Yidian SHE
Université Paul Valéry – Montpellier,
Montpellier, France
sheyidian@gmail.com
https://doi.org/10.52340/lac.2023.08.98

Svan dialectometry: applying Gabmap to Svan Phonology and Inflectional Morphology

1. Introduction

The Svan language is spoken by about 14 000 speakers in Upper and Lower Svaneti and is considered as an outlier of the Kartvelian language family (*cf.* Ethnologue 2022, Tuite 2020). Based on qualitative isoglossic parameters, i.e., *Umlaut*, *vowel length* and *reduction*, A. Shanidze (1925/1981, as cited in Tuite 2020) distinguishes four main dialects within the Svan language, that is, *Balszemouri*, *Balskvemouri*, *Lentekhuri* and *Lashkhuri*.

Table 1.1 Shanidze's classical taxonomy of Svan dialects

Dialect	*Vowel length*	*Umlaut*	*Reduction*
Balszemouri	+	+	+
Balskvemouri	-	+	+
Lentekhuri	-	+	-
Lashkhuri	+	-	+

Source: Shanidze (1925/1981, as cited in Tuite 2020, modified)

Among these four dialects, Balszemouri and Balskvemouri are spoken in Upper Svaneti, and Lentekhuri and Lashkhuri in Lower Svaneti. Some specialists claim the existence of a fifth dialect, namely *Choluruli*, spoken in the commune of Choluri (*cf.* Lip'art'eliani 1994, see also Chantladze 2018). However, Topuria (2008: 14) rather considers Choluruli as a subdialect of Lashkhuri.

Since the classical taxonomy is based on qualitative isoglossic parameters, the current study proposes a dialectometric approach to Svan dialects, applying Gabmap (i.e., Levenshtein Distance) to Svan phonology and the inflectional morphology of Svan verbs. Based on the quantitative analyses of phonological and morphological features, I try to provide an alternative and additional point of view and to highlight the details of the internal imbrication of the classification branches and sub-branches.

2. Database

The quantitative analyses of this study were performed on a micro-database obtained from the author's fieldworks in Upper and Lower Svaneti in April – June 2022 and in November 2022 – January 2023. As shown in Table 2.1, the database contains items of 35 localities (See also Figure 2.1).

Table 2.1 Surveyed dialect varieties

Regions	Villages and settlements
Upper Svaneti: 15 localities	Ušguli [USHG], Ḳala [KALA], Ipari [IPAR], Ċvirmi [CVRM], Mesṭia [MESTIA], Laṭali [LATAL];

	Uṣxvanari in Bečo [BECHO.U], Cxeḳvani in Bečo [BECHO.C], Cxumari [CXMR], Eceri [ECER], Pari [PARI], Nodaši in Laxamula [NDSH], Naḳra [NAKR], Čuberi [CHBR], Xaiši [XAISH];
Lower Svaneti: 20 localities	Ɣobi [GHOB], Mele [MELE], Čixareši [CHXRSH], Lemzagori [LMZGR], Žaxunderi [ZHXDR], Sasaši [SASA];
	Čvelpi [CHVLP], Panaga [PANAG], Leusheri [LSHR], Teḳali [TEKL], Čvelieri [CHLR], Saqdari [SAQDR];
	Babili [BABIL], Qarishi [QARSH], Çanashi [CANASH], Xeledi [XELED], Gvimbrala [GVIMB], Qvedreshi [QVEDR], Rcxmeluri [RCXML]

The abbreviations of surveyed dialect varieties, which will be used in the remainder of the report, are given in Table 2.1, within square brackets "[ ]".

