the Soviet rule, exile, regime, and oppression are revealed in the emotional prosody. The speakers repeat the same words, and the repetition both emphasizes the semantic meaning and point out how the speakers focus on their thoughts and how unconsciously they immerse in the memories while sharing their stories. Hence,

the **GDTM** is given priority to the traditional normalization/standardization methods.

## **The Georgian Dialect Translation**

### **Methodology GDTM**

#### **Translation strategies, Tags and Glosses**

Manually applied tags serve to mark the dialect used in the digitized recorded speech patterns in the BaLDAR database. Through them, we can discern the authentic linguistic patterns in the document. The non-lexicalized linguistic units have been glossed.

Throughout the process of translation, we distinguished two functions of [ak'hla] and

1. [ak'hla] as a discourse marker translated as [well] and tagged.
2. [ak'hla] as an interjection, translated as [well] and tagged.

The case below illustrates use of [Ak'hla] as a discourse marker, which indicates change of topic and move to a new exchange:

#### **Georgian text:**

<dm> ახლა</dm> მე ყორაოს გეტყვი. მე ახლა ვაკეთებ. ბეთმეზი რომ არ მაქ შაქრით ვაკეთებ, რომ <soc\_idlct> ახლა </soc\_idlct> რომ არ მაქ. თურქეთიდან <soc\_dia>გამოაქ</soc\_dia> ჩემს რძალს ბეთმეზი, იქ ყიდულობს და <soc\_dia>გადმოაქ</soc\_dia>. რომ არ <soc\_dia>მაქ</soc\_dia> იმ ყორაოს, <dm> რატომ იცი</dm>, ფერისათვის ვამღევი,

თეთრი არ იყოს, რომ მისი ფერი  
<soc\_dia>ქონდეს</soc\_dia>, ბეთმეზის  
ფერი.

### Literal translation:

**Now**, I will tell you K'orao. **Now** I cook it.  
When I don't have betmez, I cook it with  
sugar, because **now** I don't have it. My  
daughter-in-law takes betmez from Turkey.  
She buys it there and brings it over when I  
don't have it. That K'orao, you know why, I  
add for color, not to be white, to turn the  
white color into the color of betmez.

### Normalization and primary glossing:

The pragmatic meaning of [now] in the  
original equals to the SE [well]; In the process  
of normalization, we used a gambit [well] in  
English translation, as a widely used gap-filler,  
to follow the translational strategy of lexical  
change.

### Final translation, tagging and glossing:

<trglos><tr\_lexchng><dm></trglos>Wel  
l</dm></tr\_lexchng></trglos>, I will tell you  
<trglos><tr\_addnt>[about]  
</tr\_addnt></trglos> <trmglos>K'orao[sour  
plum juice boiled until thick]</trmglos>.  
<trglos><tr\_lexchng><intj>Well<tr\_lexchng>  
</intj></trglos>, I cook it...when I don't have  
<trmglos> betmez[boiled thick fruit  
jouce]</trmglos>, I cook it, <trglos>

<tr\_addnt>sometimes</tr\_addnt></trglos><tr  
glos><tr\_lexchng><intj>well</intj></tr\_lexch  
ng> </trglos> with sugar, because if I don't  
have it... My daughter-in-law brings betmez  
from Turkey when I'm short of it. You know, I  
add K'orao just for color, to turn the white color  
into the color of betmez.

In this particular pattern of speech, the  
output extralinguistic information covers: a)  
region-specific linguistic variation; 2) political  
and economic relations of Georgia and Turkey;

The pragmatic function of [ak'hla] *now* at  
the initial position of the sentence, and its  
semantic meaning differ drastically from the  
[ak'la] in the middle and in the end positions of  
the sentence; In the first case, it is a discourse  
marker, indicating change of topic, whereby the  
semantic meaning of *now* in the succeeding  
positions has a pragmatic function of  
interjection. In the second case, it also marks  
the linguistic variation and idiolect of the  
speaker, which we gloss and tag in translation.

Such cultural realia as Betmezi and  
K'orao – have been transcribed in translation,  
we have used square brackets for explications,  
they are glossed and tagged.

We distinguished all translational  
changes with the tag: <trglos></trglos>; The  
tag differs from the gloss tag which identifies  
explanations, inserted lexical units,  
clarifications, definitions, i.e. of the terms.

## Examples of lexical additions to the DRSP in the BalDAR

### Georgian sentence:

მაგ.: მასთან ცოტა ვჭამე, <soc\_dia> მარა</soc\_dia>, რომ <soc\_dia> წევდა</soc\_dia>, გადავაქციე. ვერ ვჭამე.

