Attempt at Setting Variables and Matrix Reflecting Morpho-syntactic Relations in Dialectometrical Analysis

– using negative particles in Romansh dialects as examples –

SEIMIYA Takamasa
Doctoral Program, TUFS
mail: s.takamasa1993@gmail.com

Flambeau vol.46 2020, p.119-144.
Manuscript received (2020-11-19) Manuscript accepted (2021-02-07)

Summary
In this paper, we investigated what kind of matrix should be used in the domain of dialectometrical analysis by comparing six dendrograms. We chose several maps of negative sentences from *AIS* as our data. We reached a conclusion that the dendrogram created with the variables and matrix of Pattern 3, in which each value is considered as a category, and with standardized data used, has best reflected the linguistic facts of target dialects.

Keywords: AIS, Romansh, negation, dialectometry, cluster analysis

0. Introduction

Several studies that perform dialectometrical analysis by digitizing word forms or phonetical forms written in language atlases exist. Goebl (1992), for example, classifies dialects using dummy variables[1]. Nevertheless, the criteria of correspondence were unclear, and no consideration was given on how similar the comparison point's word forms were to the those of the reference point. Yarimizu et al. (2004) and Kawaguchi (2007, 2020) investigate the process of standardization in the environs of Paris with a type of weighting method where the more the phonetic or morphology differs from the forms of Standard French, the larger the numerical value. Nerbonne et al. (1999) and Heeringa & Nerbonne (2001) study linguistic distances of Dutch dialects manipulating the Levenshtein distance, a string metric for measuring the difference between two sequences.

In the above-mentioned studies, they analyzed the linguistic distances of local dialects from the standard language or from one specific reference dialect. For clustering, they manipulated the matrices in which the sum of values obtained from the comparison of word forms were used. It means that the linguistic distances between the local dialects and the reference language/ dialect become larger in proportion to the increase in the total value. In other words, the closeness of the total value to 0 indicates that those local dialects preserved their indigenous language and were less influenced from either the adjacent dialects or the standard language. This method is useful when one analyzes the process of standardization or the linguistic distance from one specific reference dialect. However, it would be unsuitable to use this type of matrix when one analyzes the linguistic distances of dialects that have no standard language in common.

In addition to this, these studies have focused on the phonetical and morphological differences of dialects; to the best of my knowledge, there is no dialectometrical analysis that reflects the syntactical relations in the digitization. It would be worthwhile for dialectometrical analysts to attempt examining a suitable matrix and variables for the purpose of digitizing the dialects' morpho-syntactical relationships.

1. Romansh negation

Romansh, one of the Rhaeto-Romance languages (the others being Ladin and Friulian) spoken in the canton of Grisons in Switzerland, consists of five regional dialectal subgroups so called *idioms*: Sursilvan, Sutsilvan, Surmiran, Puter and Vallader[2]. The two eastern idioms, Puter and Vallader, are often referred to as *Ladin*[3]. Traditionally,

[1] Data that uses 0 where the word form of the reference point and that of the comparison point correspond and 1 where they do not.
[2] In English, they are called Surselvan, Sutselvan, Surmeiran, Puter and Vallader, respectively.
[3] One of the Rhaeto-Romance languages spoken in Northern Italy is also called Ladin, but what is

Sutsilvan and Surmiran are grouped together as *Central Romansh*, and they are grouped together with Sursilvan as *Rhenish Romansh*.

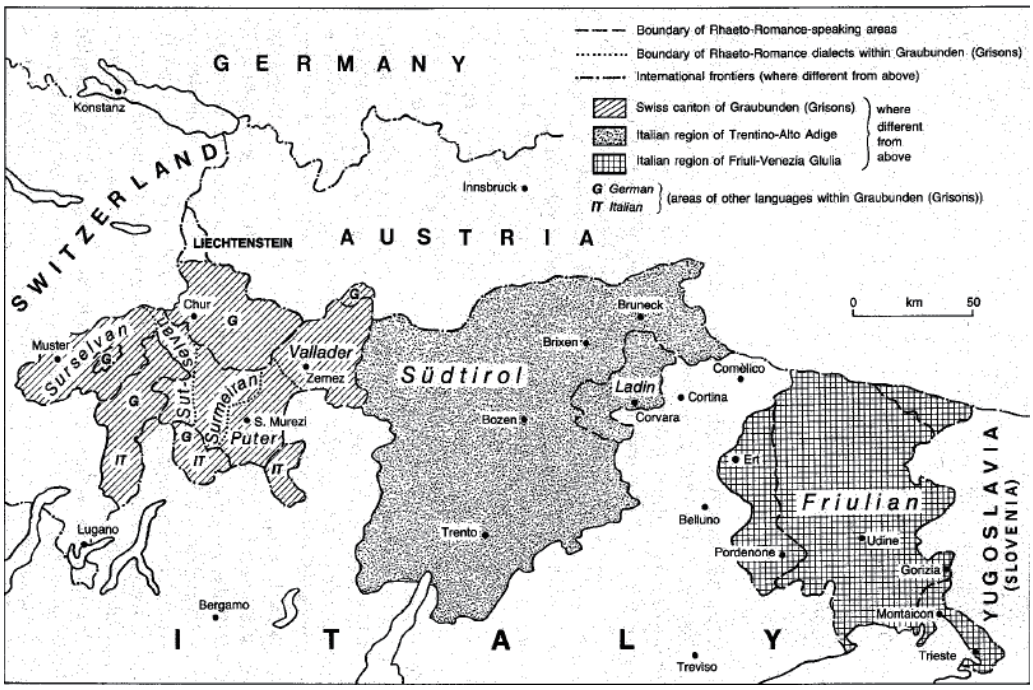

Figure 1. Distribution of the Rhaeto-Romance languages
(from Haiman & Benincà 1992: 2)

There are large regional differences among five idioms: the negative particles (NP) and their positions in negative sentences are no exception. In the two western varieties, they use the postverbal NPs *buca*/ *betg*, derived from *BICC-[4] [cf. (1) and (2)].

(1) Na, quei ei *buca* miu bab.
No, this be-IND.PRS.3SG NEG my father
No, this is not my father. (Menzli 1993: 23)

(2) Nus vagn *betga* liber suaintermiezgi.
We have-IND.PRS.1PL NEG free afternoon
We are not free in the afternoon. (Conforti & Cusimano 1997: 30)

In the two eastern idioms, the preverbal NP *nu*, derived from the Latin *nōn*, is utilized [cf. (3)]. In Sursilvan, it is also possible to use *na* to negate; however, in the modern

called Ladin in this article is the generic name for Puter and Vallader.

[4] The etymology of *buca* and *betg* is still controversial (cf. *DRG* vol. II p.507). In this research, we consider *BICC-, which is used as their possible etymology in two dictionaries (*Dicziunari Rumantsch Grischun* (DRG), *Lexicon Romontsch Cumparativ*), as their etymology. *BICC- is a phonetically reconstructed etymology whose meaning is unknown.

language, this usage is limited to poetry [cf. (4)].

(3) La stüva *nun* ais pitschna.
the room NEG be-IND.PRS.3SG small
The room is not small. (Scheitlin 1962: 27)

(4) El *na* vul.
He NEG want-IND.PRS.3SG
He does not want. (Cahannes 1924: 161)

(5) Igl frar *na* canta *betg.*
the brother NEG sing-IND.PRS.3SG NEG
The brother does not sing. (Thöni 1969: 22)

(6) Tü *nu* varast *bricha* temma?
You NEG have-IND.FUT.2SG NEG fear
Are you not going to be scared? (Liver 1991: 97)

In Surmiran, negative sentences are formed with two different NPs, *na* and *betg*, by sandwiching a verb similar to *ne…pas* in Standard French [cf. (5)]. In this idiom, there is a tendency to drop *na* and negate the sentence with only *betg.* In Vallader, just like with Surmiran, the compound negation *nu…bricha* can be used, yet this usage is less common [cf. (6)]. Table 1 summarizes the types and positions of negatives in the declarative sentences of each idiom.

Table 1. Types of negative particles and their position in the declarative of each idiom

IDIOM	NP and its position I	NP and its position II[5]
Sursilvan	V + buca	na + V
Sutsilvan	V + betga	
Surmiran	na + V + betg	V + betg
Puter	nu + V	
Vallader	nu + V	nu + V + bricha

2. Objective

In this research, “what kind of variable and matrix should be utilized in the hierarchical clustering analysis when using digitization that reflects the word form’s

[5] The use of *na* + V in Sursilvan and *nu* + V + *bricha* in Vallader is less common. Even though there is a tendency to use V + betg in Surmiran, from a viewpoint of prescriptive grammar, this usage is informal.

etymology, phonetic and syntax" is examined. In order to attain this objective, we compare and analyze six dendrograms generated with two types of data (raw data and standardized data) and three types of matrix patterns. No dialectometrical studies on the Romansh negation have been done. It can be said that this research is highly novel.