Figure 2.1 Map of surveyed dialect varieties

Table 2.2 An extract from the phonological database

	'high'	'frozen'	'hour'	'avalanche'	'far'	'ear'	'sun'	'herd'
MESTIA	ḳəltxi	ləkwrame	s̈ät	žäh	ǯwēdias	šdim	miž	ǯweg
USHG	ḳəltxi / ḳōltxi	ləkwrame	sät	žäh	ǯwēdias	šdim	miž	ǯweg
LATAL	ḳəltxi	luḳwreme	s̈ät	žäh	ǯēdias	šdim	miž	ǯeg
KALA	ḳəltxi	luḳwrame	sät	žäh	ǯwēdias	šdim	miž	ǯweg
IPAR	ḳəltxi	luḳwrame	sät	žäh	ǯwēdias	šdim	miž	ǯeg
CVRM	ḳəltxi	luḳwrame	sät	žäh	ǯwēdias	šdim	miž	ǯweg
BECHO.U	ḳəltxi	luḳwrome	s̈ät	žäx	ǯodias	šdim	miž	ǯweg
BECHO.C	ḳəltxi	luḳwrome	s̈ät	žäh	ǯodias	šdim	miž	ǯweg
ECER	ḳəltxi	luḳwreme	saät	žäh	ǯodias	šdim	miž	ǯeg
PARI	ḳəltxi	luḳwreme	saɣät	žäh	ǯwedias	šdim	miž	ǯeg
NAKR	ḳəltxi	luḳwreme	s̈ät	žäh	ǯwedias	šdim	miž	ǯeg
NDSH	ḳəltxi	luḳwreme	s̈ät	žäh	ǯodias	šdim	miž	ǯeg
CXMR	ḳəltxi	luḳwreme	s̈ät	žäh	ǯodias	šdim	miž	ǯweg
CHBR	ḳəltxi	luḳwreme	saɣät	žäh	ǯwedias	šdim	miž	ǯeg
XAISH	ḳəwtxi	luḳwreme	saɣät	žäh	ǯedias	šdim	miž	ǯeg
MELE	ḳōltxi	luḳwreme	sät	žay	ǯōdias	ṣtim	məž	ǯog
GHOB	ḳōltxi	luḳwreme	sät	žay	-	ṣtim	məž	ǯog
CHXRSH	ḳōltxi	luḳwreme	sät	žay	ǯōdias	ṣtim	məž	ǯog
LMZGR	ḳōltxi	luḳwreme	sät	žay	ǯōdias	ṣtim	məž	ǯog
ZHXDR	ḳōltxi	ləḳwreme	sät	žay	ǯōdias	ṣtim	məž	ǯog
SASA	ḳōltxi	luḳwreme	sät	žay	ǯōdias	ṣtim	məž	ǯog
PANAG	ḳəltxi	luḳwreme	sät	žay	ǯodias	ṣtim	məž	ǯweg
LSHR	ḳəltxi	luḳwreme	sät	žay	ǯodias	ṣtim	məž	ǯog

CHVLP	ḳəltxi	luḳwreme	sāt	žay	ǯodias	štim	məž	ǯog
TEKL	ḳōltxi	luḳwreme	sāt	žay	ǯōdias	štim	məž	ǯweg
CHLR	ḳəltxi / ḳlātxi	luḳwreme	sāt	žay	-	štim	məž	ǯweg
SAQDR	ḳlātxi	luḳwreme	sāt	žay	ǯodias	štim	miž / məž	ǯweg
XELED	ḳlätxi	luḳwäreme	sāt	žä	ǯodias	štim	miž	ǯweg
CANASH	ḳlätxi	luḳwäreme	sāt / sat	žä	ǯodias	štim	miž	ǯweg
BABIL	ḳlätxi	luḳwäreme	sāt	žä	ǯodias	štim	miž	ǯweg
NANAR	ḳlätxi	luḳwäreme	sāt	žä	ǯodias	štim	miž	ǯweg
QARSH	ḳlätxi	luḳwäreme	sāt	žä	ǯodias	štim	miž	ǯweg
RCXML	ḳlätxi	luḳwäreme	-	žä	ǯodias	štim	miž	ǯweg
GVIMB	ḳlätxi	luḳwäreme	sāt	žä	ǯodias	štim	miž	ǯweg
QVEDR	ḳlätxi	luḳwäreme	sāt	žä	ǯodias	štim	miž	ǯweg

