Statistics in Empirical Translation Studies (Challenges of Rustvelology in the Digital Age)

Tandaschwili Manana, PhD Goethe University of Frankfurt Frankfurt, Germany

ORCID: 0009-0005-7812-9124 DOI: https://doi.org/10.62343/cjss.2023.246

ABSTRACT

The present paper addresses the problem of data processing in multilingual parallel corpora. It focuses on the difficulties that can arise in the statistical processing of linguistic data in a multilingual parallel corpus, such as Rustaveli Goes Digital and the solutions that may be useful for overcoming the challenges of empirical translation research. During the statistical analysis of the corpus *Rustaveli Goes Digital*, we encountered certain problems that we will discuss in this article, namely the reliability of the statistical analysis in creating the index. Although many different ready-made tools are successfully used in linguistics for statistical analysis, the data processing of texts can still be very inaccurate without considering the grammatical characteristics of the languages. As empirical material, the text of the epic *The Knight in the Panther's Skin* was chosen in three languages: Georgian, Abkhazian, and Megrelian. The paper will show why ready-made tools such as KWIC and Voyant are not suitable for Caucasian languages and what problems the use of such tools can lead to.

Keywords: Caucasian languages, digital Rustvelology, translation studies, data processing

INTRODUCTION

The origins of Quantitative Linguistics date back to ancient Greece and India. One strand of tradition consists of the application of combinatorics to linguistic objects (Biggs, 1979); another is based on elementary statistical surveys, which are referred to under the keywords colometry and stichometry (Pawłowski, 2008). A thematically broader and more continuous development of quantitative linguistics (QL) began in the 19th century. Among other things, this involved sound and letter statistics as preparatory work for the development of stenographic systems and as a basis for language comparisons, the different forms of verse and the duration of sounds in relation to word length, and even the exact dating of an author's works. The studies on sound length and ideas on the interaction of other linguistic characteristics presented the first concepts that led to the development of language laws in the 20th century, most famously Zipf's Law. In the 20th century, several other topics were added: identification of anonymous authors, action quotient, language structure, language change law, type-token relation, development of children's language skills, dynamic aspects of text structure, etc. The objective of QL in the 21st century is more demanding – the formulation of language laws and, ultimately, of a general theory of language in the sense of a set of interrelated language laws. The present paper focuses on the questions of what kind of difficulties can arise in the statistical processing of linguistic data in a multilingual parallel corpus such as Rustaveli Goes Digital, and what solutions can help overcome the challenges of empirical translation research.

Parallel corpus 'Rustaveli goes digital'

Shota Rustavelis Epos *The Knight in the Panther's Skin*. The Epos by Shota Rustaveli is the most significant literary work on Georgian intangible cultural heritage. The Epos was created in the 12th century and has been handed down in over 160 different manuscripts. Its significance has gone far beyond Georgia's borders and now has a prominent place in the history of world literature: the collection of manuscripts of the Epos is included in the UNESCO World Intangible Cultural Heritage Register. The Knight in the Panther's Skin is an excellent literary work and one of the most crucial components of defining the identity of the Georgian nation. The research of this unique literary work with modern methods is not only a challenge for the Kartvelology of the 21st century. However, it will also contribute to the scientific research of Georgian intangible cultural heritage and the internationalization of modern Kartvelology. The creation of a parallel corpus of the Epos' translations is

an important step for conducting interdisciplinary research. In addition, the multilingual parallel corpus can be successfully used in bilingual/multilingual education.

Research the history of the epos and modern challenges. Scientific research on the epos began in the 18th century when King Vakhtang VI added a scientific analysis to the first printed book from 1712. This formed the basis for further research on the epos, which gradually developed into a separate field of Kartvelology – into **Rustvelology**.

The history of the Rustvelology covers more than three centuries and can be divided into several stages:

- 1. Textological research;
- 2. Textological-lexicological research;
- 3. The Soviet stage of Rustvelologian studies;
- 4. Interdisciplinary research;
- 5. Internationalization of Rustvelologian studies;
- 6. Digitization of Rustvelology.

The digitalisation of Rustavelology began in 2018 at the University of Frankfurt with the project *Rustaveli goes digital*, led by Prof. Manana Tandashvili, since 2023 by Dr. Mariam Kamarauli. The project aimed to create a **big data** in Rustvelology - a multilingual parallel corpus of translations of Shota Rustaveli's epic in 58 languages. This goal required the solution of the following tasks (Tandaschwili, 2022, p.53):

- I. Technical tasks:
 - conceptualization of the structure and design of the corpus and preparation of a technical framework;
 - digitization of the original text and its translations in 58 languages (including the digitization of several translations that co-exist in one language);
 - structural preparation of the texts for their inclusion in the parallel corpus;
 - connecting the digitized and structured texts with each other in accordance with chapters and stanzas;
- II. Methodological tasks:
 - conceptualization of the methodological framework for the study of translation strategies in the corpus;
 - development of a methodological framework for creating a basic concept of automatic processing of a poetic parallel corpus;

III. Theoretical tasks:

- aligning the multilingual parallel corpus and preparing the texts for interdisciplinary research (philosophical, religious, sociological, cultural-specific, astrological, etc. terms);
- verification of capabilities of automated translation strategies research.

METHODS

Statistical processing of the corpus

The multilingual parallel corpus *Rustaveli Goes Digital* (Beta version led by Dr. Mariam Kamarauli) currently contains 32 parallel translations of the full text of the epic in 20 languages (Georgian, German, English, Spanish, French, Italian, Turkish, Azerbaijani, Kyrgyz, Russian, Belarusian, Ukrainian, Greek, Arabic, Persian, Armenian, Ossetian, Lithuanian, Mingrelian, Svan).

We have already used statistical processing to analyze address formulas in the parallel corpus to determine and compare the strategies used by the translators. The analysis of the address formula in the translations revealed the following structures (Tandashvili & Kamarauli, 2023, pp. 99-101):

- 1. The addressee of the communication is lexically given in the address formula (sun); it acts as a vector of the communication channel and ensures the accuracy of the reference. The addressee of communication is often named directly before direct speech, in the initial position of the sentence.
- 2. An interjection in the address formula (o, sun) serves to open the communication channel and ensures its activation.
- 3. Using the second-person pronoun or possessive pronoun in the address formula (you, sun; my sun) expresses the speaker's status in the communication act.
- 4. Using both indicators of expressiveness (an interjection and a second-person pronoun or possessive pronoun) at the same time, "'O, my sun," increases the degree of expressiveness and gives more power to the information following in the direct speech.

We compared in 20 translations the statistics of equivalence degree of "sun" (as a denotative or connotative equivalence) and the address formulas in terms of the level of expressiveness. As it turned out, the frequency of use of denotative equivalents of "sun" is directly proportional to the degree of expressiveness (Tandashvi-li & Kamarauli, 2023, p. 101):

- 1. Those translators who have systematically chosen the denotative equivalent for "sun" in the address formulas are rendering them with a higher degree of expressiveness. This correlation is confirmed by a lower number in the "difference" column (especially in the case of Wardrop, de la Torre, Barea, and Martinez).
- 2. The correlation, established as the result of statistical analysis, is relevant from the point of view of a complex evaluation of the quality of a given translation because it clearly shows the translators' efforts to preserve as much as possible of the original not only the artistic language of the author but also his philosophical-religious and aesthetic worldview.
- 3. The results obtained using the corpus linguistic method indicate that the quality of the translation can be "measured" empirically. This, in turn, allows us to determine the strategies selected by the translator and the expediency and appropriateness of their application in the target text.

During the statistical analysis of the corpus *Rustaveli Goes Digital*, we encountered certain problems that we would like to discuss in this article, namely the reliability of the statistical analysis in the creation of the index.

Tokenization and accuracy of the statistical processing

In linguistics, a frequency class is a statistical measure of the frequency of use of a word in a natural language. Frequency classes can be considered on two linguistic levels: a single word form (token) or an entire lexeme with various grammatical forms. The most common statistical analysis is carried out by the type-token relation (TTR), used in quantitative linguistics and quantitative stylistics to measure linguistic diversity in a text. It is defined as the relation of unique tokens divided by the total number of tokens. When tokenizing a text, a list of tokens is created without considering its grammatical representation. In the case of inflected languages, an annotation is required not only to statistically record individual forms of the word but also to assign the various forms to the corresponding lexeme. The accuracy of the frequency of lexemes in a corpus depends heavily on how precisely the grammar of this language is mapped in the annotation system. Compare the frequency of words and lexemes in Vefxistqaosani in GNC.