3. Targets

3.1. Area to be analyzed

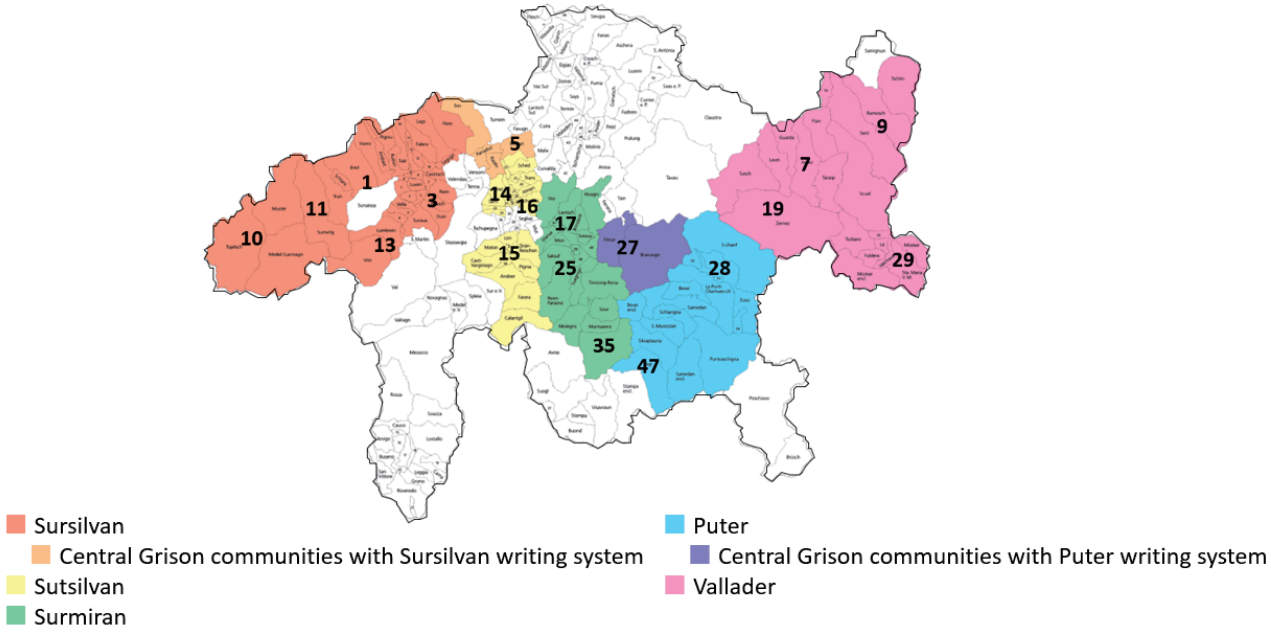

Figure 2. Target area and *AIS* points[6]
(created by the author based on *AIS* and Gross 2004: 27)

We used *Sprach- und Sachatlas Italiens und der Südschweiz* (=*AIS*, Linguistic and Ethnographic Atlas of Italy and Southern Switzerland) as a source material. There are only nineteen Romansh-speaking points in *AIS*. Therefore, we analyzed the word forms of these nineteen points. The survey in this area was conducted from 19th November 1919 to 22nd April 1920 by swiss linguist Paul Scheuermeier.

3.2. Maps to be investigated

We first selected seventeen maps that include negative sentences. For convenience, we converted the phonetic alphabet used in *AIS* to the International Phonetic Alphabet[7].

[6] The numbers on Figure 2 are those of *AIS*. The white areas on the map are Swiss German-speaking or Swiss Italian-speaking areas. For the numbering system in *AIS*, please refer to Jaberg & Jud (1928: 37-143).

[7] Some *AIS* phonetic symbols cannot be replaced by a single IPA symbol. For example, the word form of point 5 in Map 69 is "bę̣k". Regarding consonants, "b" and "k" correspond to [b] and [k], respectively, but the vowel "ę̣" is an intermediate sound between [e] and [ɪ]. For such sounds, we wrote one of them in parentheses – e.g. [be(ɪ)k]. Also, the word form of point 5 in Map 1615 is

After that, we excluded maps lacking word forms in the target area.

Table 2 shows the number and title of the maps, type of sentences – declarative (dec.), interrogative (int.) or imperative (imp.) - and their translation in English. The ellipsis in the title indicates that the whole sentence is separated and written in different maps[8]. Only the first forms in these maps were analyzed. Therefore, a total of 285 word forms (fifteen expressions × nineteen points) were analyzed, compared and digitized[9].

Table 2. Maps used

NUMBER	TITLE	TYPE	TRANSLATION
52	non vedi.. ?	int.	do not you (sg.) see.. ?
69	(perchè) non vi sposate ?	int.	why don't you (pl.) get married?
355	non vada..	imp.	Please do not go (sg.) ..
653	non dormirò..	dec.	I will not sleep..
1144	..non vadano nel giardino	dec.	..they do not go into the garden
1278	se non mangiamo..	dec.	if we do not eat..
1615CP[10]	non ha voglia di lavorare	dec.	he does not want to work
1621a[11]	non cadere	imp.	do not fall (sg.)
1621b	non cadete	imp.	do not fall (pl.)
1630	..non sarebbe contento	dec.	..he will not to be happy
1641	(mi rincresceva) che non la trovassiomo	dec.	(I was sorry) that we did not find her
1647	non ti muovere !	imp.	do not move (sg.)
1651	(mi meraviglio) che non lo troviate	dec.	(I am surprised) that you cannot find him
1658	non capisco ; capire	dec.	I do not understand ; to understand
1678	questa donna non mi piace	dec.	I do not like this woman

"bᵅgα". Sounds whose realization is ambiguous or weak are written in superscripts ("ᵅ" in this case). In order to express those sounds, we put a caret in front of it – e.g. [b^ɐgɐ].

[8] For example, the phrase "Bada che le galline non vadano nel giardino (Take care that the hens do not go into the garden.)" is separated into two different maps: Map 1143 "Bada che le galline" and Map 1144 "non vadano nel giardino".

[9] The word forms of each point on each map are shown in Appendix. The three colors used for the letters indicate that the position of the NPs placed in each expression is different: red – NP after a verb; blue – NP before a verb, and green – NPs before and after a verb. The NP in black letters indicates that NP is used, but in a different sentence structure from that of the map title.

[10] Expressions and words surveyed only in partial areas are listed as CPs (= complements) in the margins of the relevant maps. Map 1615 is a map of *lavorare ; lavora* "to work ; works".

[11] Note that since Map 1621 contains two different negative sentences, we utilized a total of fourteen maps with fifteen negative sentences. Regarding Map 1621, we name the first negative sentence as 1621a and the second 1621b in this study.

4. Methods and Procedures of the digitization of word forms and setting of matrices

4.1. Digitization of word forms

A speaker of Sursilvan, for example, would easily understand what another speaker of the same idiom speaks as they share the same language structures of Sursilvan, even if their pronunciation is different from one another. This speaker might understand what a speaker of Sutsilvan says, as the syntax of Sursilvan and Sutsilvan are quite similar, even if their pronunciation and vocabulary are slightly different from one another. This speaker, however, might hardly understand or must try to understand what a speaker of Vallader is saying as these two idioms are different in pronunciation, vocabulary, and syntax. These linguistic differences should be reflected on the dendrograms. To do so, based on Yarimizu et al. (2004), Kawaguchi (2007, 2020) and Seimiya (submitting), we set values from 0 to 10: the bigger the number, the larger the morpho-syntactic difference. In Kawaguchi (2007: 88), it is stated that "in determining the linguistic distance between geographical variants, an important distinction should be presupposed between morpho-phonological variants and lexical variants". If this statement is true, an important distinction should also be brought into the morpho-syntactical comparison. In order to emphasize the differences, we excluded values 6 and 9.

Table 3. Criteria for the digitization of word forms

VALUE	CRITERIA			EXAMPLE	
	etymology	phonetic	syntax	Point A	Point B
0	○	○	○	V + [buk]	V + [buk]
1	○	×	○	V + [buk]	V + [bec]
2	○	○	×	V + [buk]	[buk] + V
3	○	×	×	V + [buk]	[bec] + V
4	△	△	×	V + [bec]	[n] + V + [bec]
5	△	×	×	V + [buk]	[n] + V + [bec]
7	×	×	○	[buk] + V	[nu] + V
8	×	×	×	V + [buk]	[nu] + V
10	×	×	×	che S+V + [buk]	da + [buk] +INF

In Table 3, ○ indicates that the word form's etymology, phonetic and/or syntax of the points are identical, △ indicates that they partially correspond and × indicates that they are in disagreement.

When the etymology, phonetic and syntax of two points' word forms are identical (Point A: V + [buk]; Point B: V + [buk]), the value is 0. When the etymology and syntax are the same, but phonetically different (Point A: V + [buk]; Point B: V + [bec]), the value is 1. On the other hand, when the etymology and phonetic are the same, but

syntactically different (Point A: V + [buk]; Point B: [buk] + V), the value is 2. In addition, if they have the same etymology but they are phonetically and syntactically in disagreement (Point A: V + [buk]; Point B: [bec] + V), the value is 3.

When the word forms are syntactically different, but their etymologies and phonetics are partially identical (Point A: V + [bec]; Point B: [n] + V + [bec]), the value is 4. In addition, when their phonetics and syntax differ each other yet their etymology are partially identical (Point A: V + [buk]; Point B: [n] + V [bec]), the value is 5.

When they are syntactically the same, but different in etymology and phonetic (Point A: [buk] + V; Point B: [nu] + V), the value is 7. If they are etymologically, phonetically, and syntactically different (Point A: V + [buk]; Point B: [nu] + V), the value is 8. At last, although NPs are used in target and reference points, when the sentence structures differ significantly (Point A: che S + V + NEG; Point B: da + NEG + INF), the value is 10.

4.2. Comparison of word forms

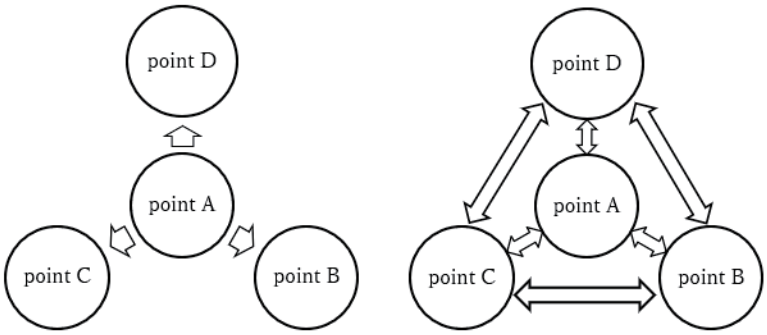

Figure 3. Simplified diagram of two different comparison types

In most of the studies described in the Introduction, the authors compared word forms of the standard language with those of target dialects, or word forms of a reference dialect with those of target dialects. In such cases, the variable used in the matrix for clustering represents how linguistically similar/different each target point is from the standard language or from the reference point. In other words, it is possible to say that only one-way comparison was performed [cf. Figure 3. Left]. However, in this research, we do not compare dialects with the standard language, but instead try to analyze how similar or dissimilar the target dialects are. Therefore, an alternating comparison was required instead of one-way comparison [cf. Figure 3. Right].