Table 2.3 An extract from the morphological database

	C1: 1st person, imperfect	C1: 1st person, present conjunctive	C1: 1st person, future inferential (Tuite 2020)	C2: 1st person, conditional	C4: 1st person, imper-fect	C4: 1st person, condi-tional
MESTIA	xwaxṭäwdäs	xwaxṭäwdēd	otxaṭwi / otxaṭwīne	xweīrōl	malṭən	melṭōl
USHG	xwaxṭäwdäs	xwaxṭäwdēd / xwaxṭäwde	otxaṭwi / otxaṭwīne	xweīrōl	malṭən	melṭōl
IPAR	xwaxṭäwdäs	xwaxṭäwde	otxaṭwi	-	malṭən	melṭōl
BECHO	xwaxṭäwidäs	xwaxṭäwide	otxaṭwine	xweirol	maläṭda	melṭol
PARI	xwaxṭäw	xwaxṭäwide	otxəṭäwisg	xweiriw	malṭən	melṭiw
NDSH	xwaxṭäw	xwaxṭäwide	otxəṭäwisg	xweirol	malṭən	melṭiw
ECER	xwaxṭäw	xwaxṭäwide	-	xweirol	malṭən	melṭiw
CHBR	xwaxṭäw	-	-	-	malṭən	melṭiw
MELE	xwaxṭawis	xwaxṭawde	otxaṭwi	xweirōl	malṭān	melṭōl
GHOB	-	-	otxaṭwi	xweirōl	malṭān	melṭōl
SASA	xwaxṭawis	xwaxṭawde	otxaṭwi	xwēirenōl	malṭān	melṭōl
PANAG	xwaxṭawis	xwaxṭawde	otxaṭwi	xweirōl	malṭān	melṭōl
LSHR	xwaxṭawis	xwaxṭawde	otxaṭwi	xweirōl	malṭān	melṭōl
CHVLP	xwaxṭawis	-	otxaṭwi	xweirōl	malṭān	melṭōl
TEKL	xwaxṭawis	xwaxṭawde	otxaṭwi	xweirōl	malṭān	melṭōl
SAQDR	xwaxaṭäwis / xwaxṭawis	xwaxaṭäwde / xwaxṭawde	otxaṭwi	xweirōl	malaṭān / malṭān	melaṭōl
XELED	-	xwaxaṭawde	atwaxaṭäwi	xweirol	malaṭən	melaṭol
QARSH	xwaxaṭäwdäs	xwaxaṭawde	atwaxaṭäwi	xweiroldäs	malaṭən	melaṭol
NANAR	xwaxaṭäwdäs	xwaxaṭäwde	atwaxaṭäwi	xweirol	malaṭən	melaṭol

The database is divided into two parts: phonological data and morphological data. The phonological database includes 58 cognates from 35 localities (see Table 2.2), while morphological database includes inflected forms of three verbs, *i.e.*, √XAṬAW (to paint), √IR (to be written) and √LAṬ (to love), from 19 localities throughout Svaneti (see Table 2.3).

3. Method

In the current study, I will apply dialectometry to the Svan diasystem with the help of *Gabmap* algorithms. The software *Gabmap* measures the distance of dialects by means of the Levenshtein algorithm. As shown in the following tables, Gabmap is able to calculate the distance

from the cognates of different dialects according to additions, absences (see Tables 3.1 b & c) or modifications (Table 3.1 c) (*cf.* Léonard 2022a, Prokić et al. 2008).

Table 3.1

a)

SAQDR — USHG

ʒ	w	e	g	
ʒ	w	e	g	
				0

b)

SAQDR LATAL

ʒ	w	e	g	
ʒ		c	g	
	l			1

c)

LATAL — MELE

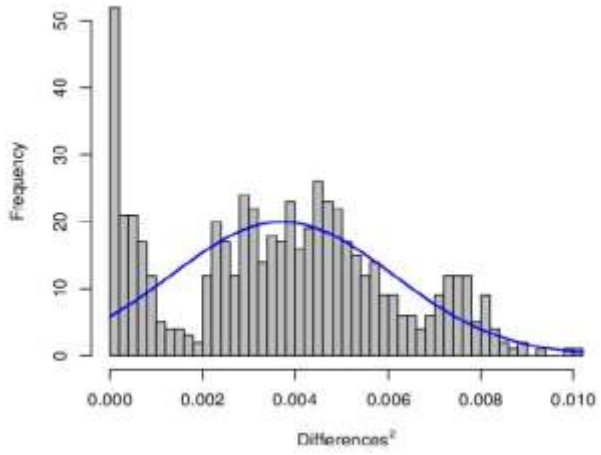
ʒ	e	g	
ʒ	o	g	
	l		1

Then, Gabmap proposes four clustering algorithms, *i.e.*, *Complete Link*, *Group Average*, *Weighted Average* and *Ward's Method*. According to Prokić et al. (2008), Ward's Method, also called the *minimal variance method*, generates in each step clusters "that result in the smallest increase in the sum of the squared distances of each individual from the mean of its cluster" (Prokić et al. 2008: 4). In the following section, I will primarily use Ward's Method for dialect cluster analyses.

4. Results

4.1 Phonological results

Figure 4.1.1 Statistics of phonological data (String Edit Distance – Tokenized)

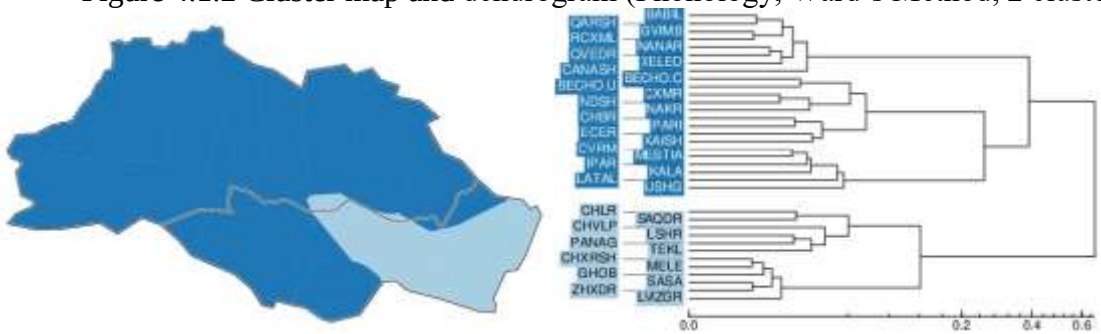


Local incoherence: 0.47; Cronbach's alpha: 0.91

Regarding the statistics of the data, despite some skewness, it is still close to a Gaussian curve, as shown in Figure 4.1.1. And Cronbach's alpha index is reassuring (> 0.7).