Table 1Frequency of word forms

\leftarrow \rightarrow C \triangle Nic	ht sicher http://gnc.	gov.ge/gnc/o	verview					☆
Transkription	🐹 für K1 CL Sprac	ntypologie un	sfs Index of /~	gjaeger/l ៰៰៰	Pragmatik_13_Satza	🔇 Texte und Ma	terialie G	"Ubungen zur
ტექსტების სია	word [სიტყვა), distinct value	s: 209 036, ტიპი:	string, scope: c	pos			
ძიება კონკორდანსი კოლოკაციები	Mean length: 4.9 This table shows Sorted by fr	97; max length: all values of the equency O alg	117 e attribute, togethe phabetically	er with their corp	ous counts.			
കവിശര	633m@n 1 , 0mm	nsos(0: 747. F	Tevious	Go to page:			- 100-	
მიმოხილვა	220491, 118471.	3199 არის 3000 კიდევ	1607 სულ 1606 შეიძლება	1122 თვალი 1118 ამას	839 ხან 837 კაცს	685 ყველაზე 682 ყოველი	563 არამედ 563 მარა	
გრამატიკული მახასიათებლები	76361 და 52003 – 25280 არ	2987 რაც 2979 თავი 2837 ისეე	1605 ყველაფერი 1580 რადგან 1564 ხექს	1060 ერთად 1056 იცოდა 1050 ისიი	834 რამე 831 მისა 830 თავშე	681 ღმერთმა 678 გამო 678 იმა	562 ვინგ 562 სახლში 560 პირველი	
გაანალიზება	19811 რომ 16874 ?	2793 კაცი 2702 თავის	1547 ორი 1538 მარტო	1040 თუმცა 1017 გინდა	828 რამ 828 ყველას	673 თურმე 673 იყვნენ	559 იქნებოდა 559 ყველაფერ	າບ
	15786 ! 10772 : 9618 მაგრამ	2629 აღარ 2589 თქვა 2526 ჯერ	1532 წინ 1512 ერთ 1491 ბოლოს	1011 კარგად 1007 კარგი 995 იმას	821 თავდაპირველად 819 ვერც 814 ვეღარ	654 მინდა 649 მართალია 645 რომელსაც	558 რასაც 555 ბიჭი 555 თუკი	
	9599 9321 თუ	2408 შენი 2360 როგორ	1478 ყველა 1473 ჩვენი	985] 984 sn	814 ისინი 814 ყოველ	644 სადაც 638 შე	550 ზედა 548 რაღა	
	9237კი 9213 მე 9102 რა	2299 როცა 2294 (2294)	1460 იყოს 1456 ხოლმე 1448 დროს	977 უკან 970 მინა 956 რომელიც	811 თოოეთ 809 მაქვს 809 შორის	633 ოასაკვიოველია 630 კაცის 629 დედა	539 საქთეს 534 ყოველთვი 533 როდის	າບ
	8885 " 8806 nym	2285 თითქოს 2273 ხელი	1447 მთელი 1432 აბა	954 თქვენი 951 უთხრა	806 გაასწორა 803 ჩვენს 802 შარს	628 კაცმა 628 მართლაც	528 განა 528 თითქმის	
	7588 უნდა 6780 ეს	2214 სხვა 2206 მხოლოდ	1418 მეორე 1410 მას	949 003 944 საერთოდ 939 იⴋიტოⴋ	799 ერთმანეთს 797 ვითომ	625 ოთახში 623 უნდოდა	526 ყოფილა 525 ძლივს	
	6680 აშ 6179 არა	2156 აზ 2092 იმის	1408 ცოტა 1383 იქნება	939 ხელში 938 ხოლო	797 თვალები 789 დათა	619 ოღოზდ 607 არიან	522 ფული 521 აღარც	

Table 2Frequency of lexemes

\leftarrow \rightarrow C \triangle Nicht	sicher http://gno	c.gov.ge/gnc/o	verview				\$
🔠 🔇 Transkription 🎽	🕻 für K1 🛛 CL Spra	chtypologie un	<mark>sfs</mark> Index of /~gj	aeger/l مسع Pi	ragmatik_13_Satza	🔇 Texte und	d Materialie G "Ubunger
ძიება კონკორდანსი კოლოკაციები სიტყვათა სია ტექსტი	slemma [a This table show Sorted	მარტივი ლემა], d is all values of the frequency	istinct values: 39 5 e attribute, together phabetically revious Next	92, ტიპი: string, r with their corpus Go to page:	multi-valued, scop s counts.	e: cpos, abbreviat	ed syntax: //
მიმოხილვა	Constant of				Transienter		
າຕະອາສຸດ	220397,	3896 306	2302 როცა	1684 თქვენი	1203 თუთაშხია	989 sn	804 đ nđn
9(1)2026(13 JC(1)	118450 .	3886 თავისი	2294 (1683 დაწყება	1198 საკუთარი	985]	803 დაპირება
00000100130@13011	76706 და	3867 წდომა	2294)	1673 სახლი	1196 მეფე	962 სი <u>კ</u> ვდილი	803 3nonma
~ ~ ~ ~	51936 -	3856 03068	2294 ცხოვრება	1657 ძარტო	1193 0063	954 bmmm	801 jasma
გაანალიზება	44886 ??	3/10 შეძლება	2285 თითქოს	1620 63633	1191 ვიდრე	940 სტუძარი	794 ძოკვლა
	33338 არ	3698 არა	2270 მეტი	1606 606	1187 მართალი	936 კიდეც	784 გაჩენა
	26263 ymo3b3	3661 საქძე	2230 სად	1580 3000	1182 დაბრუხება	931 -	783 <u>კ</u> უ
	21458 NU	3621 0300	2219 ძნოლოდ	1580 ოადგახ	11/5 სწოოედ	921 dmbcomds	782 ძოყოლა
	19925 Mmd	3617 ისე	2156 ან	1563	1168 dðs	921 ყურება	778 ხე
	17735 ງ ນ	3454 გული	2148 ფეხი	1563 ცოტა	1160 ოთახი	921 X00mds	772 დადება
	16866 ?	3320 წომ	2124 603	1546 ამბავი	1154 Xoym	919 *	771 იქნებ
	16216 რა	3290 უფრო	2115 ბატონი	1538 ადგილი	1150 თან	910 რაც	771 მიღება
	15730 !	3277 კითხვა	2078 შემდეგ	1507 ბოლოს	1148 ბევრი	909 შინება	768 ძაღლი
	12642 მე	3266 დღე	2064 უკვე	1494 ადა მია ნი	1142 ღამე	905 შემთხვევა	766 სისხლი
	12224 თქმა	3254 რომელი	2056 ხალხი	1489 ერთმანეთი	ი 1138 გასვლა	899 ბოლო	766 ქუჩა
	10771 :	3225 როგორც	2055 მეორე	1481 სახე	1127 ცხენი	898 გიორგი	765 მზე
	9569	3199;	2012 ჯერ	1459 ხოლმე	1122 ვიღაც	898 ყური	763
	9521 მაგრაშ	3092 შვილი	1979 სიყვარული	1457 ყოლა	1120 ფული	895 đ nõ	759 მთავარი
	9348 org	3082 დრო	1913 მთელი	1442 აბა	1119 ახალი	892 მხარი	748 ეზო
	8965 30	3052 არაფერი	1911 სიტყვა	1436 ცოლი	1114 გასწორება	879 საერთოდ	739 უფროსი
	8873 "	3027 ებ	1886 0ანი	1435 ალბათ	1113 ქალაქი	876	736 საუბარი
	8286 "	2937 მოსვლა	1835 მი-ცემა	1424 გაგება	1096 ლოდიწი	873 სანაშ	730 მაშიწვე
	8220 mago	2872 0210	1828 200	1418 პატარა	1093 დარწაოზ	872 *ຄຳຄົນ	729 mman

The most frequentative ten word forms (tokens) in GNC: *da* და *and*, *ra* რა *what*, *ar* არ *not*, *me* მე *I*, *tu* თუ *if*, *mas* მას *he/she/it* (*Dat.*), *iqo* იყო *was*, *rom* რომ *that*, *ese* ესე *this* (*also as an definite article*), *igi* იგი *he/she/it*.

VS.

The most frequentative ten lexemes in GNC: *da* დs *and*, *qopna* ყოფნა *to be*, *is* ob *he/she/it*, *ra* რs *an*, *ar* არ *not*, *es* ეს *this*, *misi* მისი *his/her/its*, *me* მე *I*, *kaci* კაცი *man*, *čemi* ჩემი *my*.

GNC can output statistics according to word class (noun, adjective..), semantical roles (subject, object), functionality (focus), as well as the grammatical features: case, person, TAM, genus verbi and so on.