Where N is the number of points and n is the number of maps (or the number of expressions) utilized in the analysis, the comparison can be diagrammed as shown in Figure 4. The columns show the reference points. The rows show the target points that are to be compared with the reference points. For example, in MAP a, when point A is the reference point, the value of point B vs. point A is 1, and the value of point N vs. point A is 8. Similarly, in MAP b, when point A is the reference point, the value of point B vs. point A is 2, and the value of point N vs. point A is 5. Such a comparison is carried

out for all the word forms of all the points in all the maps. Since this is a round-robin format, the number of comparisons can be calculated by the formula: $N\,(N-1) \div 2 \times n$. In this study, as we covered word forms of nineteen points in fifteen expressions, in total we compared them 2,565 times.

	MAP a				MAP b				…	MAP n			
	point A	point B	…	point N	point A	point B	…	point N	…	point A	point B	…	point N
point A		1	…	8		2	…	5	…		10	…	10
point B	1		…	5	2		…	0	…	10		…	3
⋮	⋮	⋮		⋮	⋮	⋮		⋮	…	⋮	⋮		⋮
point N	8	5	…		5	0	…		…	10	3	…	

Figure 4. Simplified diagram of digitalization of word forms

4.3. Determining variables and creating three types of matrices

In this study, we used three different matrix patterns in order to analyze what kind of matrix would be the most realistic by comparing the dendrograms with the word forms in *AIS* maps. In this section, we explain the creation procedures of each matrix pattern.

4.3.1. Pattern 1

	MAP a				MAP b				…	MAP n			
	point A	point B	…	point N	point A	point B	…	point N	…	point A	point B	…	point N
point A		(1)	…	◇8		(2)	…	◇5	…		(10)	…	◇10
point B	1		…	[5]	2		…	[0]	…	10		…	[3]
⋮	⋮	⋮		⋮	⋮	⋮		⋮	…	⋮	⋮		⋮
point N	8	5	…		5	0	…		…	10	3	…	

	$\sum_{k=m}^{n} \mathrm{MAP}_k$			
	point A	point B	…	point N
point A	0	13+α	…	23+α
point B	13+α	0	…	8+α
⋮	⋮	⋮	0	⋮
point N	23+α	8+α	…	0

Figure 5. Example of the matrix of Pattern 1

Pattern 1 is a matrix whose variables are the sum of the values obtained by comparison. For example, when point A is the reference point, the value of point A vs. point B is 1 in MAP a, 2 in MAP b, and 10 in MAP n (numbers in ○). Hence the value 13+α (=1+2+…+10), are the variables for point A vs. point B. Similarly, the variable of point A vs. point N is 23+α (=8+5+…+10), and that of point B vs. point N is 8+α (=5+0+…+3). The closer the variable is to 0, the closer the linguistic distance between

the two points and *vice versa*. This pattern is a symmetric matrix. The size of the matrix of Pattern 1 can be calculated by $N \times N$. Hence, in this study, it was a 19 × 19 matrix.

4.3.2. Pattern 2

Pattern 2 is a matrix whose variables are the values digitized by comparison without any editing. The size of the matrix of Pattern 2 can be calculated by $N \times Nn$; therefore, it was a 19 × 285 matrix.

	MAP a				MAP b				⋯	MAP n			
	point A	point B	⋯	point N	point A	point B	⋯	point N	⋯	point A	point B	⋯	point N
point A	0	1	⋯	8	0	2	⋯	5	⋯	0	10	⋯	10
point B	1	0	⋯	5	2	0	⋯	0	⋯	10	0	⋯	3
⋮	⋮	⋮	0	⋮	⋮	⋮	0	⋮	⋯	⋮	⋮	0	⋮
point N	8	5	⋯	0	5	0	⋯	0	⋯	10	3	⋯	0

Figure 6. Example of the matrix of Pattern 2

4.3.3. Pattern 3

In Pattern 3, each value in Table 3 is considered a category. It shows examples of the number of word forms, which are categorized in categories, being used as variables. This matrix needs three steps as shown in Figure 7.

STEP 1: Enter the number of occurrences of each value in regard to comparisons of the reference point with the target point. Enter the number of expressions used when the reference point and the target point are the same point.
STEP 2: Subtract the number of expressions used from the numbers in each cell of the matrix.
STEP 3: Multiply by -1 in order to convert the negative numbers into positive numbers.

On the assumption that only three maps MAP a, MAP b, and MAP n are analyzed, we explain the three steps in detail. For example, when point A is the reference point, the values of point A vs. point N are 8 in MAP a, 5 in MAP b, and 10 in MAP n (numbers in ◇). Since these values 5, 8 and 10 occur once, we put 1 in rows 5, 8 10 of point A vs. point N. In row 0, when the reference point and the target point is the same point (point A vs. point A), we put the number of expressions used. Therefore, we entered 3 in row 0 of point A vs. point A and that of point N vs. point N [cf. Figure 7-STEP 1].

The matrices of Pattern 1 and Pattern 2 are dissimilarity matrices. That is, the closer the variables in the matrix are to 0, the higher the similarity. However, Pattern 3 in STEP 1 is a similarity matrix, which means that this is the exact opposite type of those of Pattern 1 and Pattern 2. In order to unify the types of matrix in three patterns, Pattern 3 needs to be converted to a dissimilarity matrix. Therefore, after STEP 1, we subtracted 3

from all the values in the matrix [cf. Figure 7-STEP 2] and multiplied them by -1 [cf. Figure 7-STEP 3]. The size of the matrix of Pattern 3 can be calculated by $N \times N \times$ number of values. We utilized nine values (0, 1, 2, 3, 4, 5, 7, 8 and 10); therefore, it was a 19 × 171 matrix.

	MAP a				MAP b				…	MAP n			
	point A	point B	…	point N	point A	point B	…	point N	…	point A	point B	…	point N
point A		(1)	…	◇8		(2)	…	◇5	…		(10)	…	◇10
point B	1		…	[5]	2		…	[0]	…	10		…	[3]
⋮	⋮	⋮		⋮	⋮	⋮		⋮	…	⋮	⋮		⋮
point N	8	5	…		5	0	…		…	10	3	…	

STEP1

	point A									…	point N								
	0	1	2	3	4	5	7	8	10	…	0	1	2	3	4	5	7	8	10
point A	3	0	0	0	0	0	0	0	0	…	0	0	0	0	0	1	0	1	1
point B	0	1	1	0	0	0	0	0	1	…	1	0	0	1	0	1	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
point N	0	0	0	0	0	1	0	1	1	…	3	0	0	0	0	0	0	0	0

STEP2

	point A									…	point N								
	0	1	2	3	4	5	7	8	10	…	0	1	2	3	4	5	7	8	10
point A	0	-3	-3	-3	-3	-3	-3	-3	-3	…	-3	-3	-3	-3	-3	-2	-3	-2	-2
point B	-3	-2	-2	-3	-3	-3	-3	-3	-2	…	-2	-3	-3	-2	-3	-2	-3	-3	-3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
point N	-3	-3	-3	-3	-3	-2	-3	-2	-2	…	0	-3	-3	-3	-3	-3	-3	-3	-3

STEP3

	point A									…	point N								
	0	1	2	3	4	5	7	8	10	…	0	1	2	3	4	5	7	8	10
point A	0	3	3	3	3	3	3	3	3	…	3	3	3	3	3	2	3	2	2
point B	3	2	2	3	3	3	3	3	2	…	2	3	3	2	3	2	3	3	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
point N	3	3	3	3	3	2	3	2	2	…	0	3	3	3	3	3	3	3	3

Figure 7. Example of the matrix of Pattern 3

5. Method for creating and analyzing clustering results

In this study, agglomerative hierarchical clustering was used. There exist seven types of standard linkage methods in this clustering: single linkage, complete linkage, average linkage, centroid linkage, weighted average linkage, median linkage, and Ward's method[12]. Among them, as Ward's method appears to perform well (Everitt 1979: 173, Everitt et al. 2011:28, Noguchi 2018: 268), we decided on this method. As a measurement, we applied the squared Euclidean distance, which is the most compatible with Ward's method. When using the Euclidean distance, "the classifications obtained using raw and

[12] Single linkage and complete linkage are also called *nearest-neighbor method* and *farthest-neighbor method*, respectively.

standard data are usually different (Adamson & Bawden 1981: 205)", but "it is not possible to say a priori which of those will be more desirable (*Ibid*: 208)." Therefore, two types of data – raw data and standard data – were used in the matrix of three patterns: a total of six dendrograms (three patterns × two types of data) were created. R commander (R version 3.6.3.) was used for clustering analysis, standardization of data and creation of dendrograms.

The six dendrograms described above were analyzed in the following steps. First, in order to judge whether the overall linguistic tendency was reflected in the clusters, we analyzed the components (= dialect points in this case) of large clusters within each dendrogram. After that, to find out whether the linguistic similarities of dialects were reflected in the clusters, we analyzed the clustering process of the components by comparing them with the word forms observed on each map.