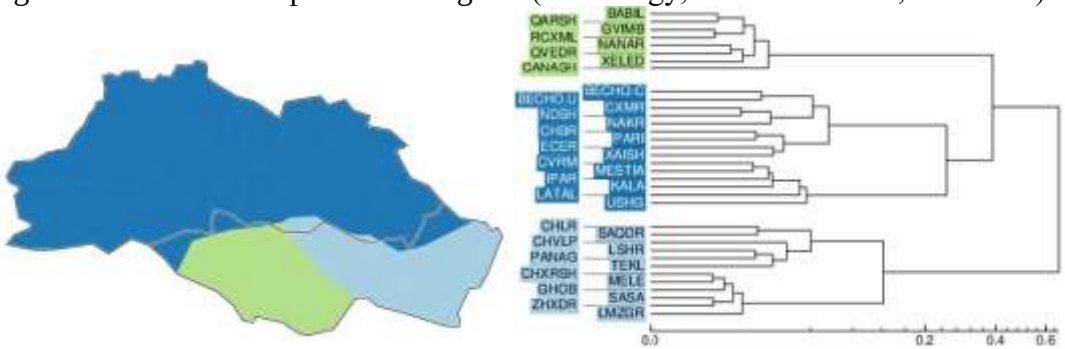
Due to space limitations, I will focus on the results generated by Ward's Method, which gives dialect clusters most often congruent with the isoglottic classifications. Let us start with the two-cluster configuration (See Figure 4.1.2).

Figure 4.1.2 Cluster map and dendrogram (Phonology, Ward's Method, 2 clusters)



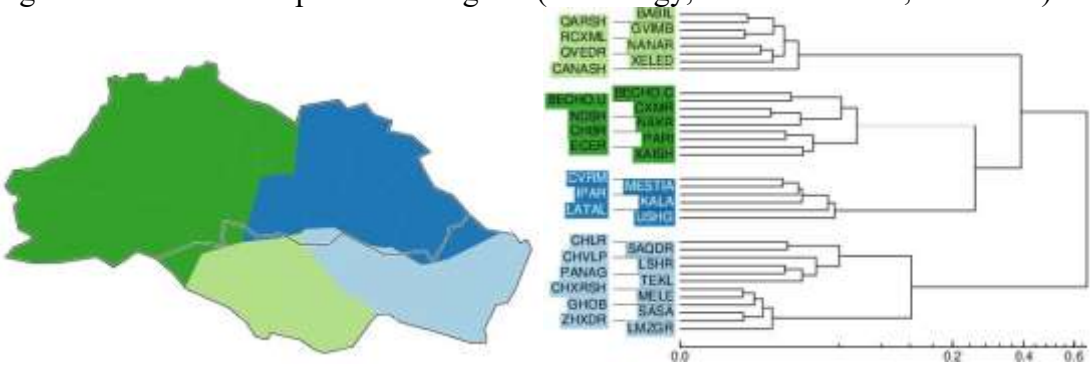
This configuration does not correspond, a priori, to two major dialects, *i.e.*, "Upper Svan" and "Lower Svan", but rather to a complex dialectal network, opposing two main areas. The internal structure of the dendrogram already suggests that there would be three dialects in the *major area* (dark blue). Then, a *minor area* (light blue) is located in the southeast, which covers the communes of Choluri and Lashkheti.