Table 3

Frequency of grammatical features

\leftrightarrow \rightarrow C \triangle Nich	t sicher http://gnc.go	v.ge/gnc/overview					\$
Transkription	🔀 für K1 🛛 CL. Sprachty	pologie un <mark>sfs</mark> Ir	idex of /~gjaeger/l.	🚧 Pragmatik_1	3_Satza 🔇 Te	exte und Materialie	G "Ubungen zur Klau
კოლოკაციები სიტყვათა სია ტექსტი მიმოხილვა	features [dashs This table shows all Sorted by frequence	ასიათებლები], distino values of the attribu uency 〇 alphabetic	t values: 18 419 (c te, together with th ally	listinct atomic value eir corpus counts.	s: 318), ტიპი: set	, multi-valued, scope:	cpos, abbreviated synta
გრამატიკული მახასიათებლები	გვერდი 1 , მთლიან	iso: 2. Previous	Next Go to pag	ge:			
	4164200	81804 Pass	30094 SV	11432 LV	4718 Area	1756 Ext	400 <der:ელ></der:ელ>
გაანალიზება	524864 N	77770 Rel:8	29007 Poss	11311 Refl	4663 DSg	1746 DDat	365 ConjPerf
• <u></u>	468669 Punct	71564 NewPl	28119 Dialect	11234 Poss3Sg	4588 IntMark	1728 Encl:მდა	352 Range
	447467 Foc	71100 Pres	27747 Loc	11135 Root	4522 DNom	1724 Dir	327 Sent
	387645 PI	70563 Prop	27582 S:1Pl	10868 Impv	4445 PP:შუა	1565 Recip	315 Zoon
	354959 Sg	69433 Advb	27459 S:1Sg	10853 Trunc	4429 DOIdPI	1486 <io:gen></io:gen>	309 Distr
	336545 V	67281 MedPass	27194 Num	10533 Cond	4415 Ord	1479 <auxtranshu< td=""><td>m> 304 Encl:dg</td></auxtranshu<>	m> 304 Encl:dg
	303677 PP	62704 PP: დან	26648 PP: J300	10442 Meas	4398 PP: aconu	1298 Foreign	297 <der:ന്ന്യാ)< td=""></der:ന്ന്യാ)<>
	296610 Encl:IndS	p3 61130 DO:2Sg	26251 Ben	10422 PP:თან	4274 IO:2	1267 [Excl]	294 <dat gen=""></dat>
	294686 Nom	60406 Erg	25531 Fut	10292 IO:1Sg	4106 PP: don6	1180 <der:ດან></der:ດან>	282 >??
	291979 Encl:8	55751 Pers	25405 PP:dcon	10033 2	4054 NegPart	1173 Imperfective	278 DAdvb
	277112 OldPl	54191 Comma	25147 SIndef	9927 Rel	3891 Poss2Sg	1106 Causal	273 Disc
	270078 >P	52950 <s:dat></s:dat>	25019 Alpha	9863 Perf	3753 Dist	1073 PP: 600	268 <gen></gen>
	228539 Pv	52659 PP:30	24979 <der:ურ></der:ურ>	9686 Pp	3749 <der:თერ< td=""><td>> 1047 Deg</td><td>261 DL</td></der:თერ<>	> 1047 Deg	261 DL
	218144 Colon	52651 <do:dat></do:dat>	24957 Encl:3)	9603 Encl:IndSp	1 3326 Abs	1016 DDSg	259 PP: მებრ
	216995 Old	51235 <s-io></s-io>	24886 <auxintr></auxintr>	9569 Dash	3199 Quote	1005 Coll	253 >ADV
	189314 L	50583 Anthr	24520 Nonhum	9115 PP:თვის	3188 PP:01	977 PP:ებრ	232 Letter
	187906 A	49558 <s-do-io< td=""><td>> 24025 Neg</td><td>8989 Place</td><td>3181 Org</td><td>973 PP:31</td><td>223 IO:1</td></s-do-io<>	> 24025 Neg	8989 Place	3181 Org	973 PP:31	223 IO:1
	182855 Adv	49260 Part	21940 Opt	8528 DO:3Sg	3034 IO:3PI	946 DDVoc	194 PP:(0)
	181804 <s:nom></s:nom>	48688 Temp	20800 Card	8435 LastName	3009 Med	929 < OldPl>	177 Anim
	179102 Dat	48569 Encl:30	20281 PP:do	8219 PP: ტან	2960 DO:1Pl	920 Symbol	166 DO:1
	153897 Pron	47562 IO:2Sa	18877 <name></name>	7816 Encl:IndSp	2 2944 PassState	880 PP:0con	141 Conj-II
	142552 <s-do></s-do>	45517 MedAct	18375 Dem	7796 Encl: 00	2937 Encl:ð1	822 LAT	132 Frac
	139287 <do:non< td=""><td>n> 44369 VN</td><td>17422 Ouant</td><td>7780 FutPart</td><td>2914 Approx</td><td>797 Iter-II</td><td>130 Compl</td></do:non<>	n> 44369 VN	17422 Ouant	7780 FutPart	2914 Approx	797 Iter-II	130 Compl
	136700 Act	43094 Impf	17401 Encl:0	7601 Indef	2818 Poss1Pl	792 [OV]	127 >0B1

The higher the annotation quality of a corpus, the more accurate and precise the results of the statistical processing of the linguistic data are.

DISCUSSION

For valuable evaluation of the statistical data, the specific characteristics of the respective languages should be taken into account; otherwise, the results of the statistical processing of the linguistic data will be inaccurate. This is particularly important for statistical analyses in parallel corpora. Below, we will present this on

the example of the multilingual parallel corpus Rustaveli Goes Digital in the case of Abkhazian and Megrelian translations of the epic The Knight in the Panther's Skin by Shota Rustaveli.

Tokenization in Abkhazian

In computational linguistics, tokenization refers to the segmentation of a text into word-level units. A token is a character string that is assigned a type by a formal grammar. The token forms the basic lexical unit for the parser. As mentioned above, the accuracy of the frequency of lexemes in a corpus depends heavily on how closely the grammar of that language is mapped in the annotation system. We will show here what happens when tokenization a non-annotated corpus of Abkhazian. As a sample text, we take two Abkhazian translations of Rustaveli's epic, which are by Dimitri Gulia and by Mushni Lasuria.

Frequency of pronouns

The frequency of use of pronouns is one of the central common statistical indicators during the automatic processing of texts. The first and second person pronouns are generally characterized by the highest frequency among the personal pronouns. The word frequency via KWIC of the Abkhazian translations is given in the table 4 below:

Table 4

Comperison of the word frequency in both Abkhasian transitions by Gulia and Lasuria

Dimitri	Gulia	Mushni	Lasuria	
Frequency	Token	Frequency	Token	
9937	,	11829	,	
1999		1564	1	
1666	;	1447	-	
1537	уи	1281		
1075	capa	888	УИ	
967	ca	810	:	
940	yc.	762	3D	
822	убри	757	«	
768	«	703	ca	
687	:	541		
680	»	395	?	
644	ya	370	зегь	
591	YPI	316	capa	
580	yapa	303	YC	
390	I. Contraction of the second s	289	ya	
360	абра	249	YPT	
348	убра	228	уара	
317	yaka	192	иара	
308	иара	187	автандил	
297	нас	172	;	
294	хәа	157	xa	
285	абри	152	нас	

The frequency of the personal pronouns (1st, 2nd and 3rd person in singular and plural) is listed individually in the corpus:

Gulia	Lasuria
capa 1075	ca 703
ca 967	capa 317
ya 644	ya 289
yapa 580	yapa 228
иара 308	иара 192
қа 176	қа 157
and so on	

The statistical processing of the texts via Voyant tools (https://voyant-tools.org/) can be visualized by five most frequently occurring words.

Figure 1

Visualization of the statistical processing of Dimitri Gulia's Abkhazian translation

\leftrightarrow \rightarrow C (a voyant-tools.org/?corpus=024ee3d8558e56	b3a68abe7f99bb21af					G	e \$ \$	🗖 😻 :
	۲	Voyant Too	ols					?
Cirrus 🖽 Corpus Terms 🔩 Links ?	Reader 🔿 Wo	rttrauben		?	2 Trends	Dokumentbegri	ffe	?
apparation	1 Алагарта 1.31 Адун 3 1 Алагарта 1.31 Адунен зегь іннеилькомур Хакаге акеранскихе гром зшаз ус кунагала, Ихратая хара вуда, адлык Ист. ит обла видь инсекци	нем зегъ и а изшаз зхат пъны илбаац л еилоу хуаг	ноилымхку еылшара мчр автны пусыахо пуштеы хоы р	га изшаз ала, лу зегьы ацеала,	0.0040 0.0035- 0.0025- 0.0020- 0.0005-	с уи убри	сара с	а
Begriffe	< >		~ ?	8 ~		× ?	Zurücksetzen	Darstellu
Zusammenfassung III Dokumente III Phrasen		?	III Kontexte	Bubblelines	E Kollokation	ien		?
WORSCHALZUICHTE: 0,344		-	Dokument		links	Begriff rechts		
Readability Index: 17.379		E	1 Алага	сыгунахара. 3. а	3 Алым иатеоул	уи икразы	аљса, аса, ус а	кулзгы
Durchschnittliche Wortanzahl pro Satz: 40.1		Œ	1 Anara	икразы аљса, а	са, ус акулзгыы.	уи ихуалы	шуа шьақар-хшл	а итеуеи
Die häufigsten Begriffe im Korpus		Œ	1 Алага	убас зызъы, Са ис	ыздырам исы	уи лыреж	уара сара ашеа	ны. 4. ә4
• yet (1868); capa (1345); ca (1241); yc (1149); y6pet (1007)		* E	1 Anara	Тамар , дхарехуал	қара лағырз	уи лзы ис	ымоуп сара аши	накуа, егь
Contraction of the second s		• R	1 Anara	пзы изывразы аши	а хаакуа Ис	ум пыны	мшыха пыблаз	W DISSERT
Elomente				× ? 1	868 Kontext 💿	erweitern	•	
	Voyant Tools , Stéfan Sincla	ir & Geoffrey Ro	ckwell (8 2023) Pi	wacy v. 2.6.10				

The frequency of personal pronouns in the Georgian-Abkhazian parallel corpus of "Vefkhistqaosani" shows much more grammatical varietes in the Abkhasian translations than in the source text itself. This is due to the differentiation of personal pronouns according to form and semantic class in Abkhasian. There are two forms of personal pronouns in the Abkhazian language: long forms such as *capa* (1pers.), *yapa, δapa* (2. pers.) (marked with -pa) and short forms - *ca, ya/δa* (without marking). The second personal pronouns in singular are additionally marked according to genus (*yapa* M and *δapa* F) and the third personal pronouns in singular also (*uapa* M, *πapa* F, *uapa* N).