6. Results and Discussion

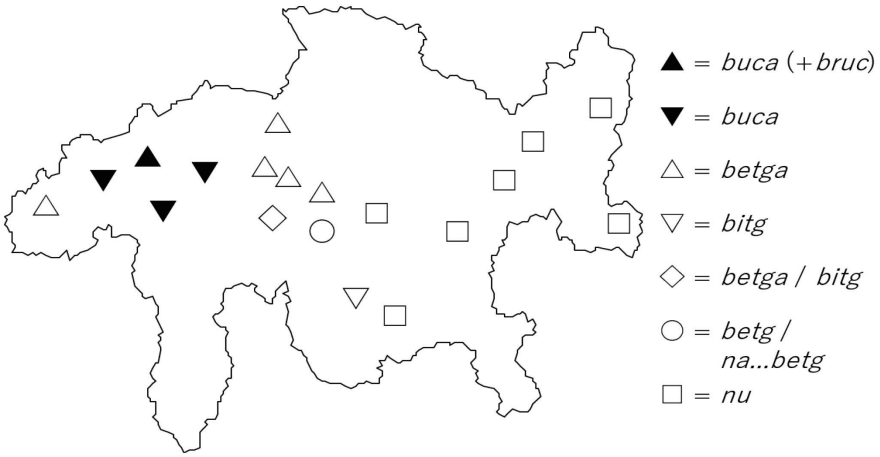

Figure 8. Summary of NPs observed in fifteen expressions

Figure 8 plots the tendency of NPs used in fifteen negative expressions at each *AIS* point. The legends, except for ▲, ◇ and ○, indicate that the word types of NPs were the same in the fifteen expressions [cf. Appendix for the word forms of each point]. ▲ is point 1, where an unidentified NP *bruc* was used in only one expression and *buca* was used in the other fourteen. ◇ is point 15, where *betga* and *bitg* were used in nine and six expressions, respectively. ○ is point 25, where two different types of negation were utilized. This distribution is roughly consistent with what has been said in section one: type *BICC- (*buca, betg, bitg*) in the west, type *nōn* + *BICC- (*na…betg*) in the central, and type *nōn* (*nu*) in the east. Even though it is possible to use *na* + V in Sursilvan and *na* + V + *bricha* in Vallader (Cahannes 1924: 161, Scheitlin 1962: 27), such negative expressions were never shown in both dialect areas. In Surmiran, the compound negation was attested only in point 25.