However, in English we added the lexical unit in the square brackets [the rest] as English syntax demands the object, e.g. I poured + [it] or I poured +[the rest]; and tagged it: with the custom-tailored tag, identifying lexical addition as a strategy of translation:

<tr\_lex.add>.... </tr\_lex.add>;

We also avoided ambiguity in translation, as far as a speaker may produce an elliptical sentence, which is clearly understandable for a listener, but it would be challenging to understand for a reader. e.g.

The implied word in the Georgian sentence is [the rest], omission of which does not make the meaning in the Georgian sentence ambiguous:

### Final translation:

<trglos> <dm> <tr\_lex.add>Well </tr\_lex.add></dm></trglos>, I ate a little together with her, but when she went, I poured <trglos> <tr\_lex.add>[the rest]</tr\_lex.add></trglos> away. I could not eat <trglos> <tr\_lex.add>[it] </tr\_lex.add></trglos>.

Along with the addition, we also tagged another object [it] added in the English sentence, and [well] as a discourse marker. As

the process of tagging is manual, we continued to build a custom system of Tagset, and introduced a pair of tags: <dm> </dm> for the Discourse Marker.

### The custom-tailored tags are added to identify code-switching local Adjarian version to the non-Georgian language:

დედამთილს ვუთხარი რომ რატომ მეთქი და, <soc\_CSturk>„ლაზლარ იაფიორ თერმონი, მუსლიმან იემეზინი“ </soc\_CSturk>, რატომ ქვია-მეთქი და ველამო არ იცოდანო, ეს ანდაზა რაც რომ <soc\_dia> გუუსახლებიან</soc\_dia> თუ წასულან თუ <soc\_dia> საცხა</soc\_dia> ძველი სიტყვა ვითომ ხალხი, <soc\_dia> ეიდა</soc\_dia> იარეს, იარეს, იარეს <soc\_dia> დაელიენ </soc\_dia> პროდუქტი...

### The tags and glosses are added in translation into Standard English when:

- a) a sentence in Turkish occurs in the exchange;
- b) non-standard parts of speech are used.
- c) translational shifts occur.

I <trglos><tr\_lexchn>asked</tr\_lexchn><trglos>my mother-in-law <trglos>[the reason]</trglos> why it was called

<soc\_CSturk>“Lazlar iapior termon, musliman iemezini” </soc\_CSturk>, and [she said] not everyone knew that proverb. When the people were exiled [or they fled themselves, as old word [legend] goes, they walked a long distance and were out of food.

**The translation of the DRSP shall reflect syntactic variation of language. Therefore, specific tags have been added to the Tagsat and used throughout the manual tagging process. In particular, we tagged dialectal lexical units used repeatedly:**

DRSP SL: იმხელა კაცი იყო მაღალი და იმ აკვანის ქვეშ <soc\_dia> მეიკუნთა </soc\_dia> ბურთივით შეძვრა და <soc\_dia> დეიმაღაო </soc\_dia>, აკვანის ქვეშო, რომ <soc\_dia>შამოვდენ</soc\_dia> არ ნახონო. მამაჩემი გადარჩა და ბიძაჩემი <soc\_dia>წეიყვანეს</soc\_dia>. <soc\_dia>წეიყვანა</soc\_dia>.

DRSP TL: He was such a tall man, and he folded himself and rolled under the cradle like a ball, to hide away, to stay invisible under the cradle if they <trglos> <tr\_addnt> Checkists </tr\_addnt> </trglos> <trmglos> [Russian: Чекист, an officer of the the All-Russian Emergency Commission for Combating Counter-Revolution and Sabotage] </trmglos> would enter. My dad escaped and my uncle was

taken. He took him <trglos> <tr\_addnt> arrested</tr\_addnt> </trglos>.

We suppose that lexical repetition shall be retained in translation. The translated DRSP shall reflect the emphatic pattern of the original, which conveys the speakers' message and the emotion simultaneously. The retained lexical repetition in translation also signals that the speaker is aware that her speech pattern differs from literary Georgian. Hence she tries to make her Speech Act as clear as possible for the recording.

The translational addition is the lexical unit <trglos> <tr\_addnt> **arrested** </tr\_addnt></trglos>, it is not in the square brackets since it does not define anything, however, it is an extra word inserted for the clarity of the meaning of the Speech Act.

The tag <trglos> identifies the strategy of translation. e.g. addition. It differs from the <trmglos> which identifies terms as untranslatable concepts.

The Soviet terms, as realia, shall be glossed, and translational shifts made, in particular, lexical additions.

Using this approach, we retain the colloquial speech pattern by means of the tags and glosses for the Corpus linguistic analysis, despite the translational shifts which transposes the sense into the SE.