Figure 2

Visualization of the statistical processing of Mushni Lasuria's Abkhazian translation

← → C (wyant-tools.org/?corpus=5695d3a9bf003	f6f97e304ec5ccd14e3							G	R 6	* *	🗖 😻 i
		Voyar	nt Too	ols							?
Oirrus	Reader O Worttra	uben			?	Image: Trends	I Dokur	nentbegriffe	0		?
закат тариел Иара зарина уча и париел Иара арри уча и париел Иара арри уча и париел Иара арри уча и париел Иара арри уча и париел Иара арри арри арри арри арри арри арри	Алагарта Адунеи ду, эх эылшарал Алагарта Адунен ду, зхваа нымтаео, аб Жеван нахмау псыт,чшалана Харт, ауаатемаса, нахоноту Ипшра-искиха изшывалоу Ипшра-искиха изшывалоу Сыхина, сухвоит, исызценерга	еаа ни ас зыл аегь ры у сахьа цеекж цеекж а ацьны	ымца потац -хкы р нтқар ш сы:	ао, абас ла изшаз, тара зылшаз, рацаа заырба цаа загь убас су загь зша! алаииааиша	a.	0.0005 0.0000 0.0005 0.0000 0.0005 0.0005 0.0005 0.0005	yc yc	уи са	apa ca	эегь 7 8 каз нымтро, г	9 10 абас зылшарал)
Begriffe (•	< >		Y	?	8 ~		2	? Zu	rücksetzen	Darstell	lung 🗸
Zusammenfassung Wortschatzeichte: U-420 Readability Index: 21.052		?		Dokument	۱۱	Bubblelines 🗄	Kollokation links	en Begriff	rechts		?
Durchschnittliche Wortanzahl pro Satz: 13.6		- 1	E	Anarap	Тамар	ал, аднынтқар-гр алтшеып, ласыр	киала	yu vu	, зыхцен-	лзыскован.	сазыразн
Die häufigsten Begriffe im Korpus		- 1	Ð	Алагар	Шеааи	, Тариел из хлар	ырзашап	уи	далсоуп,	Дарбан хата	оу уи ицна
• уи (891); Са (711); Эель (371); Сара (320); ус (310)			æ	Anarap	афырх	атца уи дапсоуп,	Дарбан	уи	ицназго?	– ус уаа дый	кам, деиба
()		•	.FR.	Anaran	ннааха	Tac wubanovn	- vanfia	VM	азханила	агаы азыбы	IIIVA – KUNH
Lighter (- 1 0al M	UTINDAL (erwe			

Stéfan Sinclair & Geoffrey Rockwell (8 2023) Privac

The comparison of the frequency between the long and short forms of the personal pronouns is shown in the following diagrams:

Translation by Gulia

Translation by Lasuria



Figure 5



Figure 7



Figure 4



Figure 6



Figure 8



The more precise statistics in both translations show the differences in the use of long and short forms of personal pronouns according to the genus. As the comparison of personal pronouns differentiated by gender shows, the masculine personal pronoun occurs more frequently than the feminine personal pronoun: 2. Pers. Pron. M - 1224 (*yapa* 580 / *ya* 644), 3. Pers. Pron. M - 310 (*uapa* 308/ *ua* 2)

cf.:

2. Pers. Pron. F - 200 (*6apa* 138 / *6a* 62), 3. Pers. Pron. F - 183 (*napa* 140/ *na* 43)

Table 5

Comparison of long and short forms of personal pronouns in the Abkhazian Translations

Personal pronoun	Genus	Gulias Translation	Lasurias Translation
I person	-	capa 1075	ca 703
		ca 967	capa 316
	М	ya 644	ya 289
II person		yapa 580	yapa 228
	F	бара 138	бара 48
		ба 62	ба 48
	М	иара 308	иара 192
III person		иа 2	иа 1
L L	F	лара 140	лара 90
		ла 43	ла 56
I person, pl.	-	қара 201	қа 157
		x a 176	қара 58
II person, pl.	-	шәара 95	шәара 33
		шәа 53	шәа 15
III person, pl.	-	дара 90	дара 59

Regarding the frequency of use of long and short forms of personal pronouns, Gulia clearly favors the longer forms (except the 2nd personal pronoun masculine). However, Lasuria presents a different picture: the longer forms are favored only for the first personal pronoun in singular and plural, as well as for the second personal pronoun feminine. The question of what causes the high frequency of long or short forms in Lasuria and Gulia's translations requires additional corpus-linguistic and contextual analysis. This is a separate research topic, and we will not address this issue here.

This statistical analysis confirms the need to account for the grammatical features of the language when annotating the corpus in order to capture both the general (part-of-speech) features and the specific characteristics. For instance, these words in Abkhazian should be annotated as personal pronouns, but also according to gender and form (long or short). We did this manually in our case, but in an annotated corpus, this should occur automatically.

Frequency of nouns in Abkhazian

The nominal morphology of Abkhazian differs from Kartvelian and the East Caucasian languages: Abkhazian has no declension, only the category of number, definiteness and possessiveness. This phenomenon is illustrated by the nouns in Abkhazian: the nouns are often marked by definiteness or possessive markers, which appear as prefixes to the nouns. The lexeme $\underline{\partial b} g$ sun, which occurs 309 times in the original text, corresponds to several inflected forms in the Abkhazian translations, which are marked by possessiveness differentiated by gender and thus result in the content of a noun phrase.

Table 6

Comparison of the "sun" in source language and target language

	Rustaveli	Gulia	Lasuria	Grammatical category	Content
		амра 134	амра 111	Definiteness	the sun
			рымра 4	Possessiveness, number (3.Pl)	their sun
Word alto sun"		сымра 17	сымра 11	possessiveness (1. Sg)	my sun
mora	0.00 ,,5111	умра 5	умра 10	Possessiveness, genus (2.Sg.M)	your (M) sun
		қамра 3	қамра 10	Possessiveness, number (1.Pl)	our sun
		имра б	имра 9	Possessiveness, genus (3.Sg.M)	his sun
Frequency	309	169	176		

For clarity, we will cite some examples from the source text of the epos and the Abkhazian translation by Gulia:

1.51. Тинатингьы амра иапылган, иатахын амра тинатинзарц! *Tinatin was more beautiful than the sun, the sun wanted to be Tinatin.*

Cf.

თინათინ **მზესა** სწუნობდა, მაგრა **მზე** თინათინებდა. *Tinatin resented the* **sun**, but the **sun** was shining.

38.920 Автандилгьы дигәалашәеит имра лаша, дызбылуа,

Avtandil also remembered his bright sun, which burns him.

Cf.

ავთანდილსცა მოეგონა **მისი მზე** და საყვარელი. Avtandil also remembered **his sun** and lover:

34.820 Аҳәынҭқар: "Иаҳцәыӡма ҳамра, имзахама, нарха змам?"

King: "Do we have lost **our sun**, it has become a moon without life?"

Cf.

მეფემან ჰკითხა: "წასრულა **მზე** დაუდგომლად, მთვარულად?" *The king asked: "Has :gone, quietly disappeared like the moon?"*

The Georgian does not have a grammatical category of definiteness. In the case of the lexical item *"sun"*, however, the reference is clearly definite (on the semantic level). This is morphologically marked in the Abkhazian translation by the **a**-prefix: **a-мpa** (1.51).

Table 7

Expression of definiteness in Abkhazian

Language	Lexem	form of language expression	Level
Georgian	მზე	Implicative expression of reference	Semantic level
Abkhasian	а-мра	explicative expression of reference	Morphologic level

The following example demonstrates the ability of the Abkhazian language to indicate the category of possession in nouns by means of prefix morphemes, which are additionally differentiated by gender in the 2nd and 3rd person singular. In our case, it is the noun *"sun"*, to which the masculine possessive prefix of the 3rd person **u**is added: **u-mpa** (38.920). This noun in Abkhasian corresponds to the noun phrase **∂olo ∂bj** *"his sun"* in the source text:

Table 8

Comparison by expression of the possessiveness in Georgian and Abkhazian

Language	Lexeme	Structure	Form of language expression	Level
Georgian	მისი მზე	NP	Explicative reference expression (with person and deixis specifi- cation)	Morphosyntaktic level
Abkhasian	и-мра	N	Explicative reference expression (with the specification of person and genus)	Morphologic level

Unlike the previous example, in this case, the reference is explicitly expressed in both languages, however, in addition to the difference in grammatical categories, they also differ from a structural point of view, which is both from the point of view of quantitative processing of the text (statistical analysis, e.g. during tokenization) and from the qualitative point of view (in translation studies, when parallelizing the text in establishing equivalence purpose) creates certain problems:

Avtandil also	remembered	his sun	bright	which burns him
Автандил-гьы	дигәалашәеит	и-мра	лаша	дызбылуа

ავთანდილსცა	მოეგონა	მისი	მზე	და	საყვარელი
Avtandil also	remembered	his	Sun	and	lover

The third example differs significantly from the two previous cases:

34.820 Аҳәынҭқар: "Иаҳцәыӡма ҳамра, имзахама, нарха змам?"

King: "Do we have lost our sun, it has become a moon without life?"

Cf.

მეფემან ჰკითხა: "წასრულა **მზე** დაუდგომლად, მთვარულად?" The king asked: "Has **the sun** gone [from us], quietly disappeared like themoon?"

In the Abkhazian translation, the noun sun xampa (34.820) is accompanied by the x- prefix of the 1st person plural. A two-person verb renders the predicate in the Abkhazian sentence:

иа - ҳ - цәыӡма	VS	წასულ-ა
$DO_{3Sg.} - S_{1pl.} - V_{tr.}$		V _{Int.} -S _{3Sg.}

The grammatical and pragmatical modification of the Georgian verb in the Abkhazian translation $(V_{Int} - S_{3Sg.} > DO_{3Sg.} - A_{1pl.} - V_{tr.})$ is conditioned by the context: the departure of Avtandil causes the regret of the king Rostevan and also the royal court of Arabia. Accordingly, in the Abkhazian translation, the translator changes the perspective of king Rostevan's statement: Avtandil's departure is told from the perspective of the king, which causes a grammatical change in the predicate of the Abkhazian sentence: an additional actant enters the verb иа-х-цәызма (1st person plural), which is also reflected in the noun through the possessive marker: х-амра.