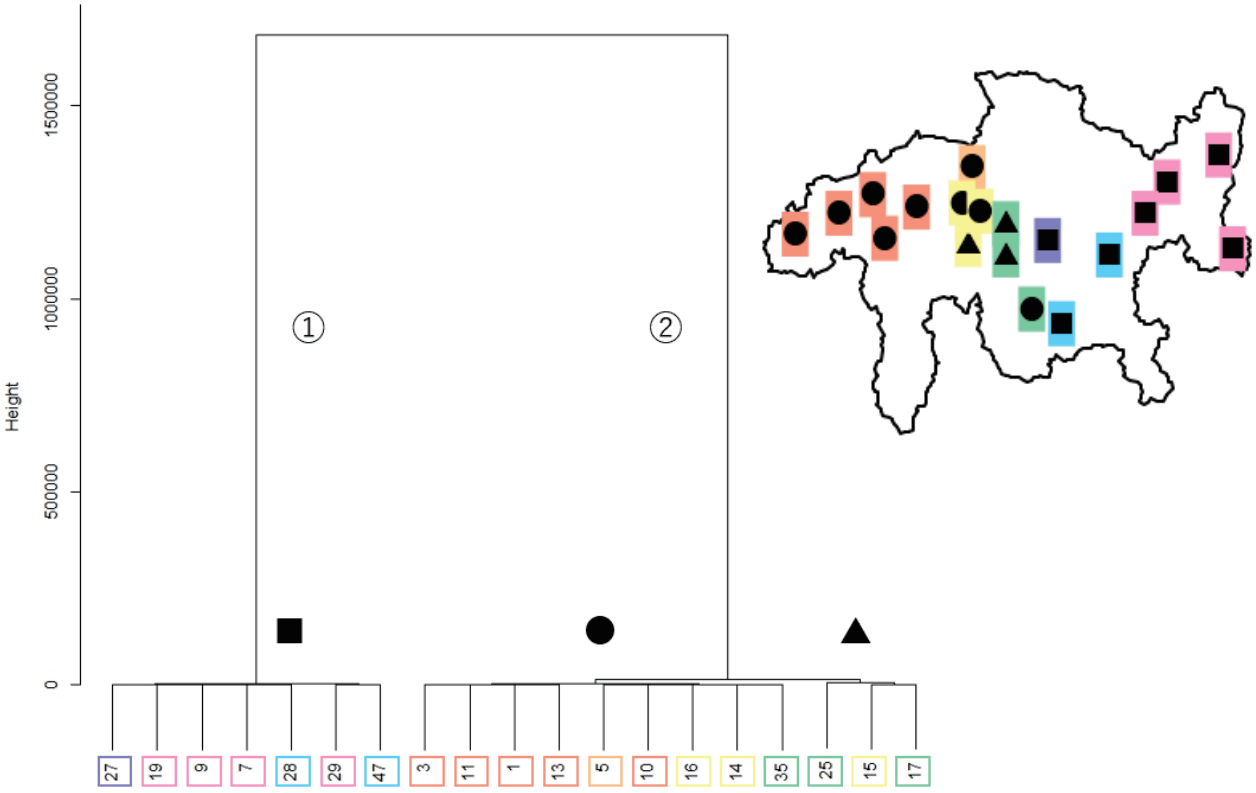

Figure 9. Dendrogram and map reflecting clustering result of Pattern 1-R

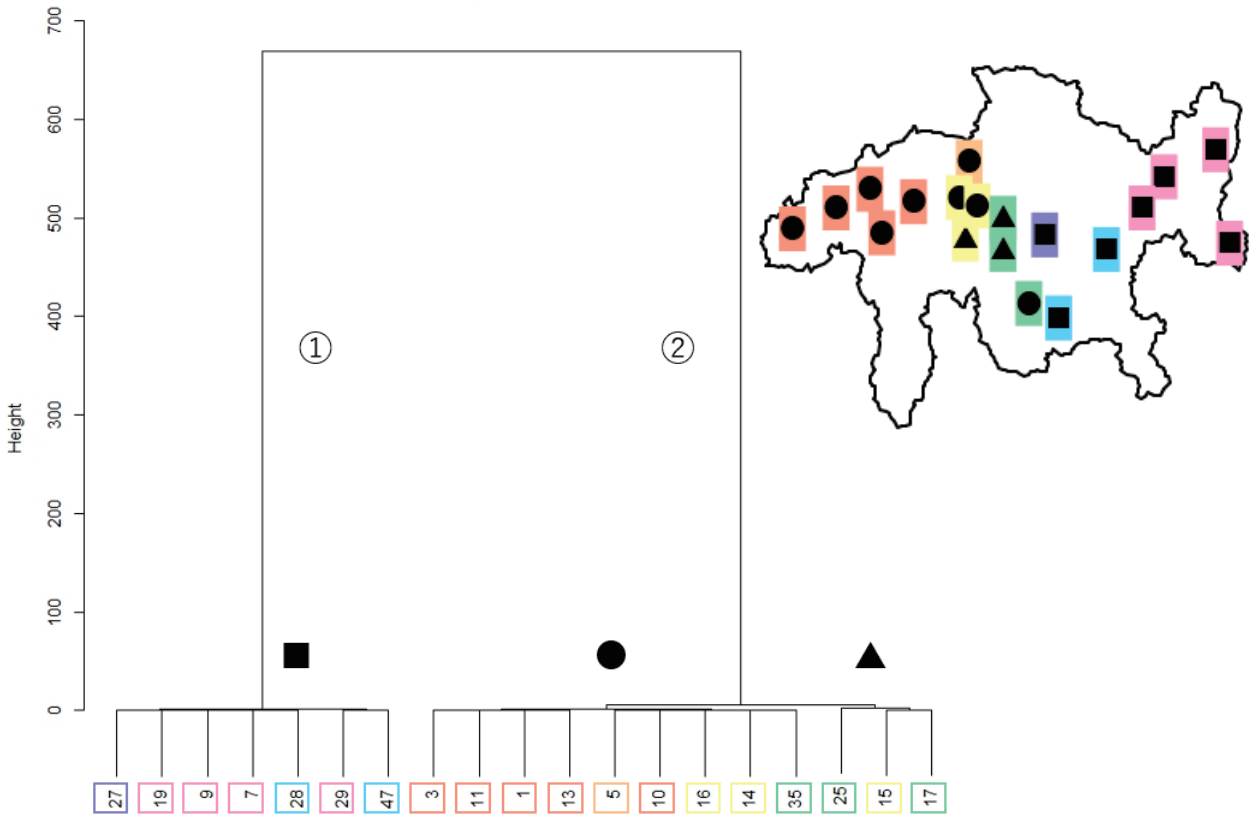

Figure 10. Dendrogram and map reflecting clustering result of Pattern 1-S

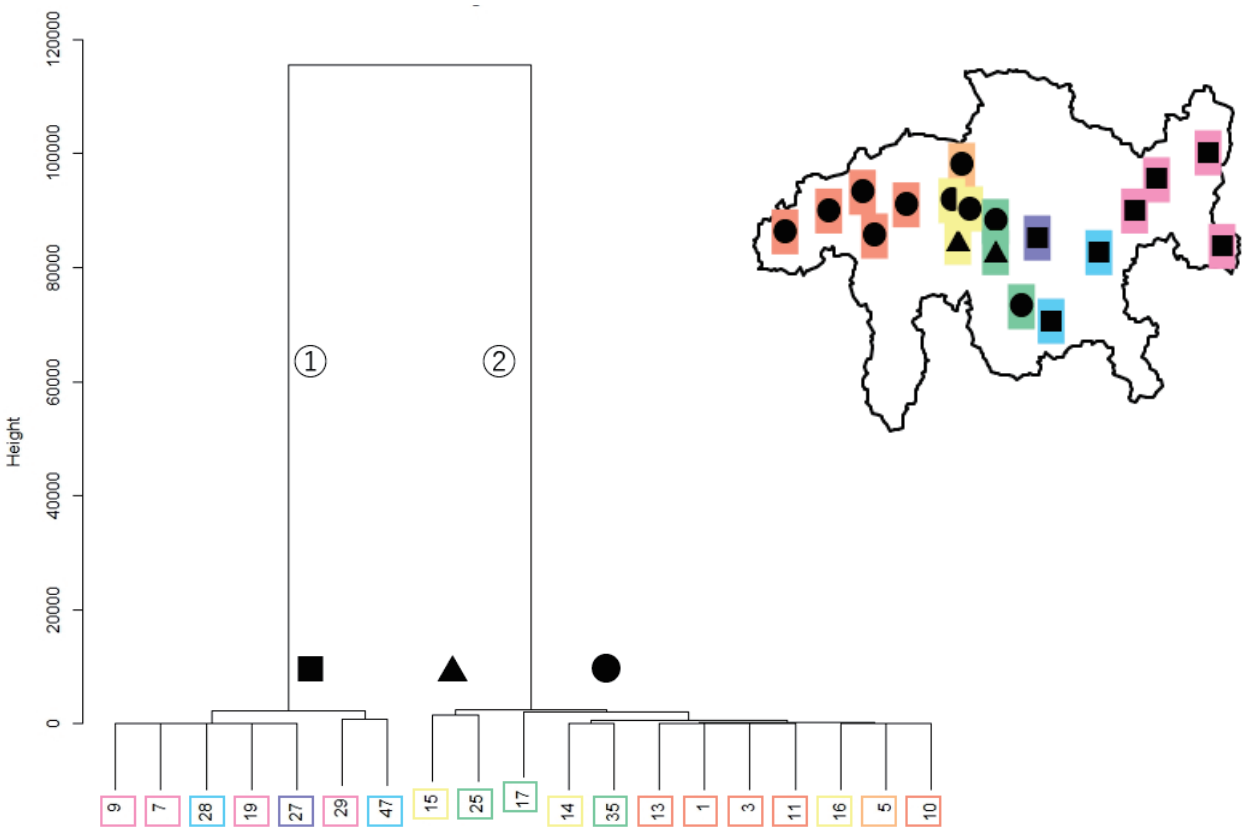

Figure 11. Dendrogram and map reflecting clustering result of Pattern 2-R

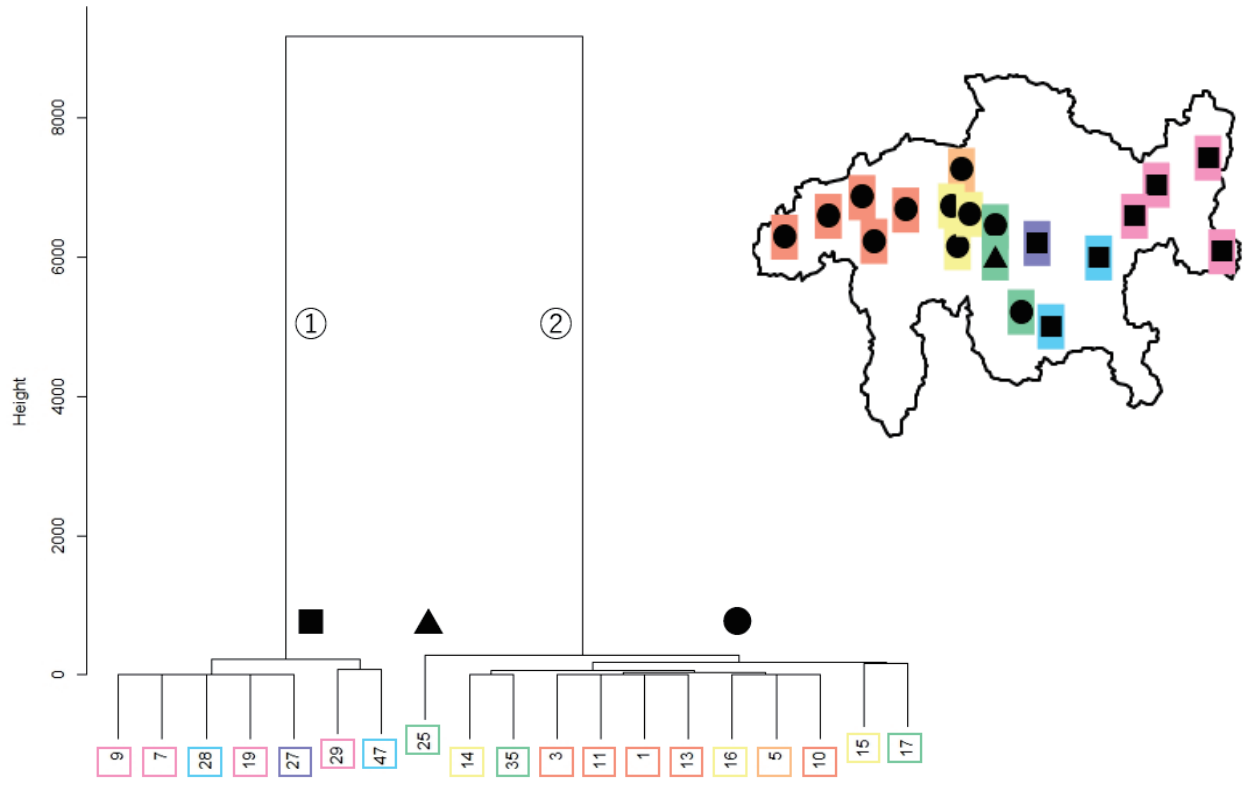

Figure 12. Dendrogram and map reflecting clustering result of Pattern 2-S

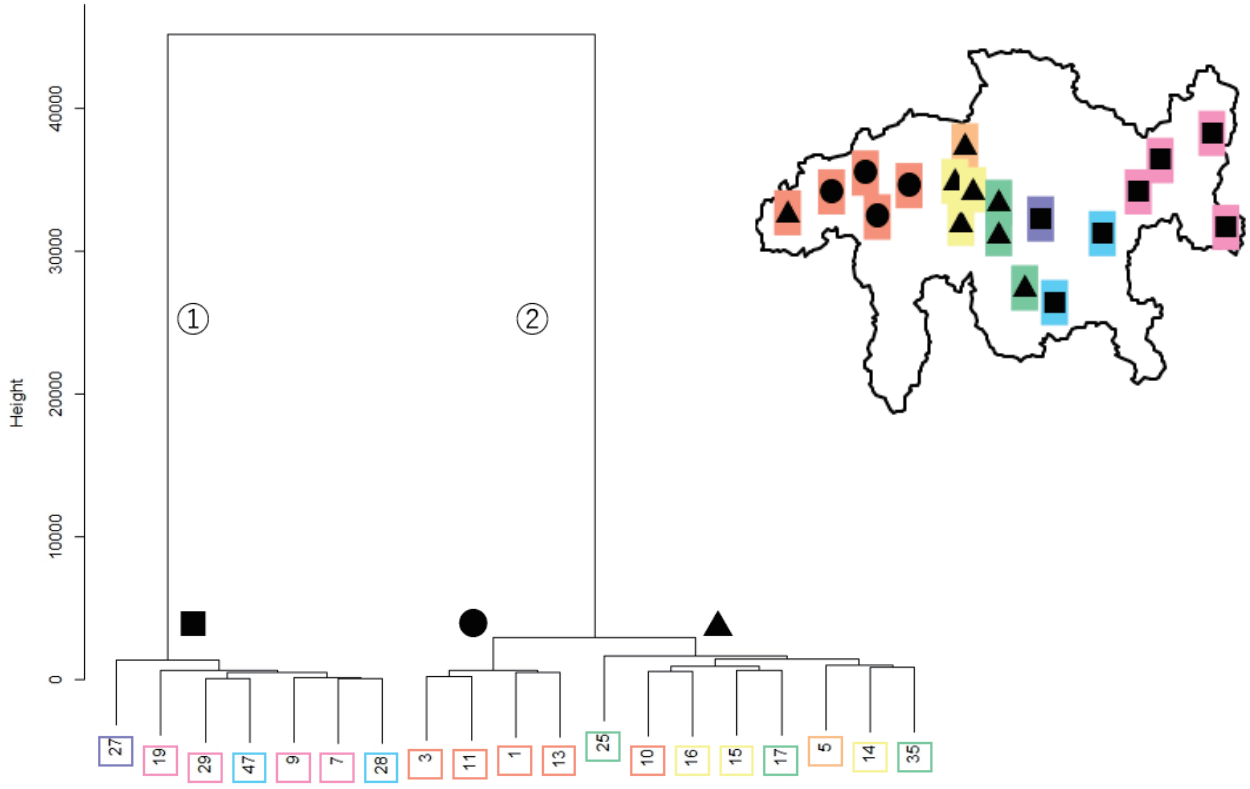

Figure 13. Dendrogram and map reflecting clustering result of Pattern 3-R

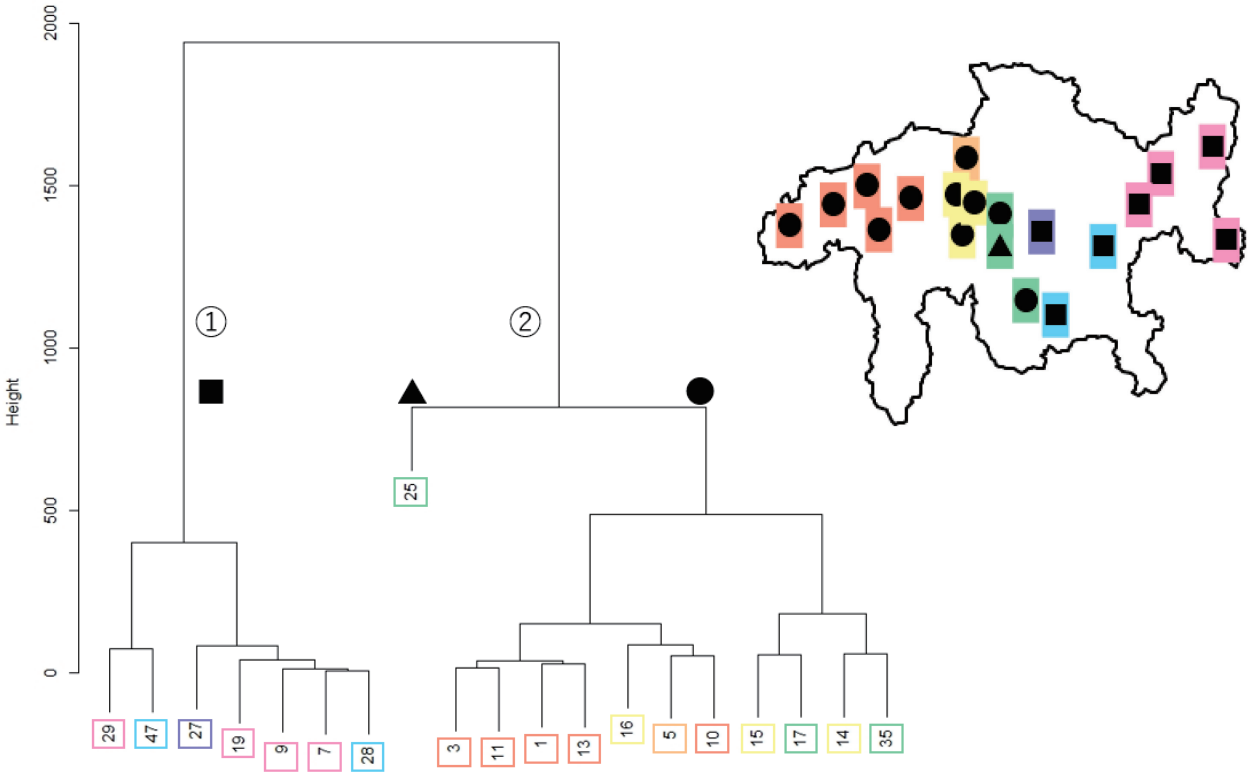

Figure 14. Dendrogram and map reflecting clustering result of Pattern 3-S

Figure 9 to 14 are the dendrograms created from each matrix pattern with each data, and the maps reflect the clustering result[13]. The dendrograms of Pattern 1-R (raw data) and Pattern 1-S (standard data) are the same from the viewpoint of the clustering process. As a result, we analyzed and compared five dendrograms: Pattern 1, Pattern 2-R (raw data), Pattern 2-S (standard data), Pattern 3-R (raw data) and Pattern 3-S (standard data).

When dividing the clustering results drawn in each dendrogram into two large clusters, the points in cluster ① and cluster ② are the same in all dendrograms. The former consisted of six Ladin points (7, 9, 19, 28, 29 and 47) and point 27, which is located in the Albula Region. The latter was composed of the other twelve points. Since it is quite difficult to make comparisons at this level, we then divided clustering results into three medium clusters ■, ▲ and ●. In all dendrograms, the points in cluster ■ were the same, although the clustering processes differed. Therefore, we first focused on the components of medium clusters ▲ and ●. In the following sections, we call clusters smaller than medium clusters, *small cluster*, and, if necessary, we place parentheses around the *AIS* points in order to indicate its components: e.g., small cluster (7 9 28).

Table 4. Components of two medium clusters ▲ and ● in each dendrogram

dendrogram	*AIS* points in the cluster ●	*AIS* points in the cluster ▲
Pattern 1	1 3 5 10 11 13 14 16 35	15 17 25
Pattern 2-R	1 3 5 10 11 13 14 16 17 35	15 25
Pattern 2-S	1 3 5 10 11 13 14 15 16 17 35	25
Pattern 3-R	1 3 11 13	5 10 14 15 16 17 25 35
Pattern 3-S	1 3 5 10 11 13 14 15 16 17 35	25

Table 4 summarizes the components of two medium clusters ▲ and ● in each dendrogram. In the dendrogram of Pattern 1, cluster ▲ consisted of points 15, 17 and 25. In the dendrogram of Pattern 2-R, medium cluster ▲ was composed of points 15 and 25 but point 17 was included in medium cluster ●. On the other hand, in the dendrograms of Pattern 2-S and of Pattern 3-S, only point 25 was the component of medium cluster ▲. The dendrogram of Pattern 3-R differed from others, consisting of seven Central Romansh points and point 10 from the Surselva Region.

We investigated the validity of dendrograms - in other words, the validity of the three matrix patterns - in the following steps. First, we investigated whether points 15 and 17 should be clustered by comparing the dendrogram of Pattern 2-R with their word forms. Second, on the hypothesis of reasonableness of the clustering of points 15 and 17, we examined if it was appropriate that points 15, 17 and 25 form a medium cluster by comparing the dendrogram of Pattern 1 with their word forms. Third, we analyzed if it was valid enough for point 5 to be clustered together with points 14 and 35 by comparing

[13] Cf. Figure 2 for the relationship between colors and idioms.

the dendrogram of Pattern 3-R with their word forms. Finally, we compared the dendrogram of Pattern 2-S and that of Pattern 3-S.

6.1. Dendrogram of Pattern 2-R

In the dendrogram of Pattern 2-R, two points 15 and 17 were in different medium cluster ● and ▲, respectively. However, they were grouped into the same cluster in the dendrograms of other Patterns [cf. Table 4].