Figure 4.1.3 Cluster map and dendrogram (Phonology, Ward's Method, 3 clusters)



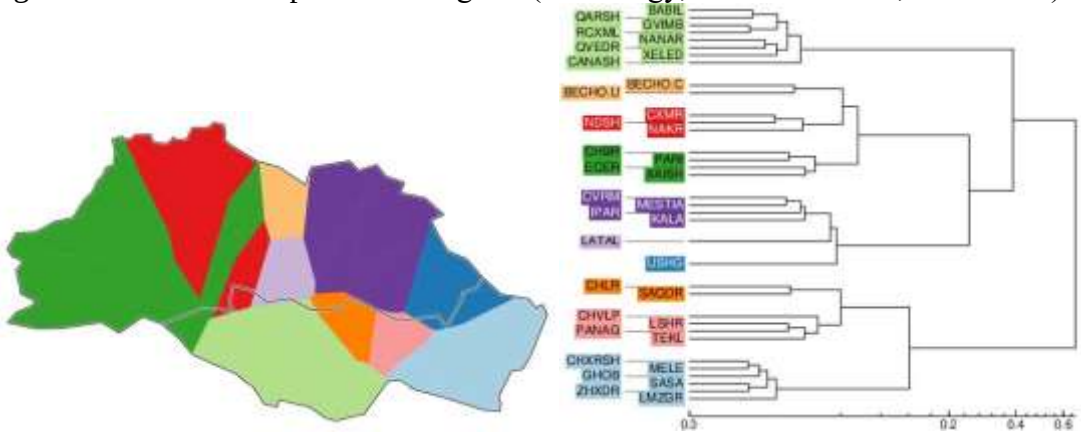
Then in the three-cluster configuration (See Figure 4.1.3), a southwestern choreme (light green, corresponding in particular to Lentekhuri), emerges from the major area. Indeed, this southwestern choreme is classified as an outlying member of the major area.

Figure 4.1.4 Cluster map and dendrogram (Phonology, Ward's Method, 4 clusters)



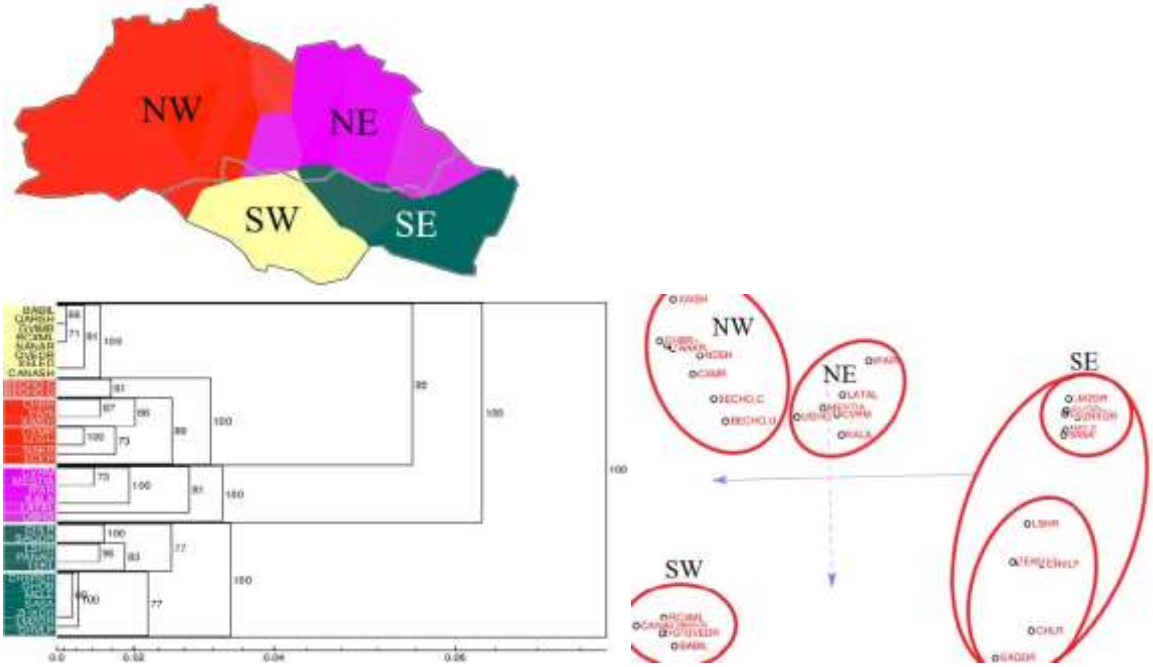
Then, the well-known quadripartition of Svan dialects appears in the four-cluster configuration, as illustrated in Figure 4.1.4. The internal structure of the northern part of the major area is further clarified: on the one hand, the northwestern choreme (dark green, corresponding to Balskvemouri), with no less than three sub-dialects; on the other hand, the *feature pool* (*cf.* Léonard 2022b, Gudschinsky 1958) of the northeastern choreme around MESTIA (dark blue, corresponding to Balszemouri). At this level of granularity, the internal structure of the minor area (light blue) remains unchanged.