Table 9

Language	Lexeme	Syntaktische Funktion	Structure	Features
Georgian	მზე	Subject	Ν	N _{Nom.Sg.}
Abkhasian	қ- амра	Direct object	PossPron+N	N _{sg.} +PossPron. _{3Pl.}

This strategy used by the translator creates certain problems when parallelizing the text (in order to establish equivalence):

King	Do we have	e lost	our sun	it has l	become a moon	life	having without
а-ҳәынҭқар	иаҳцәыӡ	ма	қамра	V	імзахама	нарха	змам
			\backslash		` ` `	· _	/
მეფემან	ჰკითხა	წას	რულა	მზე	დაუდგომლ	აად	მთვარულად
The king	asked	to ha	ive gone	sun	disappeare	ed	like the moon

A few examples given here are only a hint of the problems that can arise during the statistical processing of the Georgian-Abkhazian parallel corpus via simple statistical analyses due the texts are not annotated. Today, only a simple search in the corpus is possible:(See Figure 9)

The above problem was solved by Paul Meurer in Abkhazian National Corpus (*The Abkhaz National Corpus*, n.d.) The AbNC was developed in the years 2016–2018 in a project financed by USAID, with participants from Sukhumi, Tbilisi, Frankfurt and Bergen. It comprises more than 10 million tokens of texts from various genres and is morphologically annotated. The corpus is hosted in the Corpuscle corpus management tool, which has advanced possibilities for searching and viewing the corpus texts. Simple search allows the search to word forms, but the advanced search allows you to search by word, lemma, slemma, stlemma or grammatical features.

The search can be limited to certain subcorpus or text, as in the given case: word form is searched only in the Abkhazian translation of Shota Rustaveli's epic.

Figure 9

C s geabco.com/tools/wordsea	rch		© (☆
Georgian-Abkhazian Corpus	Translations	Tools 👻 Info	
Wordsearch			
Word		Translation	_
амра	Exact	M. Lasuria (1978), D. Gulia (1959)	 Search
168 occurences found.			
M. Lasuria (1978) [0.23.]		Амицьнур, нас, ипшра-исахьа – <mark>амра</mark> дацлабуеит, длашоит,	
D. Gulia (1959) [0.23.]		Абзиабае дкаиматзароуп, уи <mark>амра</mark> еиљш дкаљхозароуп,	
D. Gulia (1959) [1.37.]		Аецә амра изацлабуам, аха уеизгьы икуалкуацьсит.	
M. Lasuria (1978) [1.38.]		Иамеоуп уи, уҳәынҭқарра лыҭа, <mark>амра</mark> иацналгоит, дымроуп!	
D. Gulia (1959) [1.38.]		Гушрала <mark>амра</mark> зыцазыр <u>т</u> уа дќаца хәынтқарс, аус дазхоит.	
D. Gulia (1959) [1.43.]		Зегь шәлырлашап, амра дафыҧшны, шәылцәымшәан, уи шәарт аказы,	
D. Gulia (1959) [1.49.]		Агуили нас абабиреи еиљшны убасгьы амра рыкуљхоит,	

Search result in Georgian-Abkhazian corpus

The difference between statistical analysis of raw texts and annotated texts is enormous. Precise statistical processing of the data in the parallel corpus Rustaveli goes digital requires a high quality of annotation for all languages integrated into the parallel corpus so that the parallel corpus can be used efficiently for statistical data processing.

Figure 10

Advanced search of lemma a-mpa in Abkhazian National Corpus



Figure 11 Concordance of lemma *a*-mpa ,, the sun"

+ → C ==	clarino.uib.no/abnc/concordance		🕸 🖈 🙂
Ађсуа би	ызшаа амилатта корпус ihe Abkhaz National Corpus сесесе		English 👻 🗌 Use transliteration Sign in
Home	Concordance		
Using the corpus Documentation Publications	Advanced search switch to Basic search Query history [lemma = "ā-spa" %c] :: author = "PycŢaseʌu Woʒa"		
Catalogue Query	Run Query Refine window: document I Stop I Saved queries Done. Real time: 0.12 sec. (0.13 CPU sec.) I Stop I Saved queries		
Concordance	Type: kwic V Att: word V Show line filter Attributes Structures (show in	match) Page size: Context size: 500px v
Collocations Word List	Hit 1 - 30 of 375 Previous Next Go to: Download (Excel mode)	Copy que	ry URL
Text	cpos	match	
Overview Metadata	S864419 шоз, рымцәнжәаақуа ҵака ишьтахаразонт, Абинабае дКаннатзароуп, уи S864926 і динанызт, ака ипука ус дкапухари, Астар деиршын убри пршрала, зегь S865040 ак иециеонт, Убри кашеонт, ацынхурас қаға абақчасы дағак нашатүент,	амра мра Сымра	еигдш дкагдхозароуп, Кыр еилицо дбеиазароуп, днапкымзари хәшьас хәа уи дрыгдхьазон, Уи атыгдха инлыхуагдшыз иажаа лакуит, лашьцарак збоит, лашара зқум атдх ааигуоит Затдааи
Parse	5865094 ашара, атдых лашьца уи иазыгршзоуп, Хәынтқарс дкашатда сара сыгрха,	сымра	лаша убри даљсоуп!» Атакс иархаеит авазиркуа: «Уара уаж
Grammatical features	5865140 калонт, Уи нафаы-жаа нас ашатыцра уи изцрятуля, уи изанинонт, Аеца 5865191 казонт, Хын нарао игу ус каналаго уи закуанут, уи инатухонт, Бырала 5865424 касцеит сцынхурас абра хамнткарс уи шаара шааы, Зеть шалырлашап, 5865691 р хшывла, ус кунатала аусбара. Агуили нас абабирен енгушны убасты 5865518 ан, ижкуан, авара длурьен, караф ус иуал, Тинатин амра лапталараусы 5865556 быкуз уара уажаа слу иманалсыза у саказы!» Нас иннкуит нара илук, 5866556 быкуз уара уажаа слу иманалсыза у саказы!» Нас иннкуит нара илук, 5866556 быкуз уара уажаа слу иманалсыза у саказы!» Нас иннкуит нара илук, 58667046 кас какарку, и манарсара амаху факуа, Аекуа ршыатукы, нартикы за саба 5868333 атин, убра илдырт ус лаб илахь еккута марар, Диагилан днеит ашаер, 5869990 казант бара багы итаз у кантей такара. Саказы!» Кантикант баларт сара слаж сака сыс-кыз ба баз итаз у кантей казант саба баз казант у калтей калте ба баз итаз у калтей калте ба баз казант у калтей калте бала казант саба баз казанта у калтей калте ба баз казантаз у калтей калте баларт.	амра амра амра амра амра амра амра амра	изатдлабуан, ака уеизгээ икуалкузцьонт. Уара, ахаангүдар, т зактаануру аката ханжгүрэг, эку саяхогт. Хэльнгүдэр (дала даеыгушны, шөылцөымшөан, уи шөарт аказы, Шааан зегь аб рыкутухонт, Дугыэ хүүгээ инейбенушны хатыр хүтцэар урт и литальгурын, мара уака тингира ауан, Даалыгухент үн үн үзба тинатинра ауан, Даалыгухент үн иланзузен, гук-гүсэк ал зтаангуры, гүрагаразы. Автандил үз үбас ихаент: «Сара изг ашаахуа нак ихнаалуан, Ильсуган, иршуула, дааы зегь шша енгушны дкалкух убра лара, Ахтын кахига екплка даццоу д анцаа норс бахницаа зак. Харан лаширбаг ба бах иланг

Special features of the rendering of negation in Megrelian

In this section of the paper, I will further discuss the problematic aspects of statistical data processing in the multilingual parallel corpus *Rustaveli goes digital* using the example of the Megrelian translations of the epic. In particular, I will address the issue of how the category of negation is rendered in Georgian and Megrelian and the challenges of tokenization in the Georgian-Megrelian parallel corpus.

In Georgian, the category of negation is conveyed through both verb and noun morphology. The particles used in verb morphology form a three-member system: sm (not), 30m (can't), and 5m (don't). According to scientific literature, their functional-semantic distribution is as follows: sm expresses categorical negation, 30m indicates the negation of possibility, and 5m denotes prohibition. However, the intensity of the semantic function of these negation particles can be modified by combining them with verbs in different screeves. For example, in the screeves of the third series of the tense-aspect-mood (TAM) system, the particle sm loses its categorical nature and conveys a neutral negation (Kurdadze et al., 2022, p. 208). In some TAMs, the particle 5m expresses a threat (usually in combination with the particle **sds** *aba*) or a wish; it is also used in curse formulas.(See Table 10)

Functional semantics of the negation particles become much more complicated when considering semantic groups of verbs or syntactic constructions which they can build:

a) The particle არ does not express categorical negation in verbs that cannot combine with the particle 30 m. Cf.: არ მწყურია *I'm not thirsty*, არ შემიძლია *I can not*,

Table 10

	Categorical Negation	Negation of Possibility	Prohibition	Neutral Negation
I Series	არ	ვერ	ნუ	-
II Series	არ	ვერ	-	-
III Series	-	-	(ნუ)	არ

Distribution of negation particles in Georgian

არ მესმის *I don't hear* in contrast to *ვერ **მწყურია**, *ვერ **შემიძლია**, *ვერ **მესმის** (Djorbenadze 1984: 141).

b) In addition, the negation particle 30% is not used with statical verbs არ აწერია *it is not written on it*, არ ახატია *it is not painted on it* in contrast to *30% აწერია, *30% ახატია, and inversive verbs (verbs with a dative construction in the present tense): არ მშია *I'm not hungry*, არ მიყვარს *I don't love it* in contrast to *30% მშია, *30% მიყვარს (Chumburidze, 1970, p. 42), with potential: არ იჭმევა *not edible,* არ ისმევა *not drinkable* in contrast to *30% იჭმევა, *30% ისმევა (Machavariani, 2002, p. 100) and with verbs that express not having or lacking a property: არ გააჩნია/არ მოეპოვება *he/she/it does not possess/does not own* in contrast to *30% გააჩნია, *30% მოეპოვება (Chumburidze, 1970, pp. 42-43).