The dialects of these two points contained a word form whose numerical value is 10 to other points' word forms. In Map 1144, at point 15, the NP was found after the verb in the imperative mood [cf. (7)]; however, at other points, point 14 as an example, the NP was used with the verb in che-clause [cf. (8)]. Also, in Map 1641, at point 17, the NP preposed the infinite verb [cf. (9)]; on the other hand, at other points, point 10 as an example, the NP was employed with the finite verb in the subordinate clause [cf. (10)]. That is, points 15 and 17 are similar in that they are significantly different from other points.

(7) wa:rdɐ becɪ k i vɔmɐn æjn ʎ iɐrt
look-IMP.2SG NEG that they go-SBJV.PRS.3PL in the garden
Be careful that they do not go into the garden. (AIS No.1144, Pt. 15)

(8) vardɐ kɐ lɐs gɐʎe(ɪ)ɲɐs vɔmɐn be(ɪ)c ɛʎ iɐrt
look-IMP.2SG that the chickens go-SBJV.PRS.3PL NEG in the garden
Be careful that the chickens do not go into the garden. (AIS No.1143&1144, Pt. 14)

(9) dɐ bec ɐvæjr cɐto: kɐlɐ donɐ
of NEG have-INF find-PST.PTCP this woman
not having found this woman. (AIS No. 1641, Pt. 17)

(10) cɐ nus væjn bec umflaw ɛlɐ
that we have-PRS.1PL NEG find-PST.PTCP her
that we have not found her. (AIS No. 1641, Pt. 10)

In addition, between these two points, three out of fifteen word forms were etymologically, phonetically, and syntactically the same [cf. Appendix]. These linguistic facts imply that points 15 and 17 should be classified in the same cluster. Nevertheless, they were classified into separate medium clusters in Pattern 2-R [cf. Figure 11]. The dendrogram created with matrix of Pattern 2 with raw data did not fully reflect the linguistic features of target dialects.

6.2. Dendrogram of Pattern 1

Pattern 1 is a matrix whose variables are the sum of numerical values (total linguistic distance). In the dendrogram of Pattern 1, points 15 and 17 formed a medium cluster ▲ with point 25. Point 25 was the only point where the compound negation *na...betg* was used in four declarative sentence maps [cf. (11)]. It means that, from a morpho-syntactical point of view, the dialect of point 25 is clearly different from other Rhenish Romansh dialects.

(11) i *n* ɐŋklej *bec*
I NEG understand-IND.PRS.1SG NEG
I do not understand. (AIS No. 1658, Pt. 25)

Table 5 is an extract from the matrix of Pattern 1[14]. Looking at the rows of points 15, 17 and 25, it is noticeable that the values in those rows are significantly higher than those of point 1, 10 14 and 35[15]. The value of *na…betg* to other points' word forms was 5 in most cases. There was a possibility that these three points (15, 17 and 25) were judged to be similar because of their large total values compared to other dialects. As a result, in Pattern 1, they were classified into cluster ▲. However, this classification does not necessarily reflect the morpho-syntactical difference of point 25 from points 15 and 17. From this, it can be said that it is unlikely that the dendrogram of Pattern 1 matches the linguistic fact, regardless of whether the variable is raw data or standardized data.

Table 5. Extract from the matrix of Pattern 1

	Pt.1	Pt.3	Pt.5	Pt.10	Pt.11	Pt.13	Pt.14	Pt.15	Pt.16	Pt.17	Pt.25	Pt.35
Pt.1	0	11	17	17	10	11	21	32	17	32	39	23
Pt.10	17	17	15	0	17	17	19	26	12	26	35	21
Pt.14	21	21	19	19	21	21	0	26	19	26	33	17
Pt.15	32	32	30	26	32	32	26	0	28	30	35	24
Pt.17	32	32	30	26	32	32	26	30	27	0	38	24
Pt.25	39	39	37	35	39	39	33	35	36	38	0	31
Pt.35	23	23	21	21	23	23	17	24	21	24	31	0

6.3. Dendrogram of Pattern 3-R

The dendrogram of Pattern 3-R was significantly different from others in that medium cluster ● was composed of only points 1, 3, 11 and 13 [cf. Figure 13]. In all of

[14] We excerpted only rows and columns of the Rhenish Romansh dialect points from the matrix.
[15] We mentioned the morpho-syntactical closeness of points 15 and 17 in section 6. 2..

other Patterns' dendrograms, these four points formed a small cluster (1 3 11 13). Among the points of Rhenish Romansh, these points made clear the difference where the NP was always put after the verb in the fifteen expressions [cf. (12) and (13)]. Concerning this aspect, it seems that medium cluster ● in the dendrogram of Pattern 3-R is well classified.