Figure 4.1.5 Cluster map and dendrogram (Phonology, Ward's Method, 10 clusters)



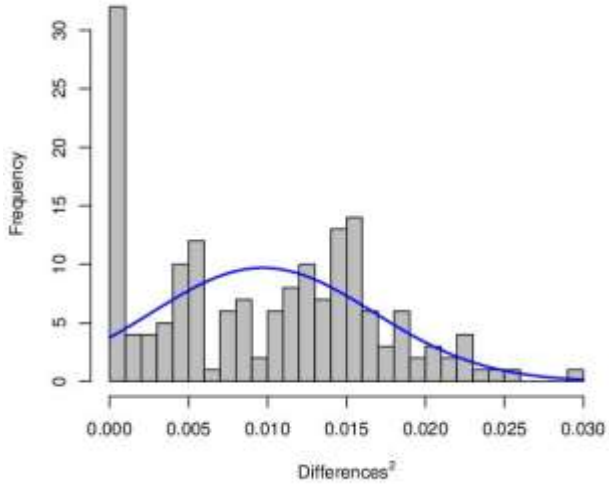
Finally, the 10-cluster configuration clearly shows the sub-dialects within each main dialect. In Upper Svaneti, the northwestern dialect (Balskvemouri) appears as an intricate group. BECHO is the structuring singleton of the core of this dialect. Similarly, it can be seen that USHG is an outlier of northestern dialect (Balszemouri). Then, LATAL is also a sub-dialect of Balszemouri which contrasts with the feature pool around MESTIA. In Lower Svaneti, the southwestern dialect (Lentekhuri) remains an unsegmented whole. However, the minor area, or southwestern dialect, is subdivided into three choremes: the light blue choreme consists of a strongly compacted group and corresponds to Lashkhuri, whereas the other two choremes, corresponding to the two sub-varieties of Choluruli (*Kveda Choluri* vs *Zeda Choluri*, *cf.* Shanidze 1925/1981), are less homogeneous. Furthermore, with respect to the probability method, the dendrogram in Figure 4.1.7 illustrates the certainty of each of the clusters. Both the dendrogram and the graph show the high certainty of a four-*modality* model (*Modalidades*, as it is called by don Ricardo, *cf.* Carballo Calero 1976, Dubert-García 2021). The MDS plots, as shown in Figure 4.1.8, further confirm the four-modality model.

Figure 4.1.7 Probabilistic graph and dendrogram and MDS plots (Phonology)



4.2 Morphological results

Figure 4.2.1 Statistics (String Edit Distance – Tokenized)



Local incoherence: 0.46; Cronbach's alpha: 0.94

The statistics of the morphological data confirms less to the Gaussian curve (see Figure 4.2.1). But Cronbach's alpha index is reassuring (> 0.7).

Table 4.2.1 Cluster map and dendrogram (Morphology, Ward's Method)