Table 11

Distribution of negation particles in Georgian by different verb types

Type of verb	Categorical Negation	Negation of Possibility	Neutral Negation
Inversive verbs	-	ვერ	არ
Statical verbs	-	ვერ	არ
Verbs with poten- tialis		ვერ არ	-
Verbs of existence	-	ვერ	არ

The distribution of negation particles gives an interesting picture in different grammatical moods, in particular, the particle 33m cannot be confirmed with imperative. It is usually used with indicative and conjunctive. The particle 53, on the contrary, is used with imperative and optative forms (Chumburidze, 1970, p. 42). (See Table 12)

The use of the particles ნუ and არ on a pragmatic level shows an interesting picture: these particles can convey identical functional content by combining with different TAM forms of the verb. For example, the negative verb form in the conjunctive II with ნუ particle - ნუ დაწერდა - has the same pragmatic content as the negative verb form in the perfect II with არ particle - არ დაეწერა.

Table 12

Distribution of negation particles in different moods

Mood	არ	ვერ	ნუ
Indicative	+	+	(+)
Imperative	+	=	+
Conjunctive	+	+	(-)

The semantic function of the particles არ and ვერ concerning the act of communication is also interesting: Although ვერსად(აც) ვერ წახვალ (*you cannot go anywhere*), formally should convey the negation of possibility: in a particular context it is used to convey the semantics of a categorical prohibition: არ წახვალ (*you will* <u>not go anywhere</u>).

Furthermore, this semantics can be seen more clearly in the idiomatic expression $\mathfrak{GDbbsg}\mathfrak{GDh}\mathfrak{Bmggmn}$, which, despite the presence of the particle \mathfrak{GDh} , expresses a clear prohibitive - "*You will not change your foot under any circumstances*" = I forbid you to move from the spot. So, the particle \mathfrak{GD} can express a categorical negation in the present tense if the action has already begun. In such a case, $\mathfrak{SM}\mathfrak{Bssg}\mathfrak{DOM}$ (NegPart \mathfrak{SM} +Imperative) and $\mathfrak{SD}\mathfrak{SDMO}\mathfrak{O}$ (NegPart \mathfrak{SM} +Present) convey the same thing functionally and semantically *don't do it*.

The three-part system of negative particles (sm ar, 33m ver, by nu) presented in the Georgian language corresponds to the two-part system in Magrelian: 3s(m) and by. The particle 3s(m) in Megrelian conveys both functions of sm and 33m particles in Georgian. 3s is not an independent element and is not written separately, it is attached to the verbal form and creates synthetic morphological forms of the negative verb. Writing the negation particle 3s together with the verb is also facilitated by the fact that it is included as an infix in verb forms which has a complex preverbs: **Qmon-3s-Qm-bmQy** dot-va-do-doxu (Qmonj-3s-Qm-bmQyb dote-va-do-doxun) he/she/it does not sit down "sm xQjbs" (Khubua, 1942, p. 744).

With the forms of potentialis, the 35 particle corresponds to the Georgian 30m particle in its function and conveys the negation of the possibility. The fact that Megrelian does not and cannot differentiate between 5m and 30m particles is compensated for in the verb form (Kiria et al., 2015, p. 623).

Cf.:

3ວ-ອ່ວທັງຄົບ *he/she/it does not write* (Pres., Act.) vs **3ງ-ງອ່ວທົງ<3ວ-ດອ່ວທ** *it cannot be written* (Fut., Pass.), **3ວ-ວອ່ວທົງ** *he/she/it cannot write* (Fut., Act.)

For the statistical analysis of Megrelian texts, the negation particle 5^o is irrelevant, as it is always written separately. Therefore, we will not discuss it here and will

instead return to the main issue: the problem that arises during tokenization due to the negation particle 35 being written with the verb.

As is known, functional elements stand out with the highest frequency in the statistical processing of data. Among them is the negation particle s6. The table shows the highest frequency words in the Georgian National Corpus diachronically:

Table 13

Old Georgian	Middle Georgian	Modern Georgian	GRC
Olu Georgiun	Midule Georgiun	Modern Georgian	ONC
და	და	და	და
იგი	ყოფნა	არ	ეს
ყოფ(ნ)ა	ის	ყოფნა	ის
რომელი	რა	ის	რომ
ის	არ	რომ	არ
არ	ეს	ეს	ყოფნა
რამეთუ	მისი	რა	რომელი
ყოველი	მე	მე	რა

Tokens with the highest frequency in comparison

As the statistical analysis of parallel texts of Georgian and Megrelian proverbs revealed, the Georgian negation particle 56 takes the second place in terms of frequency, and in the statistical analysis of Megrelian texts, the particle conveying the category of negation is not found separately at all (Jgharkava, 2024, p. 25). (See Table 14)

The same issue arises with the **Rustaveli corpus**: it is impossible to accurately measure the statistics of the negation particle **35** in the Megrelian translation of the epic. Unlike in Georgian, the Megrelian negation particle **35** forms a token only when combined with the verb, resulting in an inaccurate count from the perspective of statistical processing of negation.

We present the mentioned problem on the example of the stanza 3.90:

რა **პასუხი არა გასცა**, მონა გარე შემობრუნდა, როსტანს ჰკადრა: "შემიტყვია, იმას თქვენი **არა უნდა**; თვალნი მზეებრ გამირეტდეს, გული მეტად შემიძრწუნდა, **ვერ ვასმინე** საუბარი, მით დავყოვნე ხანი მუნ, და-".

Since he did not **answer**, the slave went back, He said to Rosten: "I understood that he **will** listen to **nothing** more from you; My eyes were dazzled as by the sun; my heart was sorely troubled.

Table 14

Comparison	of frequency	of the	Georgian	and Megrelian	proverbs
------------	--------------	--------	----------	---------------	----------

Word	Count	- ^
და	1	19
არ	1	17
ერთი		7
ვაცი		6
ვერ		4
აქვსო		3
დროზე		3
ერთხელ		3
რომ		3
უნდა		3
ღობეს		3
აკლდებოდესო		2
56		2
არც		2
გამოსაშვები		2
გამოუშვაო		2
გატყდება		2
გაჭირვება		2
ბინდ		2
ბული		2

I could not make him the conversation to hear, so I stayed there for a long time.

Translation by Kaka Jvania (3.92):

მონას მუთუნქ ვაგურთუნი : დირთ დო უწუ ხენწიფეს თაშ შევატყვი ი კოს თქვანი : მუთუნ ვაკო შხვას წუხენსટ გურქ შემეწუხ თოლქ მიდამირთ : ვაბხვალამუქ მა თენერს2 ოშ უწუენ ართ ვარჩქილე : შურო პასუხის ვერზენს2. When the slave could do nothing, he returned and said to the king. I understand it this way, he doesn't want anything from you, something else was worrying him. My heart was troubled, my eyes darkened: I've never experienced anything like it. If you've told him a hundred times, he doesn't even understand: he doesn't give any answer at all. Translation by Gedevan Shanava (3.90):

მონას **მუთუნქ ვაგშაღინუ** მუკირთ დო თეში ქმორთუა, ხენწფე თქვანი ის **ვარჩქილე** მონაქ ენა თაში თქუა, თოლქ ქამისკიდ თიშ ჯინაშა, გურქუ დახე წამირთუა, **ვაგმაგონუ** ნარაგადქ ადრეთ თიშენი **ვამმართუა**. The slave could not find out anything, turned out and came back.

King, he does not understand you - The slave said it like this.

I could not take my eyes off, my heart was almost broken.

I couldn't make him listen to what was said, and that's why I couldn't come back in time.

The statistics of the 10 most frequently occurring words in comparison are listed in the Table 14:

Table 15

Most frequently occurring words in comparison

Source text	Translation by Jvania	Translation by Shanava
და (25)	დო (23)	დო (23)
რად (9)	თეში (8)	მაფაქ (9)
რა (9)	ჩქიმი (7)	მუს (7)
არ (9)	მუში (7)	რდუ (6)
ესე (8)	ხენწიფექ (6)	მა (6)
მეფე (7)	მა (6)	ჩქიმ (5)
იყო (7)	ის (6)	უწუ (5)
იგი (7)	ვარ (6)	რე (5)
თუ (7)	გური (6)	მუთუნ (5)
ვერ (7)	აფუ (6)	კოჩი (5)

Cf.: Mapping the statistical processing of the chapter III in Voyant:

Figure 12

Georgian Text in Voyant

← → C tools.org/?corpus=39c23d8b78379	95311504434cd6de3075				\$	•	:
	😵 Voya	int Tools					?
⑦ Cirrus	Reader O Worttraub	ien	?	Market Trends III Doke	umentbegriffe		?
	ნახეს, უცხო მოყმე ვინმე ჯი საფი სენი სადავიიი ავი ლი საფი სენი სადავიიი ავი ლი საფი სენი სადავიიი ავი ლი საფი სენი სადავიი დავირთვილა, ერემლსა ვარეთ დავირთვილა, სა სენსა ვაზა ემოსა, გირეთნა ვუნის ბყავის აუბადე კირეთნა სელია სამკავი მათრანი პერძდ სატად მანია საუბრად ისი ქის წივიდა მოს საუბრად ისი ქის წივიდა მოს საუბრად ისი ქის საფია მარკა მიე მის ქის დავანამოკდებით მტირლისა ,	დი მტირალი წელისა ი მტირალი წელის არის, სა და ვითა გმიინა, სა და ვითა გმიინა, სა და გია გმიინა, გულსი ტალია განატი კულის ტალია, არქმელი თავისა, ა კული მდლიარისად, ი მკიზელი მოლიზღარის ი მკიზელი მოლიზღარის	об ностание истрие сертие истрие и и и и и и и и и и и и и и и и и и	00050 00045 00045 0005000000		a geodesic files	10 10 10
Begriffe •	< >	~ ?	Q ~ <	1	V ? Zurücksetzen	Darste	п. >
😰 Zusammenfassung 🔠 Dokumente 🖽 Phrasen	?	III Kontexte	Bubblelines	და (25) რად (9) ationen			?
Disco Konne oth Bit (Delement of income to A7, and in income	device the device Fortell	Dokument		ര്ം (9) Begriff	rechts		
vor 6 Minuten.	ainzigarugen vvoruormen, Erstellt	🗄 ნახეს, შავი	ინენი სადავით	3 ⁽⁹⁾ 3 ⁽¹⁾ 3	ვითა გმირსა, ხშირად ;	სხა მარგა	-

Figure 13 *Megrelian translation by Jvania in Voyant*

← → C 2: voyant-tools.org/?corpus=d9c90fbf1a87	534d56bef501b721a4a1						\$	🛛 :
	🎯 Voya	ant To	ols					?
Cirrus 🖽 Corpus Terms 🤸 Links ?	Reader O Worttraut	xen		?	Internet Trends	Dokumentt	pegriffe	?
	წყარპირს ბირეს დოხუნელ სადი წყარპირს ბირეს დოხუნელი უნ სადიო ვინობდნი თუმი ხპლას ბერუნა კალი ეპანწილი აღეე ვური კვილი იმგარდ იირნიუდა ვური კვილი ეპანდა დარზიუდა კიანი ხეს მათრასი უკებულ კალემ მან უკირუნა დავირუნა დაცს მიქ ძი მონა კუქეს კოკიობას მირე ფარ	ო უჩ ინებუ ალურ დლაგ მითინ თუმა ხოლო მხვა ო არუ თ მსაკი უ	ინებუ მიდგ. ინებუ მიდგ. რდ უნა ცხენი, 	არენი, ები. ,	0 0006- 0 0045- 0 0045- 0 0005- 0 00000- 0 0005- 0 0000000000	egg) • Aglada	• P30 • orgão • q	9 10 300gung6n, l
Begriffe (•)	$\leftrightarrow \rightarrow$		~ ?	8.4	(00)	×	? Zurücksetzen	Darstellu >
😰 Zusammenfassung III Dokumente IIII Phrasen	?		E Kontexte	Bubblelines	დო (23) თეში (8)	onen		?
			Dokument		B-jo∂o (7)	Begriff red	chts	
Dieses Korpus enthält 1 Dokument mit insgesamt 935 verschiedenen vor 11 Minuten.	/ einzigartigen Wortformen. Erstellt	Œ	Fyse30	ის თიცალი, მ	მუში (7)) (gen og	ენე ეკოკორობ მუში ბაღ	oo, .b
Westerbeitetighter 0.740		Đ	Fysri30	ღვა?? მონას მუთუჩ	bg6(gggd (6)	დო თე	ში ქმორთუა, ხემწფე თქ	ვანი ის
wortschatzgichte: 0.740		Đ	Fyuri30	მუთი ქორე, მეურთ	(0) cb	in the second	ბმირათ მინი ორე. მონევ	sjo
Readability Index: 15.890		Đ	Fyorido	მუთუნი ვემიორაგა	9 006 (6)	gen sy	ვემკიჯინ ვეცადუა. მონე	est bg
Durchschnittliche Wortanzahl pro Satz: 25.3		• 🕀	Fyoriao	განგარალო. ხენწფე	d amen (6)	in the	კუთხოზინუა, მათხოზი	in do cô
					00-01(0)			
<			Enviso.	malia R small month	ິ ນຫຼາ (6)	mm viv	a denter for the second of the	Rede

Figuer 14

Megrelian translation by Shanava in Voyant

	130410312000014021						
	🎯 Voyant	Tools					?
Cirrus 🖽 Corpus Terms 🔩 Links ?	Reader O Worttrauber	1	?	2 Trends	E Dokumentbeg	riffe	?
	კოჩ ქომირეს მიდგარდუნი : ინგარდ კოჩ ქომირეს მიდგარდუნი : წყირ. სახე ჯილოშა უცეფლ : ურ ინმა - ლფარც დო ინსხვრც : პარაკლ ოლაზერით შოლორ აფედ : კარაც შეს-ქენდე გადე ბავრა : პარაკ ა ფეს- გაკირულ : კარაც ურარუსერ სახუქაც : ინიტრ ქარირუსერ სახუქაც : ინიტრ	წყარ პის ხედ დ პის ხედ დო ინკარ ლურდუნი ტეფ რქვინუნდურ ეფ ხოლის გურან უფ ხალის გურანი ი დო თი ფერიში უშზვაშ ფულანდი In თიში	ო რდუნი მი	0.0056- 0.0045- 0.0045- 0.0005- 0.0005- 0.0005- 0.0005- 0.0005- 0.0005- 0.0005-	6 4 <u>6</u> 6 Eta		
8 8 S	ხემწიფექ თქუ მირე თინა : სქვამი მარს არი მათ მათას ა სამრა სამ	ჯგირი მეჯინელი - რ რ. ხ. რ.		0.0000	2 3 4 ente (კოჩ ქობირეს მი	5 6 7 დგარდუნი : წყარ	8 9 10 პის ხედ დო ინგა
Begriffe •	b)\$\$6,0923 0049 80.69 0050 : b)\$8,0920 (€) ⇒	ჯგირი მეჯინელი ∨ ?	• •	0.0000	2 3 4 ente (კო8 ქობირეს მი ~ ?	5 6 7 wgarh@rg8o : Fyarh Zurücksetzen	8 9 10 Job bgg gen offici Darstelli,
Begriffe • 5 *	byBioggi and a and a state of a s	2200fin BijiyoBijero V ?	Bubblelines	0.0000 1 Dokument Segm (0.0000 1 0,0000 1 0,000 1 0,0000 1 0,000 1 0,0000 1 0,0000 1 0,0000 1 0,0000 1 0,0000 1 0,0000 1	2 3 4 ente (¿n8 dindorigb 8 2 ?	5 6 7 დეარდუზი : წვარ Zurücksetzen	8 9 10 Job bgg gen office (): Darstelle,) ?
Begriffe •	byBineggi agg and and a signal agg	220040 8030050200	© Subblelines	0.0000 1 Dokument Segm (වි) වංදුය ව වංදුය ව ව ව ව ව ව ව ව ව ව ව ව ව ව ව ව ව ව ව	2 3 4 ente (¿m8 dmdordyb 8c ? onen Begriff recht	s 6 7 wppshargeba : Fysch Zurücksetzen	8 9 10 Job bgg gen ofge Darstell, > ?
Begriffe Begrif	byBioreggi ordga Borlig andia : Lidgadin (C.)	yganha 8yganbagena ✓ ? Kontexte Dokument ⊞ graß grad	Bubblelines	ი იიიი 1 Dokument Segn დო (23) მაფაქ (9) მუს (7) რფუ (6)	2 3 4 ente (ank dindonta) 8 ? Ponen Begriff recht gen obgat	s s z zurücksetzen s s cogeőn სახე ჯილი	s 9 10 Job bg@ @n ofgo Darstelle ? milo manual milo manual
Begriffe Begrif	byBioreggi ordg and a not si bygadn	yyyofio 8yyyofyero ? ? Markontexte Dokument B 3raß drod B 3raß drod B 3raß drod	 დ	0 0000 1 Dokument Segn (23) (3) (3) (3) (3) (3) (3) (3) (3) (3) (2 3 4 ente (კოశ ქობირეს მ: v ? onen Begriff recht დო ინგარ	5 6 7 ფიარდუნი : წყარ Zurücksetzen 5 ფუნი სახე ჯილი ყრსვ : მარგალიქ	 გ. 9 10 კის ხედ დო იმგ. ტ. Darstell. ? იმა უგუდ ეფ ირკინ
Begriffe • Begriffe • Bokumente Phrasen Dieses Korpus enthät 1 Dokument mit insgesamt 973 verschiedenen/ eir jetzt. Wortschatzdichte: 0.736	byBioeggi edgi antig anda : bigada	ჯებრი მეჯინელი	 დ Bubblelines ქობირეს მიდგარდე უგუდე : უჩა რაზი ა მანგარიში დუს გაკი 	0 0000 Dokument Segn (23) avgvd (9) avgvd (9) avgv	2 3 4 ente (ank dedorda de 2 3 4 ente (ank dedorda de 2 3 4 ente (ank dedorda de 2 3 4 ente (ank de 2 3 4 ente (ank de 2 3 4 ente (ank dedorda de 2 3 4 ente (ank de)	5 6 7 ფაკრდუნი : წყარ Zurücksetzen რელენი სახე ჯიილ ფრსც : მარვალიქ ურიში ტარ უკებე	8 9 10 30k bgg gen eðg: (): Darstellik ? miðu tjgttgg gg durhou
Begriffe Begrif	byBioeggl etgl abrig ands : bdsale	۲ ۲ ۲ ۲	۲ ۲ ۲ ۳ ۳ 8 ۳ 8 ۳ 7 ۳ 7 ۳ 7 ۳ 7 ۳ 7 ۳ 7 ۳ 7 ۳ 7	م مسلح المسلح المسلح والمسلح المسلح المسلح والمسلح المسلح ال والمسلح المسلح المسلح المسلح المسلح المسلح المسلح المسلح مسلح مسلح مسلح مسلح مسلح مسلح مسلح	2 3 4 ente (unk endorso de v ? onen Begriff recht Qen abat Qen abat Qen abat Qen abat	5 5 5 5 5 5 5 5 5 5 5 5 5 5	8 9 10 პის ხედ დო იშც: ② Darstelli. > ? ომა უგუდ ეფ იწკინ ედ მარია უდუ ი კოჩ
Begriffe Dieses Korpus enthält 1 Dokument mit insgesamt 973 verschiedenen/ eir jetzt. Wortschatzdichte: 0.736 Readability Index: 13.443 Durchschnittliche Wortanzahl pro Satz: 973.0	byBioeggl etgl abrig ands : bdgade	χχρήο δρχοδρισο	المالية <	تَعَادَ اللَّهِ اللَّهُ اللَّ	2 3 4 ente (svik jindorhyb ör ente (svik jindorhyb ör Ponen Begriff recht Qen röbjök Qen enböß Qen enböß Qen engö	5 5 7 φορ./ηφ.τβα : Εγμικ Ζυτιücksetzen 2 καρτβα Lubg χιορι σήθε : Βινήρυστιή μαθαιο ζωκ τιχορι στοβοίδου ματοςτής βαθαιο ματοςτής	8 9 10 30b byge gen oftge: Ø: Darstell, Ø: Darstell, Ø: gen oftgen, Ø: Ogen oftgen, Ø: Ogen oftgen, Ø: Ogen oftgen, Ø: Ogen oftgen,
Begriffe Begrif	bytkineggi otgg antig ands : bytgadn	ჯგირი მეჯინელი	Construction C	روم (23) محمد المحمد المحم المحمد المحمد الم محمد المحمد المحم المحمد المحمد ا	2 3 4 ente (svik jindorigi) ör ente (svik jindorigi) ör v ? onen Begriff rechti qen reblikt qen reblikt qen angi qen songi qen songi qen songi	χουνουτός - Εγωνό Ζυντοcksetzen χουνουτός - Εγωνό Ζυντοcksetzen χουτο μολο το το χροστ χουτός - Εγωνό χουνουτός - Εγωνός -	b 9 10 constell co