(12) sɐmuɛntɐ bʊk
move-IMP.PRS.2SG NEG
Do not move! (AIS No. 1647, Pt. 1)

(13) beʔ^c sɐmuajntɐ
NEG move-IMP.PRS.2SG
Do not move! (AIS No. 1647, Pt. 16)

However, medium cluster ▲ seems not to be well classified. In medium cluster ▲ of Pattern 3-R, point 5 was clustered with a small cluster (14 35). In other dendrograms, this point was clustered with points 10 and 16. ✓ in Table 6 indicates in which maps and points the inversion of NPs and verbs occur in the imperative sentences[16]. In points 5, 10 and 16, such inversion was seen only in map 1647. In point 14, it was seen in three maps: 1621a, 1621b and 1647. From the syntactical point of view, points 5, 10 and 16 are close to each other. In addition to this, while NP *betga* was used in these points, in point 35, NP *bitg*, a phonetical variety of *betga* and *buca*, was utilized.

Table 6. Occurrence of inversion of NPs and verbs in the imperative sentence maps

maps / points	Pt. 5 *betg*	Pt. 10 *betg*	Pt. 14 *betg*	Pt. 15 *betg*/*bitg*	Pt. 16 *betg*	Pt. 17 *betg*	Pt. 25 (*na*)*betg*	Pt. 35 *bitg*
355				✓		✓	✓	✓
1621a			✓	✓		✓	✓	✓
1621b			✓	✓		✓	✓	✓
1647	✓	✓	✓	✓	✓	✓	✓	✓

The dialect of point 5 is morpho-syntactically more similar to that of points 10 and 16 than that of points 14 and 35. In the dendrogram of Pattern 3-R, although the dialects were roughly classified in large and medium clusters, linguistic features do not seem to be fully reflected in the small clusters.

[16] This inversion occurred only in Rhenish Romansh dialects, because in Ladin dialects, NP *nu(n)* always precedes a verb.

6.4. Dendrogram of Pattern 2-S and of Pattern 3-S

The components of medium clusters ▲ and ● of both Pattern 2-S and Pattern 3-S were exactly the same, yet the clustering process of points in medium clusters ● was different. In the dendrogram of Pattern 2-S, two small clusters (1 3 11 13) and (5 10 16) clustered together, then another small cluster (14 35) joined into (1 3 5 10 11 13 16), and finally another small cluster (15 17) was added to (1 3 5 10 11 13 14 16 35) to form medium cluster ●. In this dendrogram, it is worth noting that the points 15 and 17, which can be regarded as outliers, were added last.

In the dendrogram of Pattern 3-S, the first clustering process was the same as that of Pattern 2-S. On the other hand, second and third clustering processes were quite different. Two small clusters (14 35) and (15 17) clustered. After that, (14 15 17 35) were added to small cluster (1 3 5 10 11 13 16) to form medium cluster ●. In this dendrogram, the clustering of four points 14, 15, 17 and 35 seems reasonable from a viewpoint of the inversion of a NP and verb [cf. Table 6].

However, it is difficult to determine which of two dendrograms best reflected the linguistic features of target dialects by only analyzing clusters ▲ and ●. It is necessary to analyze the two clusters (cluster ■) of these two Patterns.

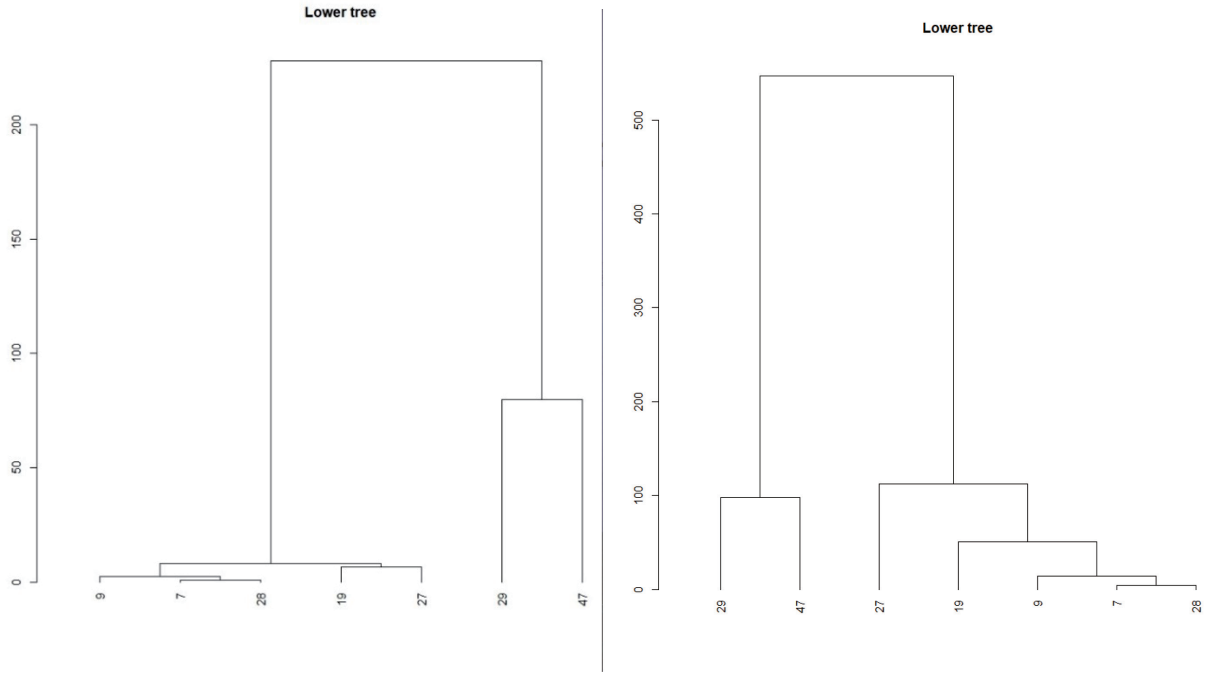

Figure 15. medium cluster ■ in the dendrogram of
Pattern 2-S (left) and Pattern 3-S (right)

Figure 15 is the enlargements of the medium cluster ■ in the dendrogram of Pattern 2-S and that of Pattern 3-S. These two medium clusters have three features in common: points 7 and 28 got clustered first; point 9 then was added to the small cluster (7 28); and

points 29 and 47 formed a small cluster.

The dialects of points 29 and 47 contained the word forms whose value is 10 to other points' word forms. In Map 355, at points 29 and 47, the imperative mood was expressed in polite form (*fuorma da curtaschia* in Vallader), cha + S + subjunctive [cf. (14) and (15)]. In the same map, at the other points, point 9 as an example, the imperative mood was expressed with a verb in imperative form [cf. (16)].

(14) cɛl nʊn jɛt
that he NEG go-SBJV-PRS-3SG
Please do not go. (AIS No. 355, Pt. 29)

(15) cɐ nʊl ɟæj
that NEG he go-SBJV-PRS-3SG
Please do not go. (AIS No. 355, Pt. 47)

(16) nʊ jɛraj
NEG go-IMP-PRS-2PL
Do not go. (AIS No. 355, Pt. 9)

The only difference between these two dendrograms is the clustering process of points 19 and 27. In the dendrogram of Pattern 2-S, these two points formed a small cluster (19 27) and then clustered with the small cluster (7 9 28). In the dendrogram of Pattern 3-S, point 19 clustered with (7 9 28), and then point 27 was added to the small cluster (7 9 19 28).

Table 7. Realization of *u* of NP *nu(n)* in points 7, 9, 19, 27 and 28

	Pt. 7	Pt.9	Pt. 19	Pt. 27	Pt. 28
[nu], [nun]	2	0	1	12	1
[nʊ], [nʊn]	13	14	14	3	14
[nɐ]	0	1	0	0	0

Table 7 shows the realization of *u* of NP *nu(n)* in five Ladin points. From this table, it is noticeable that at point 27, the vowel of the NP *nu(n)* was mostly realized as [u], a close back rounded vowel. At other points, however, it was pronounced as [ʊ], a near-close back rounded vowel, for most cases[17]. That is, point 19 is phonetically much closer to these three points than point 27. It is unlikely that points 19 and 27 form a small cluster as presented in the dendrogram of Pattern 2-S. Hence, the dendrogram of Pattern 3-S seems to reflect the linguistic characteristics of Romansh dialects among the six

[17] cf. Appendix for the word forms.

dendrograms.

7. Conclusion

In this research, “what kind of variable and matrix should be utilized in the hierarchical clustering analysis when using digitization that reflects the word form’s etymology, phonetic and syntax” was examined. In order to attain this objective, we compared and analyzed six dendrograms generated with two types of data (raw data and standardized data) and three different types of matrix patterns. Pattern 1 was the combination of the variable and the matrix which have been traditionally used in the domain of dialectometry. Pattern 2 was the matrix whose variables are the values that are digitized without any editing. Pattern 3 was the matrix in which we utilized the values as categories and whose variables were the number of occurrences in each category. In each Pattern, we used standardized data and raw data as its variable.

The dendrograms of Pattern 1-R and Pattern 1-S showed inappropriateness for grouping of point 25, the only point in which the compound negation was observed. The dendrogram of Pattern 2-R also showed inappropriateness for the grouping of point 15 and 17. Although the dialects of these points have linguistical characteristics in common, they were separated into different medium clusters in this dendrogram. Concerning the dendrogram of Pattern 3-R, we pointed out its lack of appropriateness for the grouping of point 5 with points 14 and 35 as points 10 and 16 are morpho-syntactically close to point 5. Concerning the dendrogram of Pattern 2-S, we pointed out its inappropriateness for the grouping of 19 and 27. The dialect of point 19 is phonetically close to those of points 7,9 and 28; however, in Pattern 2-S, points 19 and 27 form a small cluster. In the end, as a result, through the comparison of five dendrograms, we have reached the conclusion that the dendrogram created with the matrix of Pattern 3 and with standardized data best reflects the linguistic facts of target dialects.