	Map	Dendrogram

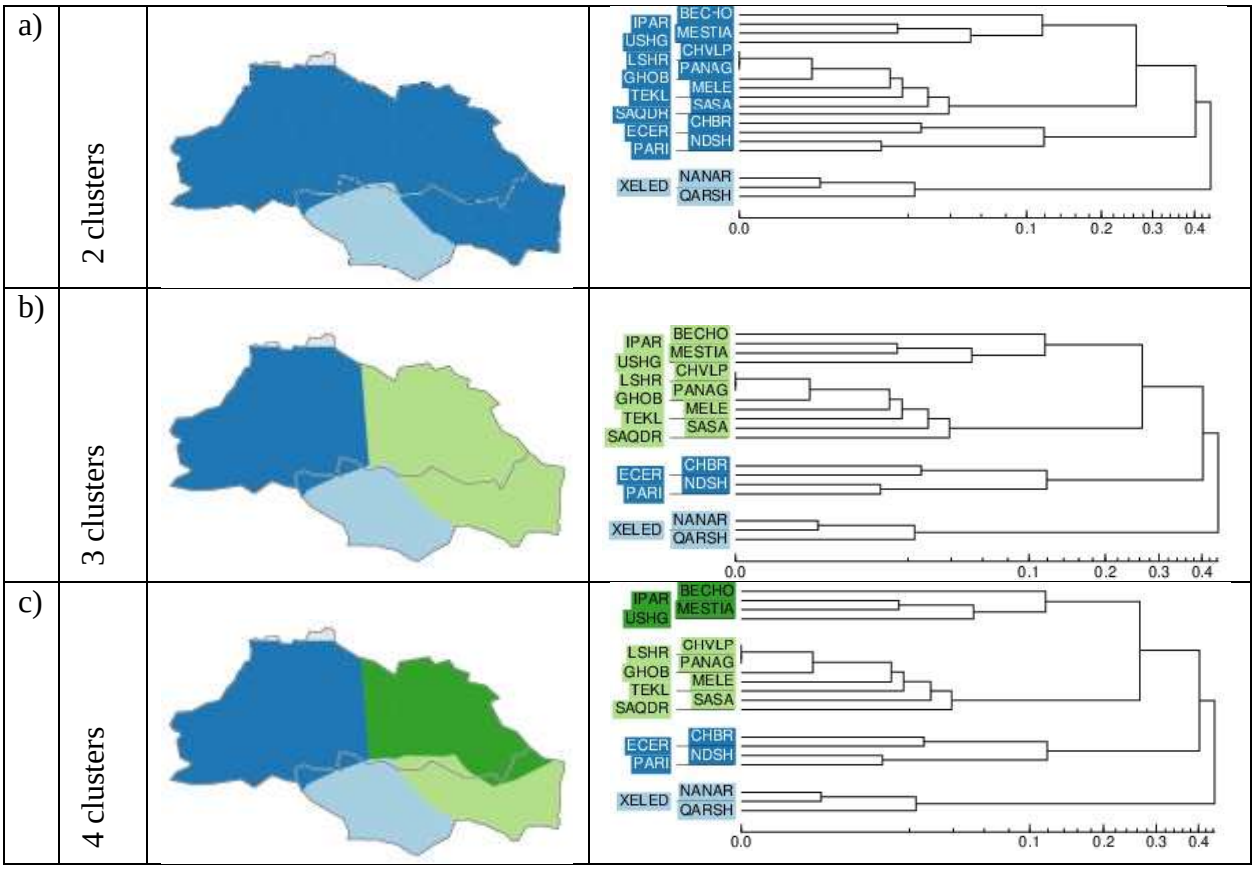
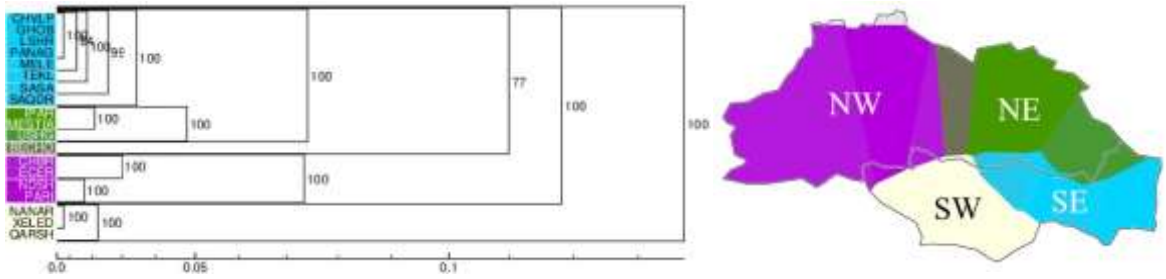


Table 4.2.1 below shows the two-cluster, three-cluster and four-cluster results in Ward's Method. Firstly, in the two-cluster configuration, it reveals once again a complex dialectal network that opposes two main areas. This time, however, it is the southwest – instead of the southeast – that opposes the rest of Svaneti (See Table 4.2.1 a). Thus, a minor area in the southwest, which corresponds to Lentekhuri, is contrasted with a major area that covers the north and southeast. This could be explained by the fact that Lentekhuri verb stems were less affected by atonic syncope than their cognates in other varieties (*cf.* Tuite 2020, Shanidze 1925/1981). Then, in the three-cluster configuration, a northwestern choreme, which corresponds roughly to Balskvemouri, emerges from the major area (See Table 4.2.1 b). According to Tuite (2020), Balskvemouri has a series of quite specific verbal exponents that distinguish it from other varieties (*e.g.*, conditional marker: Balskvemouri *iw* vs other varieties *ōl/ol*).

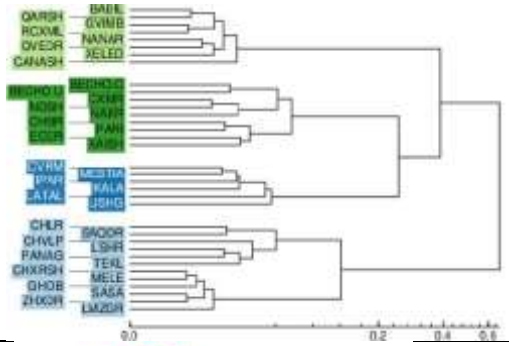
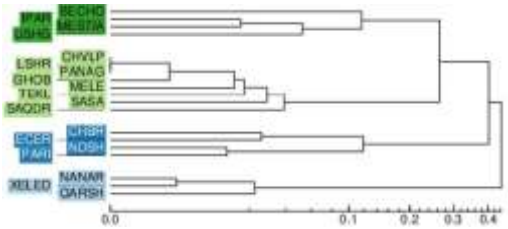
Finally, the four-cluster configuration shows again the quadripartition of the Svan dialects. Similarly, the dendrogram and the probability map propose the same pattern in four modalities (See Table 4.2.1 c and Figure 4.2.2). But it is noteworthy that in the north, BECHO is clustered in the northeast modality, instead of the northwest as seen in the phonological data results, which confirms the buffer zone character of BECHO.