### Example of an antonymous translation:

In the Speech Act below, the context makes clear that the dialectal lexical unit ‘კურთხვა’ [k’urtk’hva] has a meaning of cursing in this particular context. In the MSG the primary dictionary meaning of this word is ‘consecration’, but is used in its antonymous sense. The contextual meaning of this lexical unit overlaps the semantic meaning of the preceding lexical unit: ‘ლოცვა’ [lot’sva]; The latter lexical unit normally translates as ‘blessing’, but not in its ironic, and/or euphemistic use, which it acquires through the influence of the succeeding lexical unit and the overall context:

We applied the antonymic translation strategy in the speech pattern which revealed the irony of the speaker, or the pragmatic meaning of the Speech Act was antonymous to the semantic meaning.

DRSP SL: ამათი ლოცვა და ამათი კურთხვა ვთქვით.

DRSP TL: <trglos><tr\_antnm>I sent them my curses</tr\_antnm></trglos>.

As mentioned in the part **Brief Overview of Dialects in Adjara** of the paper, the respondents use Russian barbarisms, hence the influence of the Russian language is confirmed through the DRSP. The further manual linguistic processing of the DRSP in the BalDAR decodes significant socio-political and cultural information which otherwise would not be available for research purposes.

The following examples below showcase how the selected linguistic units are glossed and tagged manually in parallel structures which are part of the training corpus.

### Examples of the Russian barbarisms, glossed and tagged:

1.

DRSP SL: ...და <soc\_brbrus> ვსიო</soc\_brbrus> დამთავრდა.

DRSP TL: ...and <trmglos><soc\_brbrus>[vcë, vsio all, that’s it, end, stoppage] </soc\_brbrus></trmglos> that’s it.

2.

DRSP SL: პასპორტში თურქ <soc\_brbrus>პოდონი</soc\_brbrus><soc\_dia> ვინცხას </soc\_dia> ეწერა. იცი, იყო ასეთი პასპორტები გაცემული.

DRSP TL: Whoever were Turkish subjects <trmglos>[подданный, *podoni contaminated form*, citizen, subject] </trmglos> by their passports, were deported.

This case has been one of the interesting from the sociolinguistic point of view. The respondent uses подданный – a word borrowed from Russian by her, for which she cannot find an equivalent in her native tongue, and used the borrowing its phonetically contaminated form. For the diachronic research purposes of the languages of lesser diffusion, of their history and culture from the global perspective, it is significant to document the

evidential information that can be elicited through communication. The speaker's lexical repertoire, in this case, evidences that both her native and the second state language are equally of the lower register. Hence we use square brackets and tags to document in translation the same information as is available in the source text.

Other findings of Russian barbarisms include [adiali] *a blanket*, [saročka] *a shirt*, [maika] *a T-shirt*, [chorni peritsa] *black pepper*, [dukhovka] *an oven*, [banka] *a jar*, a Turkish barbarism [qener] *an edge* etc.

#### Example reflecting on the communist regime in the DRSP, glossed and tagged:

DRSP SL:

შვილები პარტიულები ყავდა,

DRSP TL:

His children were

<trglos><tr\_addnt>[communist]</tr\_addnt><trglos> party members.

We added in translation into the SE the lexical unit [communist] to avoid ambiguity, and specified the party, tagging and glossing it.

#### Example of lacunae, glossed and tagged:

DRSP SL:

<soc\_rel>ხოჯა</soc\_rel> აქ

არაფერი აღარ იყო.

DRSP TL:

<trmglos><soc\_rel>Khoja [used synonymously with Mullah]</soc\_rel></trmglos> was no one here anymore.

#### Example of an ethnonym, tagged and glossed:

DRSP SL:

დანიშნული ვიყავ, მარა  
გასახლება მოხდა, ხალხის  
გასახლება <soc\_ethn>ქურთი  
ხემში</soc\_ethn> რომ გაასახლეს,  
<soc\_ethn>თურქებიც</soc\_ethn>  
გასახლეს.

DRSP TL:

I had been engaged. But as people were deported, as it happened, <trmglos><soc\_ethn>Kurds [Hemshils]</soc\_ethn></trmglos> were deported, <soc\_ethn>Turks</soc\_ethn> were also deported.

#### Example of a toponym, tagged and glossed:

DRSP SL:

დედამთილი რომ  
<soc\_tpn>ხოჯაიდან</soc\_tpn> იყო,

DRSP TL:

Mother-in-law was from  
<trmglos><soc\_tpn>Hopa[a city in

Turkey adjacent to the border with  
Georgia] </soc\_tpn> </trmglos>

Figure 1. Tagset (non-inclusive)