It is, however, necessary to prove whether similar results can be obtained even if the number of points or expressions examined increases or decreases. In the future, we will perform a similar analysis using other *AIS* maps or maps of negative sentences covered in *Atlas Linguistique de la France* (= *ALF*, Linguistic Atlas of France). I would like to revalidate the method used and the validity of the conclusion.

Acknowledgment

This paper is based on the presentation at the 2nd conference of the Geolinguistic Society of Japan (27th September 2020) and its proceedings. I would like to thank the professors for their questions and comments.

List of abbreviations

I. Names of Atlas and Dictionary

AIS	*Sprach- und Sachatlas Italiens und der Südschweiz*
ALF	*Atlas Linguistique de la France*
DRG	*Dicziunari Rumantsch Grischun*

II. Grammatical Terms

sg./SG	singular	IND	indicative mood	PRS	present tense
pl./PL	plural	INF	infinitive	PST	past tense
dec.	declarative	IMP	imperative mood	PTCP	participle
int.	interrogative	NEG	negative	SBJV	subjunctive mood
imp.	imperative	NP	negative particle	V	finite verb

III. Technical Terms

n	number of maps	*N*	number of dialect point
Pt.	point		

References

ADAMSON, G. W. & BAWDEN, D. (1981). Comparison of hierarchical cluster analysis techniques for automatic classification of chemical structures, *Journal of Chemical Information and Computer Sciences*, 21, 204-209.

CAHANNES, G. (1924). *Grammatica romontscha per Surselva e Sutselva*, Mustér: Ligia Romontscha,

CONFORTI, C. & CUSIMANO, L. (1997). *An lingia directa 1 – Eng curs da rumàntsch sutsilvan*, Cuira: Leia Rumàntscha.

DECURTINS, A. (2012). *Lexicon romontsch cumparativ sursilvan – tudestg*, Chur: Ed. Societad Retorumantscha.

EVERITT, B. S. (1979). Unresolved problems for cluster analysis, *Biometrics* 35(1), 169-181.

EVERITT, B. S. et al. (2011). *Cluster Analysis*, 5th ed, John Wiley & Sons.

GOEBL, H. (1992). Problèmes et méthodes de la dialectométrie actuelle (avec application à l'AIS), in : Euskaltzaindia/Académie de la langue basque (ed) Nazioarteko dialektologia biltzarra. Agiriak/ *Actes du Congrès international de dialectologie (Bilbo/Bilbao 1991)*. Bilbo/Bilbao: Euskaltzaindia, 429-475.

GROSS, M. (ed) (2004). *Romanche Facts & Figures*, traduit de l'allemand par Furer, J.-J., Coire.

HAIMAN, J & BENINCA, P. (1992). *The Rhaeto-Romance Language*, London: Routledge.

HEERINGA, W. & NERBONNE, J. (2001). Dialect areas and dialect continua, *Language Variation*

and Change 13(3), 375-400.

JABERG, K. & JUD, J. (1928). *Der Sprachatlas als Forschungsinstrument: Kritische Grundlegung und Einführung in den Sprach- und sachatlas italiens und der Südschweiz,* Halle: Max Niemeyer Verlag.

—— (1960). *Index zum sprach- und sachatlas italiens und der Südschweiz: Ein propädeutisches etymologisches Wörterbuch der italienischen mundarten*, Bern: Stämpfli.

JABERG, K., JUD, J. et al. (1928-1940). *Sprach- und sachatlas italiens und der Südschweiz*, Zofingen (Schweiz: Ringier & Co.). (http://www3.pd.istc.cnr.it/navigais-web/).

KAWAGUCHI, Y. (2007). Is it possible to measure the distance between near languages? A case study of French dialects, *Langues proches – Langues collatérales*, Paris: L'Harmattan, 81-88.

—— (2020). Standardization and distance: A case study of the linguistic atlas of Champagne and Brie (ALCB), *Bamberger Beiträge zur Englischen Sprachwissenschaft* Bd.59, Berlin: Peter Lang, 269-276.

LIVER, R. (1991). *Manuel pratique de romanche sursilvan-vallader : précis de grammaire suivi d'un choix de textes*, Cuira: Ediziun Lia Rumantscha /Ligia Romontscha.

MENZLI, G. (1993). *Cuors da romontsch sursilvan 1*, Ligia Romontscha.

NERBONNE, J. et al. (1999). Edit distance and dialect proximity, *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison* 15, V-XVI.

NOGUCHI, H. (2018). *Zukai to Suuchiretsu de manabu Tahenryoukaiseki nyuumon – Big data Jidai no data Bunseki* (Introduction to Multivariate Analysis), Japanese Standards Association.

PLANTA, R. et al. (1939 -). *Dicziunari Rumantsch Grischun*, Cuera: Bischofberger.

SCHEITLIN, W. (1980). *Il pled puter: grammatica ladina d'Engadin' ota*, Ediziun da l'Union dals Grischs.

SEIMIYA, T. (submitting). Romanshugo to Furankopurovansugo no Ruijisei no Keiryou-hougengaku teki Bunseki (Dialectometrical Analysis of proximity of Romansh and Francoprovençal), To appear in *Studia Romanica* 53.

THÖNI, G. (1969). *Rumantsch – Surmeir: grammatica per igl idiom surmiran*, Ligia Romontscha.

YARIMIZU, K. et al. (2004). Multi Analysis in Dialectology – A Case study of the Standardization in the Environs of Paris, *Linguistic Informatics 3*, Tokyo University of Foreign Studies, 99-119.

Appendix. List of word forms at each point in each AIS map

No.	type of sentences	point 1	point 3	point 5	point 7	point 9	point 10	point 11	point 13	point 14	point 15	point 16	point 17	point 19	point 25	point 27	point 28	point 29	point 35	point 47
52	interrogative	bʊkɐ	buk	bekɐ	nʊ	nɐ	becɐ	buk	bug	be(ɪ)c	bɪcɐ	bekɐ	bec	nʊ	bec	nu	nʊ	nʊ	bic	nʊ
69	interrogative	brʊk	buk	be(ɪ)k	nʊ	nʊ	bec	buk	buk	be(ɪ)c	bec	bę̣?c	bec	nʊ	bec	nu	nʊ	nʊ	bic	nu
355	imperative	bugɐ	bukɐ	bek	nʊ	nʊ	bec^i	bug	bugɐ	be(ɪ)c	bɪcɐ	bec	bę̣c	nun	bec	nu	nʊ	nʊn	bic	nʊl
1621a	imperative	bʊgɐ	bugɐ	b^ʊ(ɐ)gɐ	nʊ	nʊ	be(ɪ)ɟɐ	bukɐ	bʊgɐ	be(ɪ)	becɐ	beʔ	be	nʊ	be^t	nu	nʊ	nʊ	bit	nʊ
1621b	imperative	bʊgɐ	bugɐ	b^ɐgɐ	nʊ	nʊ	be(ɪ)ɟɐ	bugɐ	bʊgɐ	be(ɪ)	becɐ	beʔ	be	nʊ	be^t	nu	nʊ	nʊ	bic	nʊ
1647	imperative	bʊk	bʊk	be(ɪ)kɐ	nʊ	nʊ	be(ɪ)cɐ	bʊk	bʊk	be(ɪ)c	bɪcɐ	beʔ^c	bet	nʊn	be^t	nu	nʊ	nʊ	bic	nʊ
653	declarative	bʊk	bʊkɐ	bekɐ	nʊ	nʊ	be(ɪ)cɐ	bugɐ	bugɐ	be(ɪ)ca	bɪcɐ	becɐ	bec	nʊn	be^t	nu	nʊ	nʊ	bic	nʊ
1615	declarative	kɐ	buk	b^ɐgɐ	nʊn	nʊn	^b^ecɐ	bugɐ	bʊgɐ	be(ɪ)c	becɐ	beʔ	bec	nʊ	be^t	nʊ	nʊ	nʊ	bic	nʊ
1630	declarative	bugɐ	bugɐ	be(ɪ)kɐ	nʊ	nʊ	becɐ	bugɐ	bukɐ	be(ɪ)^c	bɪcɐ	beʔ	bec	nʊn	be	nu	nʊ	nʊn	bi	nu
1641	declarative	buk	buk	bɐgɐ	nʊ	nʊ	bec	buk	bug	be(ɪ)	becɐ	beʔ	bec	nʊn	bec	nun	nʊ	nʊ	bi	nʊ
1651	declarative	bukɐ	buk	pek	nu	nʊ	bec	bug	buk	be(ɪ)c	bec	bec	bec	nʊn	nu bec	nu	nʊ	nʊ	bic	nʊ
1658	declarative	bʊk	buk	bek	nʊn	nʊn	bec	bʊk	bʊ(o)k	be(ɪ)c	bec	bec	bec	nʊn	n bec	nʊn	nʊn	nʊn	bic	nʊn
1678	declarative	bʊk	bʊk	be(ɪ)k	nu	nʊ	becɐ	bʊk	buk	be(ɪ)c	bɪc	beʔ^c	bec	nʊn	nɐ bec	nu	nu	nʊ	bic	nu
1144	declarative	bʊk	bukɐ	be(ɪ)k	nʊn	nʊ	bec	bugɐ	buk	be(ɪ)c	becɪ	bec	bec	nʊn	nɐ bec	nʊ	nʊn	nʊ	bic	nʊ
1278	declarative	bugɐ	buk	be(ɪ)k	nʊ	nʊ	becɐ	bʊʔk	bʊgɐ	be(ɪ)c	becɐ	beʔ	bec	nʊ	be^t	nu	nʊ	nʊn	bic	nʊn
		red : V+Neg				blue : Neg+V				green : Neg+V+Neg				black : da + Neg + INF etc.						