Figure 4.2.2 Probabilistic dendrogram (Inflectional Morphology)



5. Conclusion

Table 5.1 Quadripartition of Svan dialects

	Phonology	**Inflectional Morphology**
Ward's Method – 4 clusters		
		
Probabilistic clusters		

In the current study, both phonological and morphological results highlight the robust quadripartition of Svan dialects. The four-modality model, to which geographic criteria (NW, NE, SW, SE) can be applied for convenience, is consistent with the classical taxonomy accepted by most scholars. Nevertheless, although this quadripartition is retained, the results of the quantitative analyses in this study suggest different groupings (see Table 5.1). According to the phonological results, there are two equipollent blocks, with very different dimensions: a major block, relatively

unitary but endemically intricate in structure, which combines the NW, NE and SW modalities; and a minor block, relatively homogeneous in structure. As for the morphological results, there are also two blocks. But this time it is the SW modality that is the outlier, which contrasts with the major block that brings together NW, NE and SE. Thus, if the term *Upper Svan* (ზემოსვანური) is still relevant, as in the phonological results NW and NE are united in one block, the term *Lower Svan* (ქვემოსვანური) turns out to be less linguistic than geographical, since the two modalities SW and SE exhibit significant linguistic distance.

As for the status of Choluruli, phonological results suggest that Lashkhuri and Choluruli consist of two sub-dialects of the modality (or dialect) SE.

The present study has provided an alternative viewpoint to the classification of Svan dialects, and especially to the details of the internal imbrication of branches and sub-branches, based on the quantitative analyses of phonological and morphological features, by means of the clustering algorithms in Gabmap. However, I did not consider the lexical data in this study. Since there are indeed a number of lexicons that distinguish the varieties of Upper Svaneti from those spoken in Lower Svaneti, further research should take into account lexical variation, in order to better apprehend the dynamics of Svan diasystem.

Acknowledgement: I would like to express my sincere thanks to my supervisor, Prof. Jean Léo Léonard, for his precious advice and his help throughout the analysis of the data of the current study and the writing of this report.

Bibliography

1. Carballo Calero, Ricardo. 1976. *Gramática elemental del gallego común*. Editorial Galaxia.
2. Chantladze, Iza. 2018. “Towards Qualification of the Kodorian Svan Speech” 12 (2).
3. Dubert-García, Francisco. 2021. “Ricardo Carvalho Calero, Dialectólogo.” Presented at the II Congresso Internacional em Variação Linguística nas Línguas Românicas, Universidade de Aveiro, Portugal. https://www.academia.edu/49547260/Ricardo_Carvalho_Calero_dialect%C3%B3logo.
4. Ethnologue 2022. *Ethnologue: Languages of the World, Twenty-fifth edition* Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). Dallas: SIL International. Accessed January 29, 2023. https://www.ethnologue.com/language/sva.
5. Gudschinsky, Sarah C. 1958. “Mazatec Dialect History: A Study in Miniature.” *Language* 34 (4): 469–81. https://doi.org/10.2307/410694.
6. Léonard, Jean Léo. 2022a. “Gabmap appliqué aux données de la morphologie d’ALTO - Atlas linguistique du Tseltal occidental.” Diaporama/Tutoriel, July 28.
7. Léonard, Jean Léo. 2022b. “Hints at Georgian Dialect History: A Study in Miniature.” In *Issues in Kartvelian Studies*, 3–29. Wilmington, Delaware: Vernon Press.
8. Lip’art’eliani, Aleksandre. 1994. სვანურ-ქართული ლექსიკონი *:* ჩოლურული კილო *[Dictionnaire Svane-Géorgien : Dialecte de Choluri]*. Tbilissi.
9. Prokic, Jelena, and John Nerbonne. n.d. “Recognizing Groups Among Dialects.”
10. Shanidze, Ak’ak’i. 1925. “უმლაუტი სვანურში [Umlaut in Svan].” In *Txzulebani*, 321–74. II. Tbilisi: Tbilisi State University Press, as cited in Tuite (2020).
11. Topuria, Varlam. 2008. სვანური ენის სახელმძღვანელო *[A Course of the Svan Language]*. Tbilisi: Ivane Javakhishvili Tbilisi State University.

12. Tuite, Kevin. 2020. “Svan Grammatical Sketch.” Université de Montréal.