Comparison of word frequency in source and target text of the epic:

Figure 15

1	Rustaveli	Jvania	s translation	Shanavas translation		
WordList - VEPXU.TXT - ×		🔐 WordList - ສູລູລຣົດລ.t	xt – 🗆 X	WordList - گەڭەغە.txt — → × Copy ☐ Font Frequency - ↓ ٩		
		, - copy the role				
Word	Count 🗸 ^	Word	Count 🗸 ^	Word	Count ~ ^	
Qυ	902	gom	1051	QQM	810	
არ	784	მა	391	მა	450	
ვითა	237	რე	346	ĥjnðn	423	
30 ^m	214	მუ	262	თემი	290	
১চ	197	მარა	218	მუ	247	
არა	179	მუთ	216	նո	247	
ესე	155	6Jnð	205	მუჭო	242	
ვით	129	n5nEn	203	მუში	221	
306	129	მუჭო	186	გური	180	
ავთანდილ	117	ოკო	186	თემ	175	
രൗლറ	114	նո	175	თამი	173	
ვარ	106	თქუ	157	სქანი	162	
გოლსა	101	რენი	137	თიში	137	
ანუ	79	-9 6 -9	136	ათე	131	
ამბავი	78	თემი	132	ვართი	125	
3160	77	უკულ	132	აფუ	124	
ასრე	71	თიმ	129	თუ	121	
defe	69	ბური	119	ორე	121	
არცა	69	მუჭოთ	116	ართი	116	
amola	61	მუს	115	ვართ	113	
26000	58	ດາກ	113	თიმ	112	
20000	56	ასე	108	ართო	110	
0000	55	ხოლო	102	მოთ	108	
1900	50	646	101	3660	107	
and	49 4	1600	98 ~	3.6.	106 ¥	
Total Token: 48280	TTR: 0.3084921	Total Token: 42040	TTR: 0.355471	Total Token: 43330	TTR: 0.3396723	
Total Type: 14894		Total Type: 14944	X a	Total Type: 14718	N 0	
	X Cancel		× Cancel		× Cancel	

Comparison of word frequency in source and target text in KWIC

In these examples, it is also interesting that the particle **35** *va*- sometimes preceded by the indefinite pronoun **∂yoyb** *mutun* 'someone'. In combination with the negative verb (verb with the negative particle), the indefinite pronoun **∂yoyb** becomes a negative pronoun. The negative pronouns are present in Megrelian and Laz (**∂nos** *mita* 'nobody', **∂yos** *muta* 'nothing'), but they are less productive. It is also interesting to note that the verb in combination with the negative pronouns **∂nos** *mita* and **∂yos** *muta* is always formed in the positive form (e.g. **mita murtumu** 'no one came' and not **mitas vamurthumu**). From the point of view of these two different ways of conveying the negation, it will be interesting to check statistically which translator chooses which strategy. This requires an annotated corpus of Megrelian, so that a precise statistical processing of negative verbs would be possible despite the peculiarity of the negation category in Megrelian.

RESULTS

The multilingual parallel corpus *Rustaveli goes digital*, which currently contains 32 parallel translations of the full text of the epic in 20 languages (Georgian, Ger-

man, English, Spanish, French, Italian, Turkish, Azerbaijani, Kyrgyz, Russian, Belarusian, Ukrainian, Greek, Arabic, Persian, Armenian, Ossetian, Lithuanian, Mingrelian, Svan) is an important digital resource for translation studies. Although nowadays, there are many different ready-made tools that are successfully used in linguistics for statistical analysis, the data processing of texts can still be very in-accurate without considering the grammatical characteristics of the languages. As shown in the article, it is necessary to develop a suitable tool for each language to be able to carry out a cross-linguistic analysis in a parallel corpus such as *Rustaveli goes digital*.

To use the multilingual parallel corpus for multidisciplinary research, it is necessary to incorporate a two-level statistical data processing: at the low level, statistical processing of the text must consider the linguistic features because inaccurate statistical results can lead to wrong statistics, and thus to wrong conclusions. In this way, we will get the accurate statistical data obtained at the low level of the statistical data processing respective languages as a result, which we can compare at the second level with the statistical results of the other languages that were also statistically processed at the low level.

AbNC	Abkhazian National Corpus	Nom	nominative
Act.	active	Pass.	passive
DO	direct object	Pers.	person
GNC	Georgian National Corpus	Pl	plural
F	feminine	PossPron	possessive pronoun
Fut.	future	Pres.	present
KWIC	Key Word in Context	S	subject
М	masculine	Sg	singular
Ν	neuter	TAM	tempus, aspect, mood
NP	nominal phrase	Vtr.	transitive verb
Ν	noun	Vint.	intransitive verb

ABBREVIATIONS:

REFERENCES

Chkheidze, M., & Taktakishvili, L. (2016). Vefxistqaosnis gamocemata bibliografia: 1712–2015 [Bibliography of "The Knight in the Panther's Skin": 1712– 2015]. Tbilisi: Sezani.

Chumburidze, Z. (1970). Uarkhopiti nats'ilakebi kartulshi da mati khmarebis st'iluri taviseburebani [Negation particles in Georgian and stylistic features of their use]. School and Life, 2(17), 41-46.

Gippert, J. (2024). A zoological riddle from Medieval Georgia. In F. Mühlfried (Ed.), Languages and Cultures of the Caucasus: A Festschrift for Kevin Tuite (pp. 85–103). Wiesbaden: Reichert.

Jgharkava, G. (2024). Kartvelur enata andazebis kvleva ts'ifrul ep'ok'ashi – teoretiuli da teknologiuri charcho [Digital processing of proverbs in Kartvelian languages – theoretical and technological framework]. Millennium, 2, 5-43. https://doi.org/10.62235/mln.2.2024.7991

Koller, W. (1992). Einführung in die Übersetzungswissenschaft. Heidelberg.

Khubua, M. (1942). Uarkhopis nats'ilaki va megrulshi [The negation particle va in Megrelian]. Moambe (Journal of the Academy of Science of Georgia), 3(7), 743–745.

Kurdadze, R., Lomia, M., Margiani, K., & Tchumburidze, N. (2022). Uarkhopa kategorias kartvelur enebshi [The category of negation in the Kartvelian languages]. Tbilisi: Universali.

Machavariani, G. (2002). Kartvelur enata shedarebiti gramat'ika [The comparative grammar of the Kartvelian languages]. Kartvelologische Bibliothek, 9. Tbilisi: Publishing house of Tbilisi Staatliche University.

Sinz, J. (2017). Translationstheorien: Die Äquivalenz nach Werner Koller und die Adäquatheit in der Skopostheorie. Grin Verlag.

Tandaschwili, M. (2022). Dighitaluri rustvelologia [Digital Rustvelology]. Tbilisi: Iverioni.

Tandaschwili, M., & Kamarauli, M. (2023). Vepkhistqaosnis stiluri sakhelebis targmani (mts'eris p'irveli maghals) [Translating the stylistic devices of "The Knight in the Panther's Skin" (The case of the 'sun')]. Digital Kartvelology, 2, 82-105.

The Abkhaz National Corpus. (n.d.). *CLARINO Bergen Center*: https://clarino.uib.no/abnc/page