Tags	Linguistically stratified speech pattern	Sample	Translation
<soc_brbrus> </soc_brbrus>	Barbarism from Russian	Prosta	just merely
<soc_brbturk> </soc_brbturk>	Barbarism from Turkish	Lazlar iafar termini and Musliman iemez onu.	Transliterated Glossed isoglossed
<soc_dltheadex> </soc_dltheadex>	Marks change of topic, or functions as a head exchange in a dialect	[ax'la], [k'hoda], [da], [h'oda]	Glossed and tagged
<soc_idlct> </soc_idlct>	Marks idiolect in speech variation; is a pragmatic connective in the Speech Act sequences;	[ak'hla]	Glossed and tagged
<soc_phnmbrb> </soc_phnmbrb>	Phonemic barbarism	[Ag'ranom]	Tagged and glossed

<soc_spchvar>  </soc_spchvar>	Speech variation	[bavs'hvi] [bag'hana]	Tags with the Literary English mark for the researcher that a non- literary form was used; Isogloss glosses; footnotes or isoglosses maybe used alternatively
<stl_diavar>  </stl_diavar>	Dialect variation	[bovs'hi] [bag'hana]	Tags with the Literary English mark for the researcher that a non- literary form was used;
<trmglos>  </trmglos>	Term	[kulak]	Transliterated Glossed Glossing of a term
<soc_spchrhyme>  </soc_spchrhyme>	Rhymed pairs of lexical units	[p'ek'hi] [M'ek'hi]	Tagged, glossed
<tr_addnt>  </tr_addtn>	Translational shifts, addition		Tagged, glossed

## Conclusion

The Georgian Dialect Translation Methodology GDTM and the translation strategies have been applied to the BalDAR DigiArchive, based on the task-specific custom-tailored tagset elaborated individually;

We applied the tags and glosses in translation into Standard English for transposition of Russian, Turkish, etc. barbarisms occurring in the speech patterns, non-standard parts of speech are used, and for marking the implemented translational shifts.

The applied translation, tagging and glossing methodology of the DRSP observe the linguistic variation, through which much of the extralinguistic SL information remains intact.

We analyzed translation, glossing and tagging approaches to the DRSP in the BaLDAR, lexical The examples include lexical additions, antonymous translation, transposition of Russian barbarisms, terms

originating from the communist regime, examples of lacunae, an ethnonym, and a toponym.

The paper confirms the influence of the Russian language in the Adjarian dialects. Further research will be carried out to discern the traces of the Pontic Greek language in the Adjarian dialects archived in the DRSP.

### **List of abbreviations:**

DRSP digitized recorded speech patterns

BaLDAR Batumi Linguocultural Digital Archive

GDTM Georgian Dialect Translation Methodology

POS tagging part of speech tagging

SE – Standard English

MSG Modern standard Georgian

## REFERENCES

Dijk, 1977: Teun A. van Dijk, “Text and Context: Explorations in the Semantics and Pragmatics of Discourse”. London: Longman.

Tandaschwili, Khalvashi, Beridze, Khakhutaishvili, Tsetskhladze 2017: Manana Tandaschwili, Ramaz Khalvashi, Khatuna Beridze, Mzia Khakhutaishvili, Nana Tsetskhladze, Batumi Linguoculturological Digital Archive (contemporary Technological Achievements for the Database Arrangement of the Folklore Resources) TEMPUS International Journal of Multilingual Education. pp. 52 - 68;

Giorgadze, Abuladze, 2019: Giorgadze Marine, Abuladze Teona, “Diaspora and Migration Studies: Pontic Greeks in the Post Soviet Period”. Proceedings of the international conference: East European Multicultural Space - ISBN 978-9941-462-72-6.

Links:

The Armazi Project

<http://armazi.uni-frankfurt.de/armaziII/enebi/kartuli/dialekt/acaruli.html>

English-Norwegian Parallel Corpus: Manual Stig Johansson Jarle Ebeling Signe Oksefjell Department of British and American Studies University of Oslo, 1999/2002

<https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/ENPCmanual.pdf>

Multilingual Education, journal <http://www.multilinguaeducation.org/en/article/37>

Social Variation in Language, ch. 19; [http://www.dsglynn.univ-](http://www.dsglynn.univ-paris8.fr/Yule_2014_The_Study_of_Language_Chpt19.pdf)

[paris8.fr/Yule\\_2014\\_The\\_Study\\_of\\_Language\\_Chpt19.pdf](http://www.dsglynn.univ-paris8.fr/Yule_2014_The_Study_of_Language_Chpt19.pdf)

The Leipzig Glossing Rules:

<https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>

The Reichenau Glosses:

[https://is.muni.cz/el/1421/podzim2008/LJMgrB23/um/texty/glosy\\_Reichenau.pdf](https://is.muni.cz/el/1421/podzim2008/LJMgrB23/um/texty/glosy_Reichenau.pdf)

The Scottish Corpus of Texts & Speech (SCOTS)

<http://www.helsinki.fi/varieng/CoRD/corpora/SCOTS/index.html>

Scots Corpus <https://www.scottishcorpus.ac.uk/>